

# Modern Statistics for Spatial Point Processes\*

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.

*Key words:* 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 (Ripley, 1981, 1988; Diggle, 1983; Stoyan, Kendall & Mecke, 1995; Stoyan & Stoyan, 1995) usually deal with relatively small point patterns, where the assumption of stationarity is central and 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 *et al.* (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,

\* This paper was presented at the 21st Nordic Conference on Mathematical Statistics, Rebild, Denmark, June 2006 (NordStat 2006).

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 <http://www.math.aau.dk/~rw/sppcode>. As we shall often refer to our own monograph, please notice the comments and corrections to Møller & Waagepetersen (2003b) at <http://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 sections. 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, as it is only possible to observe whales within the vicinity of the ship. Moreover, whales within sighting distance may fail to be observed because of 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. The data in Fig. 1 do not reflect the fact that the whales move and that the whales are observed at different points in time. However, the observations from different transect lines may be considered approximately independent due to the large spatial separation between the transect lines. More details on the data set and analysis of line transect data can be found in Skaug, Øien, Schweder & Bøthun (2004), Waagepetersen & Schweder (2006), and Buckland *et al.* (2004). 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 (2006) is described in example 2 and analyzed in examples 8 and 9.

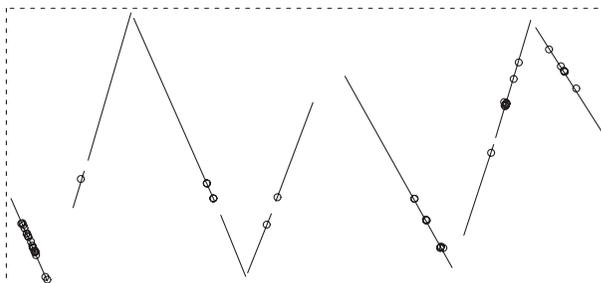


Fig. 1. Observed whales along transect lines. The enclosing rectangle is of dimensions 263 km by 116 km.

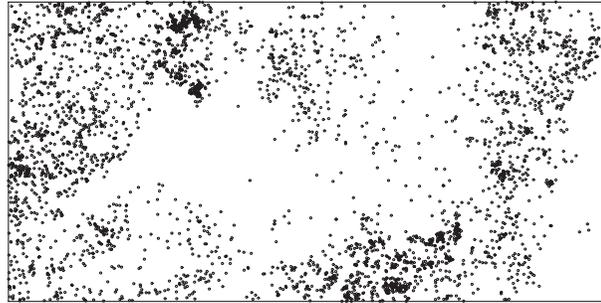


Fig. 2. Locations of *Beilschmiedia pendula Lauraceae* trees observed in a 1000 m by 500 m rectangular window.

In studies of biodiversity of tropical rainforests, 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 rainforest 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 hundreds of species; see Hubbell & Foster (1983), Condit *et al.* (1996), and Condit (1998). In addition to the tree positions, covariate information on altitude and norm of altitude gradient is available, see Fig. 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 introduced in examples 1 and 3 and further considered in Fig. 9 and examples 10, 11, 13 and 14.

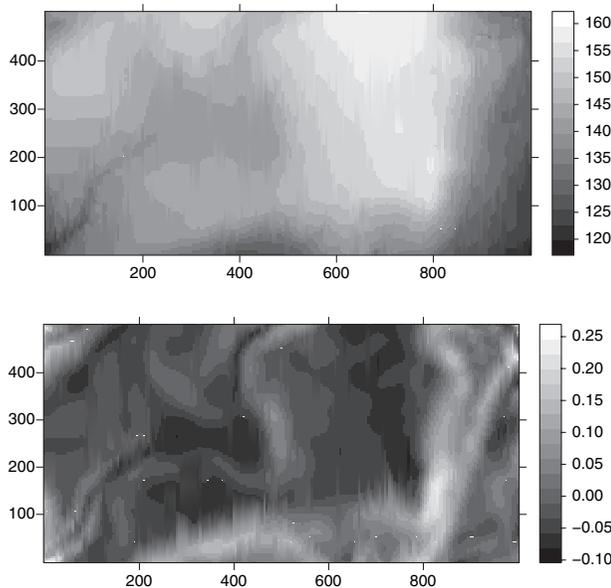


Fig. 3. Altitude (upper plot) in meter and norm of altitude gradient (lower plot).

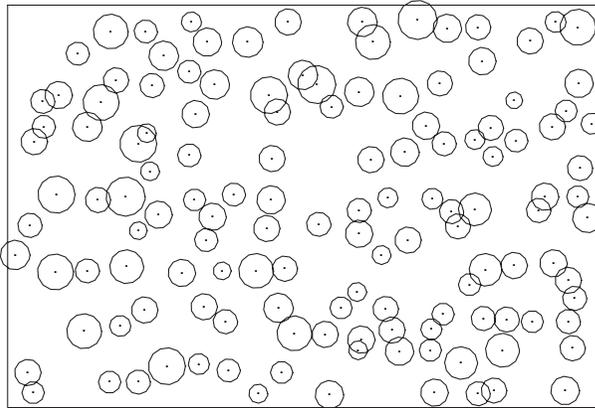


Fig. 4. Norwegian spruces observed in a rectangular 56 m by 38 m window. The radii of the discs equal five times the stem diameters.

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 collected in Tharandter Forest, Germany, by the forester G. Klier and was first analyzed using point process methods by Fiksel (1984). The data set is an example of a marked point pattern, with points given by the tree locations and marks by the stem diameters. The discs in Fig. 4 are of radii five times the stem diameters and may be thought of as 'influence zones' of the trees; see Penttinen *et al.* (1992) and Goulard *et al.* (1996). The regularity in the point pattern is to a large extent due to forest management. 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 observing a tree at a given location and of given stem diameter conditional on the neighbouring trees. In example 4, 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. The spruces are also considered in Fig. 10 and example 6.

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 Fig. 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 *Cataglyphis* ants, on the other hand, feed on dead *Messors* and hence the positions of *Messor* nests might affect the choice of sites for *Cataglyphis* nests. Högmänder & Särkkä (1999) 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, Figs. 7–8, and examples 7 and 12.

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 rainforest trees and the whales), *inhomogeneity*, e.g. caused by a thinning mechanism or covariates (as for the whales and tropical rainforest trees), and *interaction* between points, where the interaction possibly depends on marks associated with the points

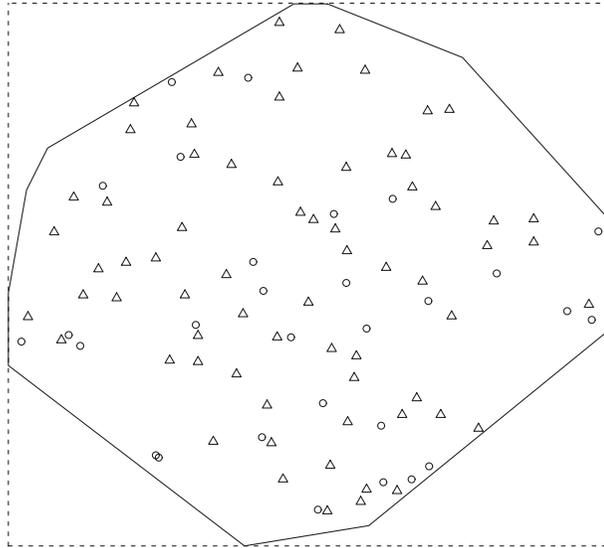


Fig. 5. Locations of nests for *Messor* (triangles) and *Cataglyphis* (circles) ants. Enclosing rectangle for observation window is 414.5 ft by 383 ft.

(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*  $\mathbf{x} = \{x_1, \dots, x_n\}$  of  $n \geq 0$  points contained in  $S$ . We say that the *point process is defined on*  $S$ , and we write  $\mathbf{x} = \emptyset$  for the empty point pattern. To specify the distribution of  $\mathbf{X}$ , one may specify the distribution of the number of points,  $n(\mathbf{X})$ , and for each  $n \geq 1$ , conditional on  $n(\mathbf{X}) = n$ , the joint distribution of the  $n$  points in  $\mathbf{X}$ . An equivalent approach is to specify the distribution of 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 them 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. As  $\mathbf{X} \setminus W$  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*

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

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

$$\mu(B) = \int_B \rho(u) du$$

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

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

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

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  $C \subseteq \mathbb{R}^2$ ,

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

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

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

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

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

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

$$E \sum_{\substack{\neq \\ u_1, \dots, u_n \in \mathbf{X}}} h(u_1, \dots, u_n) = \int h(u_1, \dots, u_n) \rho^{(n)}(u_1, \dots, u_n) du_1, \dots, du_n \tag{4}$$

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 *et al.*, 2000 and section titled ‘second-order summary statistics’), and in applications it is often assumed that  $g(u, v) = g(\|u - v\|)$ , i.e. that  $g$  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; Diggle, 2003, section 5.8.3).

Suppose  $\pi(u) \in [0, 1]$ ,  $u \in \mathbb{R}^2$ , are given numbers. An *independent  $\pi$ -thinning* of  $\mathbf{X}$  is obtained by independently retaining each point  $u$  in  $\mathbf{X}$  with probability  $\pi(u)$ . It follows easily from (4) that  $\pi(u_1) \dots \pi(u_n) \rho^{(n)}(u_1, \dots, u_n)$  is the  $n$ th 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, for example the radius of a disc as in Fig. 4, the type of ants as in Fig 5, or another point process (e.g. the clusters in a shot noise Cox process, see section titled ‘Shot noise Cox processes’). The process  $\Phi = \{(u, m_u) : u \in \mathbf{X}\}$  is called a *marked point process*, and indeed marked point processes are important for many applications; see Stoyan & Stoyan (1995), Stoyan & Walder (2000), Schlather *et al.* (2004), Schlather (2001), Moller & Waagepetersen (2003b). For the models presented later in this paper, the marked point process model of discs in Fig. 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 Fig. 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;

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

$z(u) = (z_1(u), \dots, z_k(u))$  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);

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

$\theta$  is the vector of all parameters (including  $\beta$ ) 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 by 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  $\mu$  and intensity function  $\rho$  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,

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

reflecting the lack of interaction. Stationarity means that  $\rho(u)$  is constant, and implies isotropy 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 (Cox, 1972),

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

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  $\Lambda = (\Lambda(u))_{u \in S}$ , such that conditional on  $\Lambda$ ,  $\mathbf{X}$  is a Poisson process with intensity function  $\Lambda$  (Cox, 1955; Matérn, 1971; Grandell, 1976; Daley & Vere-Jones, 2003).

Three points of statistical importance should be noticed. First, though  $\Lambda$  may be modelling a random environmental heterogeneity,  $\mathbf{X}$  is stationary if  $\Lambda$  is stationary. Second, we cannot distinguish the Cox process  $\mathbf{X}$  from its corresponding Poisson process  $\mathbf{X}|\Lambda$  when only one realization of  $\mathbf{X}_W$  is available, cf. 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.

*Log Gaussian Cox processes.* In analogy with random effect models, as an extension of the log linear model (7), take

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

where  $\Psi = (\Psi(u))_{u \in S}$  is a zero-mean Gaussian process. Then we call  $\mathbf{X}$  a *log Gaussian Cox process* (Møller *et al.*, 1998). The covariance function  $c(u, v) = \text{cov}[\Psi(u), \Psi(v)]$  typically depends on some lower-dimensional parameter; see, e.g. example 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

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

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  $\rho$  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 1 (Log Gaussian Cox process model for tropical rainforest trees).* For the tropical rainforest trees in Fig. 2, we consider in example 10 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 Fig 3. An exponential covariance function  $c(u, v) = \sigma^2 \exp(-\|u - v\|/\alpha)$  is used for the Gaussian process, where  $\sigma$  and  $\alpha$  are positive parameters.

*Shot noise Cox processes.* A shot noise Cox process  $\mathbf{X}$  has random intensity function

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

where  $c \in \mathbb{R}^2$ ,  $\gamma > 0$ ,  $\Phi$  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  $\gamma k(c, \cdot)$ ,  $(c, \gamma) \in \Phi$ , where we interpret  $\mathbf{X}_{(c, \gamma)}$  as a cluster with centre  $c$  and mean number of points  $\gamma$ . Thus  $\mathbf{X}$  is an example of a *Poisson cluster process* (Bartlett, 1964), 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 2 (Shot noise Cox process for minke whales).* In Waagepetersen & Schweder (2006), the positions of minke whales in Fig. 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  $\Lambda(u) = p(u) \sum_{(c, \gamma) \in \Phi} \gamma k(c, u)$ . The cluster centres are assumed to form a stationary Poisson process with intensity  $\kappa$ , the  $c$ 's are independent of the  $\gamma$ 's, and the  $\gamma$ 's are i.i.d. gamma random variables with mean  $\alpha$  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  $\kappa$  and the  $\gamma$ 's are all equal to a positive parameter  $\alpha$  (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 (*modified*) *Thomas process* (Thomas, 1949). A Neyman–Scott process is stationary with intensity  $\rho = \alpha\kappa$ , and the Thomas process is also isotropic with

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

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  $\Phi$  by a Poisson or non-Poisson process model for the points  $(c, \gamma, b)$  (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

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

(Waagepetersen, 2007). 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  $\beta$ .

*Example 3 (Inhomogeneous Thomas model for tropical rainforest trees).* In addition to the log Gaussian Cox process model for the tropical trees (example 1), we consider an inhomogeneous Thomas process of the form (12),

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

where now  $\Phi$  denotes a stationary Poisson process with intensity  $\kappa$ . 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  $\mu$  and intensity function  $\rho$ . 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,

$$P(\mathbf{Y}_\rho \in F) = \sum_{n=0}^{\infty} \frac{e^{-\mu(S)}}{n!} \int_{S^n} \mathbf{1}[\mathbf{x} \in F] \rho(x_1) \dots \rho(x_n) dx_1 \dots dx_n \tag{14}$$

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

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

$$\begin{aligned} P(\mathbf{X} \in F) &= E[\mathbf{1}[\mathbf{Y}_1 \in F] f(\mathbf{Y}_1)] \\ &= \sum_{n=0}^{\infty} \frac{e^{-|S|}}{n!} \int_{S^n} \mathbf{1}[\mathbf{x} \in F] f(\mathbf{x}) dx_1 \dots dx_n \end{aligned} \tag{15}$$

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

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

for any  $n \in \mathbb{N}$  and pairwise different points  $u_1, \dots, u_n \in S$ . 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

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

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 expressible in closed form. For the Poisson process  $\mathbf{Y}_\rho$ , (14) gives

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

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

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

which in general is not expressible in 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.

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

for finite point configurations  $\mathbf{x} \subset S$ . This 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  $\lambda$  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),

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

Recall that the conditional probability  $P(A | \mathbf{x})$  of an event  $A$  given  $\mathbf{X} = \mathbf{x}$  satisfies  $P(A) = E[P(A | \mathbf{X})]$ . Thus, due to the infinitesimal interpretation of  $\rho(u) \, du$  (section 3.2), it follows from (22) that  $\lambda(u, \mathbf{x}) \, du$  may be interpreted as the conditional probability that there is a point of the process in an infinitesimally small 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  $Eh(\mathbf{Y}_1)$  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

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

for all  $u \in S$  and finite  $\mathbf{x} \subset S$ . Local stability implies both that  $h$  is hereditary and integrable with respect to the 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

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

and higher-order potentials are of the form

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

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

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

where

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

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  $\|u - v\| < r$ .

The *Hammersley–Clifford theorem* for Markov random fields was modified to the case of spatial point processes by Ripley & Kelly (1977), stating that 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 to parameter estimation and simulation.

The property (I) 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 4 (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 Fig. 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  $\psi$  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  $\beta_k$  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 (Hierarchical model for ants' nests).* The hierarchical model in Högmänder & Särkkä (1999) for the positions of ants' nests is based on so-called *Strauss processes with hard cores* as described below.

For distances  $t > 0$ , define

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

where  $R$  is the interaction range and  $r \geq 0$  denotes a hard core distance (or no hard core if  $r = 0$ ). For the *Messor* nests, the Strauss 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  $\beta_M$  and  $\psi_M$  times the number of neighbouring nests within distance  $R$ . 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, which are biased maximum likelihood estimates.

Using positive hard cores  $r_M$  and  $r_C$  may be viewed as an *ad-hoc* approach to obtain a model that is well-defined for all real values of the parameters  $\beta_M, \beta_C, \psi_M, \psi_{CM}$ , and  $\psi_C$ , whereby both repulsive and attractive interaction within and between the two types of ants can be modelled. 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 use maximum likelihood estimates  $r_M = 9.35$  and  $r_C = 2.45$  (distances are measured in ft) but in contrast to Högmänder & Särkkä (1999) 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 covariate function  $z_2(u) = \sum_{v \in \mathbf{x}_M} V(\|u - v\|; 0)$  for the *Cataglyphis* model with  $r_{CM} = 0$ . To facilitate comparison with previous studies we fix  $R$  at the value 45 used in Högmänder & Särkkä (1999). Note however, that profile-pseudo-likelihood computations indicate that more appropriate intra-species interaction ranges would be 15 and 30 for the *Messor* and *Cataglyphis* ants, respectively.

#### 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*

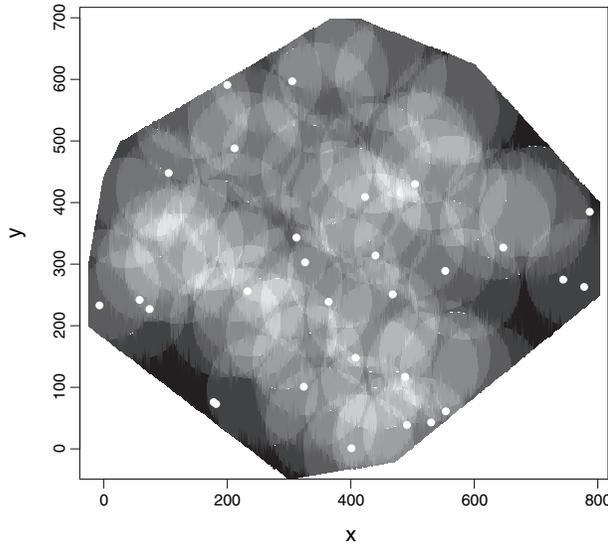


Fig. 6. The covariate function  $z_2(u) = \sum_{v \in \mathbf{x}_M} V(\|u - v\|; 0)$  for the *Cataglyphis* model (light greyscales correspond to high values of  $z_2(u)$ ). The dots show the locations of *Cataglyphis* nests.

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

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

for all non-negative functions  $h$ , then  $\lambda$  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*

$$E \sum_{x_1, \dots, x_n \in \mathbf{X}}^{\neq} h(x_1, \dots, x_n, \mathbf{X} \setminus \{x_1, \dots, x_n\}) = \int \dots \int E[\lambda(x_1, \mathbf{X})\lambda(x_2, \mathbf{X} \cup \{x_1\}) \dots \lambda(x_n, \mathbf{X} \cup \{x_1, \dots, x_{n-1}\})h(x_1, \dots, x_n, \mathbf{X})] dx_1 \dots dx_n \tag{32}$$

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

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

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

For instance, for a Cox process driven by  $\Lambda$ ,

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

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  $u \in \mathbb{R}^2$ ,

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

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  $\lambda$  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, to deal with edge effects, we may 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 section titled ‘Edge effects’ 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 that 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 criticize 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  $\rho$  is constant on  $S$ , a natural unbiased estimate is  $\hat{\rho} = n/|W|$ . In the inhomogeneous case, a non-parametric kernel estimate is

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

where  $k$  is a kernel with finite bandwidth, and where the denominator is an edge correction factor ensuring that  $\int_W \hat{\rho}(u) du$  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  $\theta$  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 generalized linear model (GLM), while the higher order potentials are roughly analogous to the distribution of the errors in a GLM. Recently, Baddeley *et al.* (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 expressible in 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.

*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 titled 'Edge effects'.

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

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

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- $\lambda$*  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  $dI(u) = -\lambda(u, \mathbf{X})h(u, \mathbf{X}) du$  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

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

The *h-weighted residual* is defined by

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

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- $\lambda$  residuals are

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

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

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

respectively. In order that the Pearson and inverse- $\lambda$  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 *et al.* (2006).

Similarly, we define innovations and residuals based on  $\rho$ , where in all expressions above we replace  $\lambda$  and  $\hat{\lambda}$  by  $\rho$  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  $\int_W h(u)\rho(u) du < \infty$ , so that (37) also holds in this case.

*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  $\lambda$ ; if we instead consider residuals based on  $\rho$ , 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 (example 5) fitted in example 12 and based on raw residuals ( $h \equiv 1$ ). The plots are corrected for edge effects, cf. section titled ‘Edge effects’.

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 greyscales should correspond to low/high values of the conditional intensity, and in regions of the same greyscale the circles should be uniformly distributed. The upper left plot in Fig. 7 is a mark plot for the raw residuals obtained from the model fitted to the *Cataglyphis* nests in example 12. In this case, the circles are of the same radii and just show the locations of the nests.

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

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

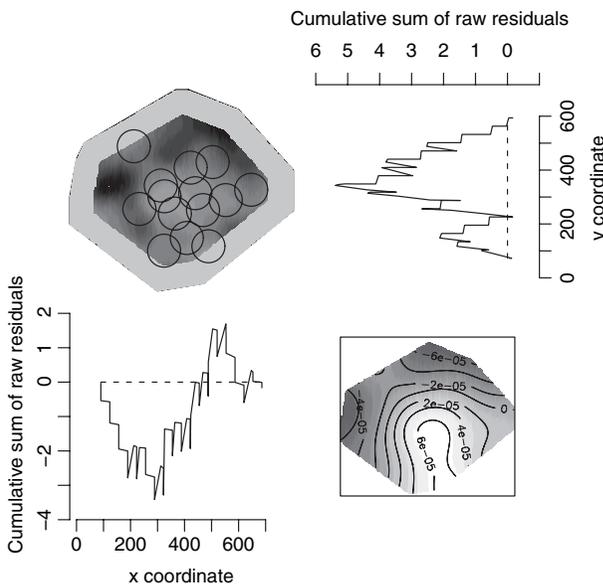


Fig. 7. Plots for *Cataglyphis* nests based on raw residuals: mark plot (upper left), lurking variable plots for covariates given by  $y$  and  $x$  coordinates (upper right, lower left), and smoothed residual field (lower right). Dark greyscales correspond to small values.

where  $k$  is a kernel and the denominator is an edge correction factor. For example, for raw residuals, the numerator of (39) has mean  $\int_W k(u - v) \mathbb{E}[\lambda(v, \mathbf{X}) - \hat{\lambda}(v, \mathbf{X})] dv$ , 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. For example, the lower right plot in Fig. 7 suggests some underestimation of the conditional intensity at the middle lower part of the plot and overestimation in the left and 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’  $A(t) = R_{\hat{\lambda}}(W(t))$  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 Fig. 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 Fig. 7 might be related to inhomogeneity; the observation window consists of a ‘field’ and a ‘scrub’ part divided by a boundary that 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  $s(u, \mathbf{x}^{(1)}), \dots, s(u, \mathbf{x}^{(n)})$ , where  $\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(n)}$  are simulations from the fitted model. This is done using a grid of fixed locations  $u_j \in W, j = 1, \dots, J$ . For each  $k = 0, \dots, n$ , where  $\mathbf{x}^{(0)} = \mathbf{x}$  is the data, we sort  $s_j^{(k)} = s(u_j, \mathbf{x}^{(k)})$ ,  $j = 1, \dots, J$  to obtain the order statistics  $s_{[1]}^{(k)} \leq \dots \leq s_{[J]}^{(k)}$ . We then plot  $s_{[j]}^{(0)}$  versus the estimated expected empirical quantile  $\sum_{k=1}^n s_{[j]}^{(k)} / n$  for  $j = 1, \dots, J$ . The Q-Q plot in Fig. 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.

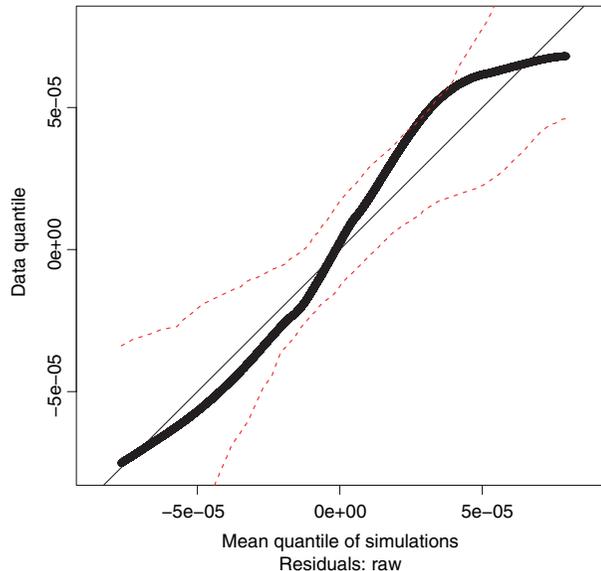


Fig. 8. Q-Q plot for *Cataglyphis* nests based on smoothed raw residual field. The dotted lines show the 2.5% and 97.5% percentiles for the simulated order statistics.

*Edge effects.* Substantial bias and other artifacts in the diagnostic plots for residuals based on  $\lambda$  may occur if edge effects are ignored. We therefore use the border method as follows (see also Baddeley *et al.*, 2007a). 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 Fig. 7.

For residuals based on  $\rho$  instead, we have no edge effects, so no adjustment of the diagnostic tools in the section ‘Diagnostic plots’ is needed.

### 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 *et al.* (2007b) who develop residual versions of such summary statistics.

*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  $g$  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)

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

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

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

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  $K(r) = 2\pi \int_0^r sg(s) \, ds$ , so that  $g$  and  $K$  are in a one-to-one correspondence. In the case of a stationary point process, 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 bandwidth. Various edge corrections have been suggested, the simplest and most widely applicable being

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

where  $W_u$  is  $W$  translated by  $u$ , and  $\hat{\rho}$  is an estimate of the intensity function. The choice of  $\hat{\rho}$  is crucial since this determines how much variation is attributed to inhomogeneity rather than random clustering. In general we prefer to use a parametric estimate of the intensity

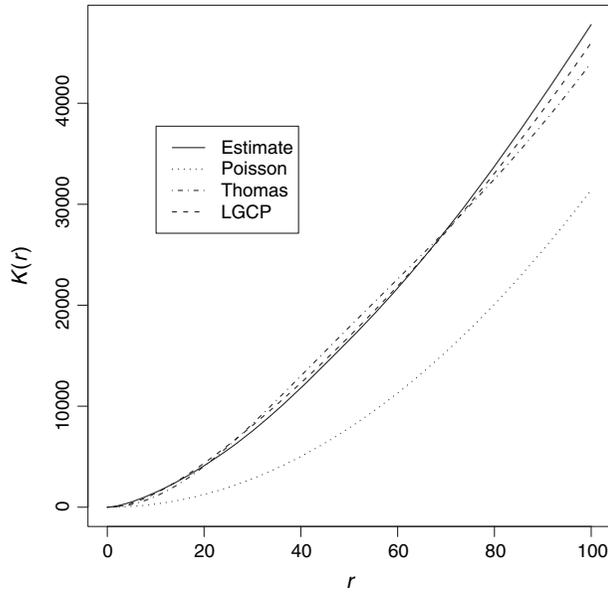


Fig. 9. Estimated  $K$ -function for tropical rainforest trees and theoretical  $K$ -functions for fitted Thomas, log Gaussian Cox, and Poisson processes.

function. Another possibility is the non-parametric estimate of  $\rho$  given in (35) but the degree of clustering exhibited by the resulting estimate  $\hat{K}(r)$  is very sensitive to the choice of kernel band-width.

An estimate of the  $K$ -function for the tropical rainforest trees obtained with a parametric estimate of the intensity function (see example 11) is shown in Fig. 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 13). 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 at least for a stationary Poisson process 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 Figs 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), Schladitz & Baddeley (2000) and Guan, Sherman & Calvin (2006).

*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  $\mathbf{X}$ ,

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

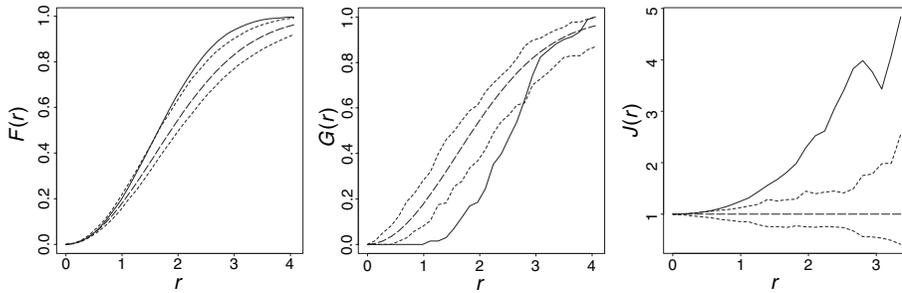


Fig. 10. Left to right: estimated  $F$ ,  $G$ , and  $J$ -functions for the Norwegian spruces (solid lines) and 95% envelopes calculated from simulations of a homogeneous Poisson process (dashed lines) with expected number of points equal to the observed number of points. The long-dashed curves show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

The *nearest-neighbour function* is defined by

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

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  $G$  accounting 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  $J$ . Estimates of  $F$ ,  $G$ , and  $J$  for the positions of Norwegian spruces shown in Fig. 10 provide evidence of repulsion.

**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  $\rho_\theta$ , the log likelihood function is

$$l(\theta) = \sum_{u \in X} \log \rho_\theta(u) - \int_W \rho_\theta(u) du, \tag{42}$$

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 of 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 (2007) 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 11.

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

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

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

$$u(\theta) = t(\mathbf{x}) - E_\theta t(\mathbf{X}), \quad j(\theta) = \text{var}_\theta t(\mathbf{X}),$$

where  $E_\theta$  and  $\text{var}_\theta$  denote expectation and variance with respect to  $\mathbf{X} \sim f_\theta$ .

Consider a fixed reference parameter value  $\theta_0$ . The score function and observed information may then be evaluated using the *importance sampling formula*

$$E_\theta k(\mathbf{X}) = E_{\theta_0} [k(\mathbf{X}) \exp(t(\mathbf{X})(\theta - \theta_0)^\top)] / (c_\theta / c_{\theta_0}) \tag{43}$$

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

$$c_\theta / c_{\theta_0} = E_{\theta_0} \left[ \exp\left(t(\mathbf{X})\theta_0^\top\right) \right]. \tag{44}$$

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  $E_{\theta_0}[\dots]$  using MCMC samples from  $f_{\theta_0}$ , see section 9.2.

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

$$\log(c_\theta / c_{\theta_0}) = \int_0^1 E_{\theta(s)} t(\mathbf{X}) (d\theta(s)/ds)^\top ds \tag{45}$$

provides an alternative and often numerically more stable way of computing a ratio of normalizing constants. Here  $\theta(s)$  is a differentiable curve, e.g. a straight line segment, connecting  $\theta_0 = \theta(0)$  and  $\theta = \theta(1)$ . The log ratio of normalizing constants is approximated by evaluating the outer integral in (45) 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

$$L(\theta) = E_{\theta} f_{W, \theta}(\mathbf{x} | \mathbf{X}_{\partial W})$$

may be computed using a missing data approach, see Geyer (1999) and Møller & Waagepetersen (2003b). A simpler but less efficient alternative is the border method, considering the conditional likelihood function

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

where the score, observed information, and likelihood ratios may be computed by analogy with the  $W = S$  case, cf. sections 5.3–5.4. These and other approaches for handling edge effects are discussed in Møller & Waagepetersen (2003b).

For a fixed  $R$ , the approximate (conditional) likelihood function can be maximized with respect to  $\theta$  using Newton–Raphson updates. In our experience the Newton–Raphson updates converge quickly, and in the examples below, the computing times for obtaining an 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 & 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 6 (Maximum likelihood estimation for overlap interaction model).* For the overlap interaction model in example 4, 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  $W \ominus 2b$ , as trees with locations outside  $W$  do not interact with trees located inside  $W \ominus 2b$ . The conditional MLE is given by  $(\hat{\beta}_1, \dots, \hat{\beta}_6) = (-1.02, -0.41, 0.60, -0.67, -0.58, -0.22)$  and  $\hat{\psi} = -1.13$ . Confidence intervals for  $\psi$  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 Fig. 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 Fig. 11 (right plot).

*Example 7 (Maximum likelihood estimation for ants' nests).* Högmander & Särkkä (1999) consider a subset of the data in Fig. 5 within a rectangular region, and they condition on the observed number of points for the two species when computing MLE's and MPLE's for the hierarchical model described in example 5, whereby the parameters  $\beta_M$  and  $\beta_C$  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 12 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  $\hat{\psi}_M = -0.06$  indicate a weak repulsion within the *Messor* nests, and the MLE's  $\hat{\beta}_C = -9.24$ ,  $\hat{\psi}_{CM} = 0.04$ , and  $\hat{\psi}_C = -0.39$  indicate positive association between *Messor* and *Cataglyphis* nests, and repulsion within the *Cataglyphis* nests. Confidence intervals for  $\psi_{CM}$  are  $[-0.20, 0.28]$  (based on

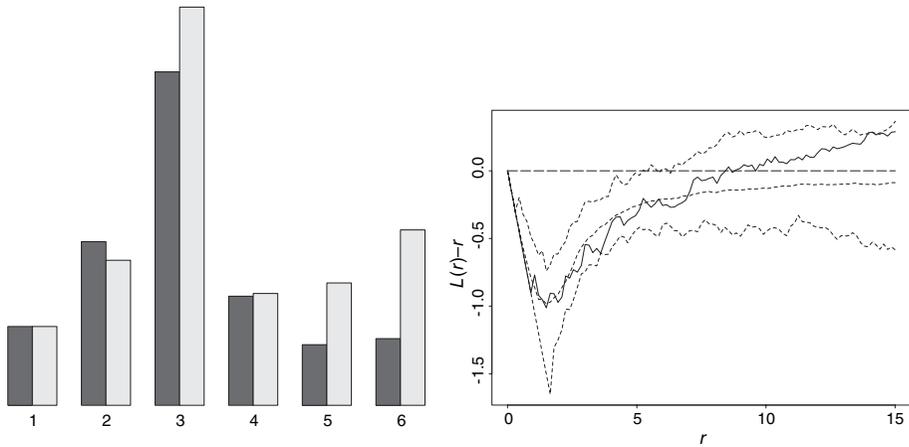


Fig. 11. Model assessment for Norwegian spruces (example 6). Dark grey bars: frequencies of trees for the six size classes (scaled so that light and dark bars are of the same height for the first class). Light grey bars: MLE of  $\exp(\beta_k)$ ,  $k = 1, \dots, 6$ . Right plot: estimated  $L(r) - r$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

observed information) and  $[-0.16, 0.30]$  (parametric bootstrap). To avoid the phase transition property of the Strauss hard core process (example 5), we restrict  $\psi_C \leq 0$  in the Monte Carlo computations for the bootstrap simulated data sets. The results in Högmänder & Särkkä (1999) differ from ours, since they estimate a stronger repulsion within the *Cataglyphis* nests and a weak repulsion between the two species. This seems partly due to the fact that Högmänder & Särkkä (1999) use a smaller observation window which excludes a pair of very close *Cataglyphis* nests, see also example 12.

7.3. Cox process models

We consider MLE for shot noise Cox processes and log Gaussian Cox processes.

In the case of a shot noise Cox process (section titled ‘Shot noise Cox processes’), suppose that the parameter vector  $\theta = (\alpha, \omega)$  consists of components  $\alpha$  and  $\omega$  parameterizing respectively the intensity function  $\zeta_\alpha$  of  $\Phi$  and the kernel  $k(c, \cdot) = k(c, \cdot; \omega)$ . Let  $f(\mathbf{x} | \Lambda)$  denote the Poisson density of  $\mathbf{X}_W$  given  $\Lambda(\cdot) = \Lambda(\cdot; \Phi, \omega)$ . 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

$$L(\theta) = E_\alpha f(\mathbf{x} | \Lambda(\cdot; \Phi, \omega)) = E_\alpha f(\mathbf{x} | \Lambda(\cdot; \Phi_{\tilde{W}}, \omega))$$

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

$$L(\theta)/L(\theta_0) = E_{\alpha_0} \left[ \frac{f(\mathbf{x} | \Lambda(\cdot; \Phi_{\tilde{W}}, \omega)) f_{\tilde{W}}(\Phi_{\tilde{W}}; \alpha)}{f(\mathbf{x} | \Lambda(\cdot; \Phi_{\tilde{W}}, \omega_0)) f_{\tilde{W}}(\Phi_{\tilde{W}}; \alpha_0)} \Bigg| \mathbf{X}_W = \mathbf{x} \right] \tag{46}$$

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

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

The score function and observed information are given by

$$u(\theta) = E_{\theta}[V_{\theta, \mathbf{x}}(\Phi_{\tilde{W}}) | \mathbf{X}_W = \mathbf{x}]$$

and

$$j(\theta) = -E_{\theta} \left[ dV_{\theta, \mathbf{x}}(\Phi_{\tilde{W}}) / d\theta^T | \mathbf{X}_W = \mathbf{x} \right] - \text{var}_{\theta} [V_{\theta, \mathbf{x}}(\Phi_{\tilde{W}}) | \mathbf{X}_W = \mathbf{x}].$$

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 (see section titled ‘Log Gaussian Cox processes’), 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 8 (Maximum likelihood estimation for North Atlantic whales).* For the shot noise Cox process model in example 2, the unknown parameters are the intensity  $\kappa$  of the cluster centres, the mean number  $\alpha$  of whales per cluster, and the standard deviation  $\omega$  of the Gaussian density. As it is difficult to evaluate the components of the score function and observed information corresponding to the parameter  $\omega$ , Waagepetersen & Schweder (2006) compute the profile log likelihood function  $l_p(\omega) = \max_{(\kappa, \alpha)} \log L(\theta)$  for a finite set of values  $\omega_l$ . This is done using (46) 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 Fig. 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/km<sup>2</sup> 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 9.

#### 7.4. Bayesian inference

To compute posterior distributions for  $\theta$  in a fully Bayesian approach to inference, we need to know the likelihood function for all values of  $\theta$ . For a Gibbs point process, the computational problems which arise because of the need to evaluate the unknown normalizing constant are therefore even harder than for finding the MLE (section 7.2) or the maximum  $a$

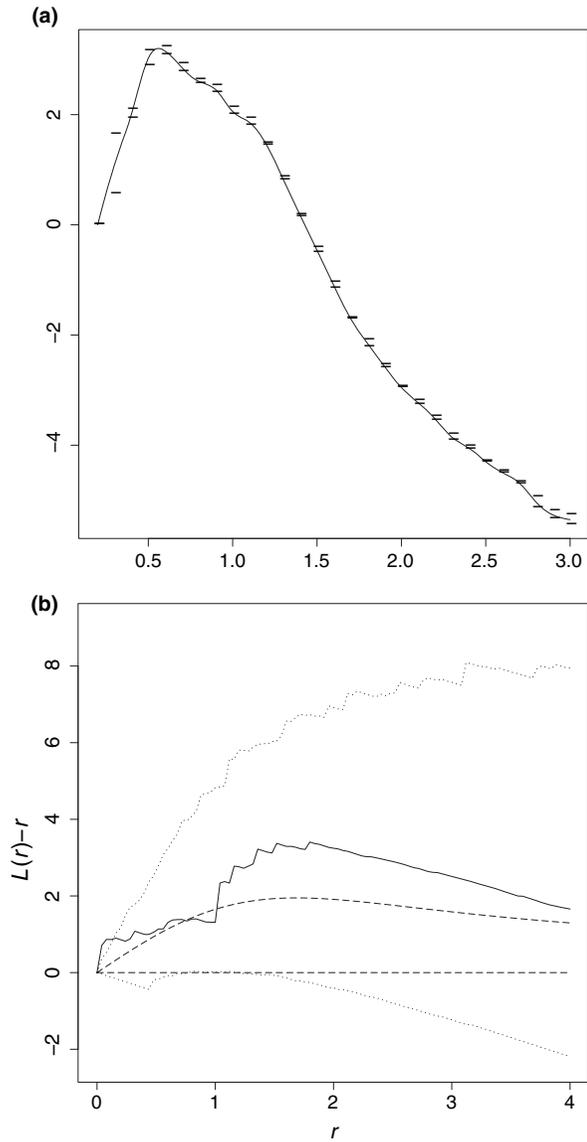


Fig. 12. Fitting a shot noise Cox process model to the North Atlantic whales data set. Upper plot: profile log likelihood function  $l_p(\omega) = \max_{(\kappa, \alpha)} \log L(\theta)$  obtained by cumulating estimated log likelihood ratios, see text. The small horizontal bars indicate 95% Monte Carlo confidence intervals for the log likelihood ratios. Lower plot: 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).

posteriori estimate (Heikkinen & Penttinen, 1999). Based on perfect simulation (section 9.3) and auxiliary variable MCMC methods (Møller *et al.*, 2006), progress on Bayesian inference for Markov point processes has been made in Berthelsen & Møller (2003, 2004, 2006, 2007).

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  $\theta$ . This simplifies computations, since the likelihood of the data given  $\theta$  and the random intensity function is just a Poisson likelihood function. References on Bayesian inference for Cox processes include Heikkinen & Arjas (1998), Wolpert & Ickstadt (1998), Best, Ickstadt & Wolpert (2000), Cressie & Lawson (2000), Møller & Waagepetersen (2003a, 2003b), Beněš *et al.* (2005) and Waagepetersen & Schweder (2006).

*Example 9 (Bayesian inference for North Atlantic whales).* In Waagepetersen & Schweder (2006), the unknown parameters  $\kappa$ ,  $\alpha$  and  $\omega$  (examples 2 and 8) are assumed to be *a priori* independent with uniform priors on bounded intervals for  $\kappa$  and  $\omega$  and an informative  $N(2, 1)$  (truncated at zero) prior for  $\alpha$  (the whales are *a priori* believed to appear in small groups of one to three 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  $\kappa$ ,  $\alpha$  and  $\omega$ . The posterior means for  $\kappa$ ,  $\alpha$  and  $\omega$  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 8) and the 95% central posterior interval [0.04, 0.08] for the whale intensity.

*Example 10 (Bayesian inference for tropical rainforest trees).* Considering the log Gaussian Cox process model for the tropical rainforest trees (example 1), we assume that  $\beta = (\beta_1, \beta_2, \beta_3)$ ,  $\sigma$ , and  $\alpha$  are *a priori* independent, and use an improper uniform prior for  $\beta$  on  $\mathbb{R}^3$ , an improper uniform prior for  $\sigma$  on  $[0.001, \infty)$ , and a uniform prior for  $\log \alpha$  with  $1 \leq \alpha \leq 235$ . For a discussion of posterior propriety in similar models, see Christensen *et al.* (2000). The Gaussian process is discretized to a  $200 \times 100$  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  $\beta$ ,  $\log \sigma$ , 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  $\beta_2$ , 8.76 and [6.03, 11.37] for  $\beta_3$ , 1.61 and [1.44, 1.85] for  $\sigma$ , 42.5 and [32.1, 56.45] for  $\alpha$ . Figure 13 shows the posterior means of the systematic part  $\beta_1 + \beta_2 z_2(u) + \beta_3 z_3(u)$  (upper plot) and the random part  $\Psi(u)$  (lower 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. Fig. 3. The fluctuations of the random part may be caused by small scale clustering due to seed dispersal and covariates concerning soil properties. The fluctuation may also be due to between-species competition.

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  $\rho_\theta(u) = \exp\left(z(u)\beta^\top + \sigma^2/2\right)$  for  $\mathbf{X}$  given  $\theta$ . 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 accommodate clustering for small values of  $r$  less than 10 m.

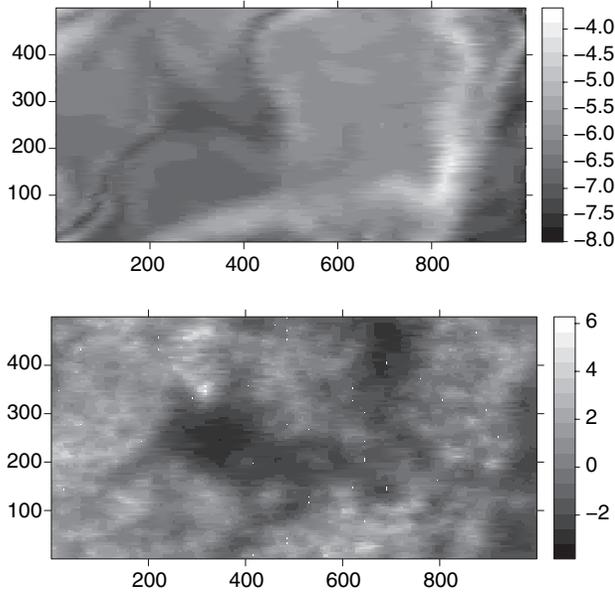


Fig. 13. Posterior mean of  $\beta_1 + \beta_2 z_2(u) + \beta_3 z_3(u)$  (upper) and  $\Psi(u)$  (lower),  $u \in W$ , under the log Gaussian Cox process model for the tropical rainforest trees.

**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. Other approaches for obtaining estimating equations for spatial point process models are studied in Takacs (1986) and Baddeley (2000).

In section 8.1, estimating functions based on the (conditional) intensity function are motivated heuristically as limits of composite likelihood functions (Lindsay, 1988) for Bernoulli 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.

*8.1. Estimating functions based on intensities*

For a given parametric model with parameter  $\theta$ , suppose that the intensity function  $\rho_\theta$  is expressible in 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  $p_i(\theta) = P_\theta(N_i = 1)$ . 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

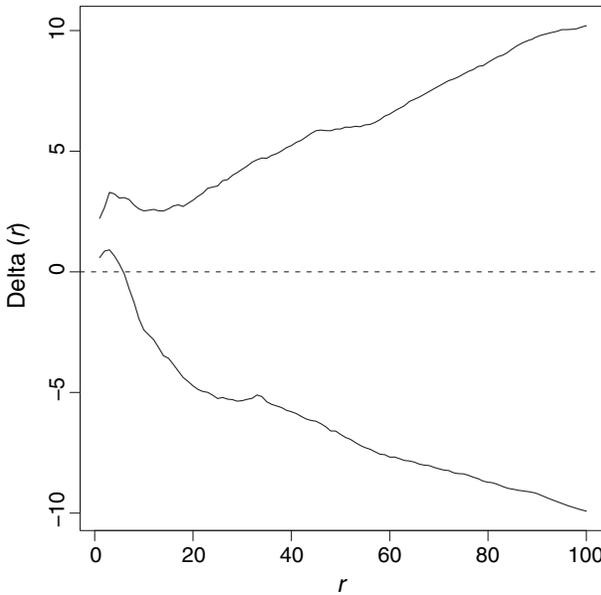


Fig. 14. Tropical rainforest trees: 95% central envelopes obtained from posterior predictive simulations of  $\Delta(r)$ .

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

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

$$\psi_1(\theta) = \sum_{u \in \mathbf{x}} d \log \rho_{\theta}(u) / d\theta - \int_W (d \log \rho_{\theta}(u) / d\theta) \rho_{\theta}(u) \, du. \tag{47}$$

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  $\rho_{\theta}$  is on a log linear form. For Cox processes, as exemplified in example 11 below, the solution may only provide an estimate of one component of  $\theta$ , while the other component may be estimated by another method. Asymptotic properties of estimates obtained using (47) are derived in Schoenberg (2004) and Waagepetersen (2007).

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

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

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

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

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

*Example 11 (Estimation of the intensity function for tropical rainforest trees).* For both the log Gaussian Cox process model in example 1 and the inhomogeneous Thomas process model in example 3, the intensity function is of the form  $\exp(z(u)(\hat{\beta}_1, \beta_2, \beta_3)^\top)$ , where  $\hat{\beta}_1 = \sigma^2/2 + \beta_1$  for the log Gaussian Cox process and  $\hat{\beta}_1 = \log(\kappa\alpha)$  for the inhomogeneous Thomas process. Using the estimating function (47) and `spatstat`, we obtain  $(\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3) = (-4.989, 0.021, 5.842)$ , where  $\hat{\beta}_2$  and  $\hat{\beta}_3$  are smaller than the posterior means obtained with the Bayesian approach in example 10. The estimate of course coincides with the MLE under the Poisson process with the same intensity function. Estimates of the clustering parameters, i.e.  $(\sigma^2, \alpha)$  respectively  $(\kappa, \omega)$ , may be obtained using minimum contrast estimation, see example 13.

Assuming  $(\hat{\beta}_2, \hat{\beta}_3)$  is asymptotically normal (Waagepetersen, 2007), we obtain approximate 95% confidence intervals  $[-0.018, 0.061]$  and  $[0.885, 10.797]$  for  $\beta_2$  and  $\beta_3$ , respectively. Under the Poisson process model much more narrow approximate 95% confidence intervals  $[0.017, 0.026]$  and  $[5.340, 6.342]$  are obtained.

*Example 12 (Maximum pseudo-likelihood estimation for ants' nests).* For the hierarchical model in example 5, we first correct for edge effects by conditioning on the data in  $W \setminus W_{\infty 45}$ . Using `spatstat`, the maximum pseudo-likelihood estimate (MPLE) of  $(\beta_M, \psi_M)$  is  $(-8.21, -0.09)$ , indicating (weak) repulsion between the *Messor* ants' nests. Without edge correction, a rather similar MPLE  $(-8.22, -0.12)$  is obtained. The edge corrected MPLE of  $(\beta_C, \psi_{CM}, \psi_C)$  is  $(-9.51, 0.13, -0.66)$ , indicating a positive association between the two species and repulsion within the *Cataglyphis* nests. If no edge correction is used, the MPLE for  $(\beta_C, \psi_{CM}, \psi_C)$  is  $(-9.39, 0.04, -0.30)$ . As mentioned in example 7, Högmänder & 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 Högmänder & Särkkä's model.

The differences between the MLE in example 7 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 (see section titled 'Log Gaussian Cox processes') 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 discrepancy 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 (47). This is illustrated in example 13 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 processes in Heinrich (1992) and Guan & Sherman (2007).

Alternatively, we may consider an estimating function based on the second order product density  $\rho_\theta^{(2)}(u, v)$  (Waagepetersen, 2007):

$$\psi_2(\theta) = \sum_{u, v \in X}^{\neq} d \log \rho_\theta^{(2)}(u, v) / d\theta - \int_{W^2} \left( d \log \rho_\theta^{(2)}(u, v) / d\theta \right) \rho_\theta^{(2)}(u, v) du dv. \tag{48}$$

This is the score of a limit of composite log likelihood functions based on Bernoulli 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 (48) typically must be evaluated using numerical integration. In the stationary case, Guan (2006) considers a related unbiased estimating function, where the integral in (48) is replaced by the number of pairs of distinct points times  $\log \int_{W^2} \rho_\theta^{(2)}(u, v) du dv$ .

*Example 13 (Minimum contrast estimation of clustering parameters for tropical rainforest trees).*

The solid curve in Fig. 9 shows an estimate of the  $K$ -function for the tropical rainforest trees obtained using (41) with  $\hat{\rho}$  given by the estimated parametric intensity function from example 11. For the inhomogeneous Thomas process, a minimum contrast estimate  $(\hat{\kappa}, \hat{\omega}) = (8 \times 10^{-5}, 20)$  is obtained by minimizing

$$\int_0^{100} (\hat{K}(r)^{1/4} - K(r; \kappa, \omega))^{1/4} dr \tag{49}$$

where

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

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

$$K(r; \sigma, \alpha) = 2\pi \int_0^r s \exp(\sigma^2 \exp(-s/\alpha)) ds$$

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 Fig. 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 (49). Typically, these parameters are chosen on an *ad-hoc* basis.

*Example 14 (Simultaneous estimation of parameters for tropical rainforest trees).*

To estimate the parameters  $(\tilde{\beta}_1, \tilde{\beta}_2, \tilde{\beta}_3)$  and  $(\kappa, \omega)$  for the inhomogeneous Thomas process (see example 11) simultaneously, we apply the estimating function  $\psi_2$  (48). We solve  $\psi_2(\theta) = 0$  using a grid search for  $\omega$  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  $\omega$  within a finite set of  $\omega$ -values. The resulting estimates of  $(\tilde{\beta}_1, \tilde{\beta}_2, \tilde{\beta}_3)$  and  $(\kappa, \omega)$  are respectively  $(-5.001, 0.021, 5.735)$  and  $(7 \times 10^{-5}, 30)$ . The estimate of  $\omega$  differs considerably from the minimum contrast estimate in example 13, while the remaining estimates are quite similar to those obtained previously for the inhomogeneous Thomas process in examples 11 and 13. The numerical computation of  $\psi_2$  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 often 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 often needed, see e.g. Fig. 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), Lantuejoul (2002), and Møller & Waagepetersen (2003b). For a shot noise Cox process, edge effects may occur since the Poisson process  $\Phi$  in (10) may be infinite or extend beyond  $S$ , 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 (Ripley, 1977, 1979) 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).

Without conditioning on  $n(\mathbf{X})$ , the standard algorithms 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,  $\lambda$  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, 1976; 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 *et al.*, 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 particularly 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  $\phi$  is a deterministic function and the  $R_t$  are i.i.d. random variables. The updating function  $\phi$  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  $n=1, 2, \dots$  by  $U_{t+1}^n = \phi(U_t^n, R_t)$  and  $L_{t+1}^n = \phi(L_t^n, R_t)$ ,  $t = T_n, T_n + 1, \dots, -1$ , 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  $n=1, 2, \dots$ . Note that the  $R_t$  are re-used for all  $n=1, 2, \dots$ . 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  $L_t^n \prec L_t^{n+1} \prec U_t^{n+1} \prec U_t^n$ ,  $t = T_n, \dots, 0$ . 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 *et al.*, 1999; Møller & Waagepetersen, 2003b, 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 (2003b, 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 *et al.*, 2006).

Many scientific problems call for new spatial point process methodology for analyzing complex and large data sets (often with marks and possibly in space-time). For example, in the tropical rainforest example, the data for the *Beilschmiedia* trees are just a very small part of a very large data set containing positions and diameters at breast height for around 300 species recorded over several instances in time, and between-species competition (which is ignored in our analysis) should of course be taken into account (see, e.g. Illian, Møller & Waagepetersen, 2007). Another example is the Sloan Digital Sky Survey with millions of galaxies, where it is of interest to model the clustering of galaxies. One question of practical importance is, in which cases might the quick non-likelihood approaches be sufficient and how to choose between them.

### 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) does 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.*, 2007b).

Often we need to specify an inhomogeneous point process model. For Poisson and Cox processes, log linear modelling as in examples 1 and 3 can be a useful approach, but it may not always be scientifically reasonable. Furthermore, how much of the inhomogeneity should be explained by varying intensity (first order effect) and how much by clustering (second order property)? The same problem of unidentifiability is encountered in geostatistics (trend versus spatial correlation).

For Gibbs point processes, as exemplified by the model for the *Cataglyphis* ants' nests in example 5, a simple way of modelling inhomogeneity is to introduce a non-constant first-order potential (Ogata & Tanemura, 1986; Stoyan & Stoyan, 1998; Berthelsen & Møller, 2007). Another possibility is to consider an independent, but possibly inhomogeneous, 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, Van Lieshout, Jensen & Nielsen (2003), and Nielsen & Jensen (2004). It is an open problem to extend asymptotic results for MLE (and to some extent 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, as the likelihood function depends on the covariate at any location

in the observation window. In the tropical rainforest 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. For the case of Poisson processes, the missing covariate problem is considered in Rathbun (1996) and Rathbun, Shiffman & Gwaltney (2007).

### 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. In spatial statistics, Ripley (1977) noted that Markov point processes can be considered as the equilibrium state of a reversible spatial birth–death process (section 9.2). We question their popularity in spatial statistics for the following reasons.

Indeed, Markov point processes provide a flexible framework for modelling repulsive spatial interaction as exemplified by the overlap model for the Norwegian spruces. However, although 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.

The estimation of the interaction range  $R$  is a tricky issue which seems to require computation of a profile likelihood (or 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, 2006, 2007).

There is a lack of satisfactory Gibbs point processes modelling attractive spatial interaction. As discussed in example 5, Strauss point processes with a hard core are not so flexible due to a kind of phase transition behaviour; see also Gates & Westcott (1986) and Møller (1999). 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 and Illian, Møller & Waagepetersen (2007).

As mentioned several times, closed form expressions for spatial point process densities are rare. Shirai & Takahashi (2003) and McCullagh & Møller (2006) study a large model class of non-Poisson point process models, called the permanental and determinantal processes in McCullagh & Møller (2006), where both the density of the process and the product densities are of an analytic form. The processes possess many other appealing properties, and the permanental process models aggregation of points (in some cases it is a Cox process), while the determinantal 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. Because of 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, 2006, 2007). Also the techniques in Møller & Mengersen (2007a, 2007b) 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 determinantal processes mentioned in section 10.3 have not yet 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 user-friendly 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

Because of space constraints, apart from sections 9 and 10.3, we have not considered space-time point processes. One challenge is to develop tractable and yet interesting continuous-time models, when spatial point process data are 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), and the comparison of discrete and continuous time point processes in Rasmussen *et al.* (2007). For space-time point processes in general, we refer to Daley & Vere-Jones (2003), Møller & Waagepetersen (2003b, section 2.4), Diggle (2005), Lawson (2006), and the references therein.

### 10.6. Conclusion

In conclusion, spatial point processes and their applications 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 the least, the challenges of developing new tools.

### Acknowledgements

We are grateful to Adrian Baddeley for assistance with `spatstat`, and in particular to Peter Diggle for many detailed comments to an earlier manuscript. We also acknowledge helpful comments by an anonymous associate editor, an anonymous referee, Yongtao Guan, Andrew Lawson, Antti Penttinen, and Eva Vedel Jensen. Supported by the Danish Natural Science Research Council grant 272-06-0442 'Point process modelling and statistical inference'.

## References

- Baddeley, A. & Møller, J. (1989). Nearest-neighbour Markov point processes and random sets. *Int. Stat. Rev.* **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. *Aust. NZ J. Stat.* **42**, 283–322.
- Baddeley, A. & Turner, R. (2005). Spatstat: an R package for analyzing spatial point patterns. *J. Stat. Softw.* **12**, 1–42. Available at <http://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 & D. Stoyan), 23–74, Springer Lecture Notes in Statistics 185, Springer-Verlag, New York.
- Baddeley, A., Møller, J. & Waagepetersen, R. (2000). Non- and semi-parametric estimation of interaction in inhomogeneous point patterns. *Statist. Neerlandica* **54**, 329–350.
- Baddeley, A., Turner, R., Møller, J. & Hazelton, M. (2005). Residual analysis for spatial point processes (with discussion). *J. Roy. Statist. Soc. Ser. B* **67**, 617–666.
- Baddeley, A., Gregori, P., Mateu, J., Stoica, R. & Stoyan, D., eds. (2006). *Case studies in spatial point process modeling*, Springer Lecture Notes in Statistics 185, Springer-Verlag, New York.
- Baddeley, A., Møller, J. & Pakes, A. G. (2007a). Properties of residuals for spatial point processes. *Ann. Inst. Statist. Math.* (to appear).
- Baddeley, A., Møller, J. & Waagepetersen, R. (2007b). Residual versions of summary statistics for spatial point processes (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. *Ann. Inst. Statist. Math.* **47**, 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 & D. Stoyan), 153–167, Springer Lecture Notes in Statistics 185, Springer-Verlag, New York.
- 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 Anal. Stereol.* **24**, 159–168.
- Berman, M. & Turner, R. (1992). Approximating point process likelihoods with GLIM. *Appl. Stat.* **41**, 31–38.
- Berthelsen, K. K. & Møller, J. (2002). A primer on perfect simulation for spatial point processes. *Bull. Braz. Math. Soc.* **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. *Scand. J. Stat.* **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. Proceedings of the International Conference on Spatial Point Process Modelling and its Applications* (eds A. Baddeley, P. Gregori, J. Mateu, R. Stoica & D. Stoyan), Colección Trabajos de Informática y Tecnología, Num 20. Editorial Universitat Jaume I, Castellón, Spain.
- Berthelsen, K. K. & Møller, J. (2006). Bayesian analysis of Markov point processes. In *Case studies in spatial point process modeling* (eds A. Baddeley, P. Gregori, J. Mateu, R. Stoica & D. Stoyan), 85–97, Springer Lecture Notes in Statistics 185, Springer-Verlag, New York.
- Berthelsen, K. K. & Møller, J. (2007). *Non-parametric Bayesian inference for inhomogeneous Markov point processes*. Technical Report R-2007-09, Department of Mathematical Sciences, Aalborg University. Available at <http://www.math.aau.dk/~jm>.
- Besag, J. (1977a). Some methods of statistical analysis for spatial data. *Bull. Inst. Stat. Inst.* **47**, 77–92.
- Besag, J. E. (1977b). Discussion of the paper by Ripley (1977). *J. Roy. Stat. Soc. Ser. 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. *J. Amer. Stat. Assoc.* **95**, 1076–1088.
- Blackwell, P. G. & Møller, J. (2003). Bayesian analysis of deformed tessellation models. *Adv. Appl. Probab.* **35**, 4–26.
- Boots, B., Okabe, A. & Thomas, R., eds. (2003). *Modelling geographical systems: statistical and computational applications*, Kluwer Academic Publishers, Dordrecht.
- Brix, A. & Chadoeuf, J. (2002). Spatio-temporal modeling of weeds and shot-noise G Cox processes. *Biometrical J.* **44**, 83–99.

- Brix, A. & Diggle, P. J. (2001). Spatio-temporal prediction for log-Gaussian Cox processes. *J. Roy. Stat. Soc. Ser. B* **63**, 823–841.
- Brix, A. & Kendall, W. S. (2002). Simulation of cluster point processes without edge effects. *Adv. Appl. Probab.* **34**, 267–280.
- Brix, A. & Møller, J. (2001). Space-time multitype log Gaussian Cox processes with a view to modelling weed data. *Scand. J. Stat.* **28**, 471–488.
- Buckland, S. T., Anderson, D. R., Burnham, K. P., Laake, J. L., Borchers, D. L. & Thomas, L. (2004). *Advanced Distance Sampling*, Oxford University Press.
- 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*. Tech. rep. 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. *J. Trop. Ecol.* **12**, 231–256.
- Cox, D. R. (1955). Some statistical models related with series of events. *J. Roy. Stat. Soc. Ser. B* **17**, 129–164.
- Cox, D. R. (1972). The statistical analysis of dependencies in point processes. In *Stochastic point processes* (ed P. A. W. Lewis), 55–66, Wiley, New York.
- Cressie, N. & Lawson, A. (2000). Hierarchical probability models and Bayesian analysis of minefield locations. *Adv. Appl. Probab.* **32**, 315–330.
- Daley, D. J. & Vere-Jones, D. (2003). *An introduction to the theory of point processes. Volume I: elementary theory and methods*, 2nd edn. Springer-Verlag, New York.
- 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. *Appl. Stat.* **34**, 138–147.
- Diggle, P. J. (2003). *Statistical analysis of spatial point patterns*, 2nd edn. Arnold, London.
- 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 Kybernetik* **20**, 270–278.
- Gates, D. J. & Westcott, M. (1986). Clustering estimates for spatial point distributions with unstable potentials. *Ann. Inst. Statist. Math.* **38**, 123–135.
- Gelman, A. & Meng, X.-L. (1998). Simulating normalizing constants: from importance sampling to bridge sampling to path sampling. *Stat. Sci.* **13**, 163–185.
- Gelman, A., Meng, X. L. & Stern, H. S. (1996). Posterior predictive assessment of model fitness via realized discrepancies (with discussion). *Stat. Sinica* **6**, 733–807.
- Georgii, H.-O. (1976). Canonical and grand canonical Gibbs states for continuum systems. *Commun. Math. Phys.* **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 & M. N. M. van Lieshout), 79–140, Chapman & Hall/CRC, Florida.
- Geyer, C. J. & Møller, J. (1994). Simulation procedures and likelihood inference for spatial point processes. *Scand. J. Stat.* **21**, 359–373.
- Geyer, C. J. & Thompson, E. A. (1995). Annealing Markov chain Monte Carlo with applications to pedigree analysis. *J. Amer. Stat. Assoc.* **90**, 909–920.
- Gouldard, M., Särkkä, A. & Grabarnik, P. (1996). Parameter estimation for marked Gibbs point processes through the maximum pseudo-likelihood method. *Scand. J. Stat.* **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. *J. Amer. Stat. Assoc.* **101**, 1502–1512.
- Guan, Y. & Sherman, M. (2007). On least squares fitting for stationary spatial point processes. *J. Roy. Statist. Soc. Ser. B* **69**, 31–49.
- Guan, Y., Sherman, M. & Calvin, J. A. (2006). Assessing isotropy for spatial point processes. *Biometrics* **62**, 119–125.
- 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. *Adv. Appl. Probab.* **35**, 319–336.
- Harkness, R. D. & Isham, V. (1983). A bivariate spatial point pattern of ants' nests. *Appl. Stat.* **32**, 293–303.
- Heikkinen, J. & Arjas, E. (1998). Non-parametric Bayesian estimation of a spatial Poisson intensity. *Scand. J. Stat.* **25**, 435–450.
- 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* (eds S. Kubik & J. A. Visek), 479–492, Academic Publishing House, Prague.
- 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 & A. C. Chadwick), 25–41, Blackwell Scientific Publications, Oxford.
- Illian, J. B., Møller, J. & Waagepetersen, R. P. (2007). Spatial point process analysis for a plant community with high biodiversity. *Environ. Ecol. Stat.*, conditionally accepted.
- Jensen, A. T. (2005). *Statistical inference for doubly stochastic Poisson processes*. Ph.D. 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. *Annal. Appl. Probab.* **3**, 445–461.
- Kendall, W. S. (1998). Perfect simulation for the area-interaction point process. In *Probability towards 2000* (eds L. Accardi & C. Heyde), 218–234, Springer Lecture Notes in Statistics 128, Springer Verlag, New York.
- Kendall, W. S. & Møller, J. (2000). Perfect simulation using dominating processes on ordered spaces, with application to locally stable point processes. *Adv. Appl. Probab.* **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 & D. Stoyan), 36–71, Lecture Notes in Physics, Springer, Berlin.
- Kingman, J. F. C. (1993). *Poisson processes*. Clarendon Press, Oxford.
- Lantuejoul, C. (2002). *Geostatistical simulation: models and algorithms*. Springer-Verlag, Berlin.
- Lawson, A. (2006). *Statistical methods in spatial epidemiology*, 2nd edn. Wiley, New York.
- van Lieshout, M. N. M. (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. *Statist. Neerlandica* **50**, 344–361.
- Lindsay, B. G. (1988). Composite likelihood methods. *Contemp. Math.* **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. *Adv. Appl. Probab.* **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. *Ann. Inst. Statist. Math.* **53**, 661–680.
- Matérn, B. (1971). Doubly stochastic Poisson processes in the plane. In *Statistical ecology* (eds G. P. Patil, E. C. Pielou & W. E. Waters), volume 1, 195–213, Pennsylvania State University Press, University Park.
- McCullagh, P. & Møller, J. (2006). The permanental process. *Adv. Appl. Probab.* **38**, 873–888.
- Møller, J. (1989). On the rate of convergence of spatial birth-and-death processes. *Ann. Inst. Statist. Math.* **3**, 565–581.
- Møller, J. (1994). Contribution to the discussion of N. L. Hjort and H. Omre (1994): Topics in spatial statistics. *Scand. J. Stat.* **21**, 346–349.
- Møller, J. (1999). Markov chain Monte Carlo and spatial point processes. In *Stochastic geometry: likelihood and computation* (eds O. E. Barndorff-Nielsen, W. S. Kendall & M. N. M. van Lieshout), 141–172, Monographs on Statistics and Applied Probability 80, Chapman and Hall/CRC, Florida.
- Møller, J. (2003). Shot noise Cox processes. *Adv. Appl. Probab.* **35**, 4–26.
- Møller, J. & Mengersen, K. (2007a). Ergodic averages for monotone functions using upper and lower dominating processes. In *Bayesian Statistics 8* (eds J. Bernardo, S. Bayarri, J. Berger, A. Dawid, D. Heckerman, A. Smith & M. West), Oxford University Press, Oxford.

- Møller, J. & Mengersen, K. (2007b). Ergodic averages via dominating processes. *Bayesian Anal.* (conditionally accepted).
- Møller, J. & Torrisi, G. L. (2005). Generalised shot noise Cox processes. *Adv. Appl. Probab.* **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), 143–198, Springer Lecture Notes in Statistics 173, Springer-Verlag, New York.
- Møller, J. & Waagepetersen, R. P. (2003b). *Statistical inference and simulation for spatial point processes*, Chapman and Hall/CRC, Florida.
- Møller, J., Syversveen, A. R. & Waagepetersen, R. P. (1998). Log Gaussian Cox processes. *Scand. J. Stat.* **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**, 451–458.
- Neyman, J. & Scott, E. L. (1958). Statistical approach to problems of cosmology. *J. Roy. Statist. Soc. Ser. 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. *Scand. J. Stat.* **31**, 131–142.
- Norman, G. E. & Filinov, V. S. (1969). Investigations of phase transition by a Monte-Carlo method. *High Temp.* **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 & F. C. Lam), 150–154, Elsevier, Amsterdam.
- Papangelou, F. (1974). The conditional intensity of general point processes and an application to line processes. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* **28**, 207–226.
- Penttinen, A., Stoyan, D. & Henttonen, H. M. (1992). Marked point processes in forest statistics. *Forest Sci.* **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 Struct. Algor.* **9**, 223–252.
- Rasmussen, J. G., Møller, J., Aukema, B. H., Raffa, K. F. & Zhu, J. (2007). Continuous time modelling of dynamical spatial lattice data observed at sparsely distributed times. *J. Roy. Statist. Soc. Ser. B* **69**, 701–713.
- 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. *Adv. Appl. Probab.* **26**, 122–154.
- Rathbun, S. L., Shiffman, S. & Gwaltney, C. J. (2007). Modelling the effects of partially observed covariates on Poisson process intensity. *Biometrika* **94**, 153–165.
- Ripley, B. D. (1976). The second-order analysis of stationary point processes. *J. Appl. Probab.* **13**, 255–266.
- Ripley, B. D. (1977). Modelling spatial patterns (with discussion). *J. Roy. Statist. Soc. Ser. B* **39**, 172–212.
- Ripley, B. D. (1979). Simulating spatial patterns: dependent samples from a multivariate density. Algorithm AS 137. *Appl. Stat.* **28**, 109–112.
- 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. *J. London Math. Soc.* **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.
- Schlادitz, K. & Baddeley, A. J. (2000). A third-order point process characteristic. *Scand. J. Stat.* **27**, 657–671.
- Schlatter, M., Riberio, P. & Diggle, P. (2004). Detecting dependence between marks and locations of marked point processes. *J. Roy. Statist. Soc. Ser. B* **66**, 79–93.
- Schlather, M. (1999). *Introduction to positive definite functions and unconditional simulation of random fields*. Tech. rep. ST 99-10, Lancaster University.
- Schlather, M. (2001). On the second-order characteristics of marked point processes. *Bernoulli* **7**, 99–117.

- Schoenberg, F. P. (2005). Consistent parametric estimation of the intensity of a spatial-temporal point process. *J. Stat. Plan. Infer.* **128**, 79–93.
- Shirai, T. & Takahashi, Y. (2003). Random point fields associated with certain Fredholm determinants I: Fermion, Poisson and Bose point processes. *J. Funct. Anal.* **205**, 414–463.
- Skare, Ø., Møller, J. & Jensen, E. B. V. (2007). Bayesian analysis of spatial point processes in the neighbourhood of Voronoi networks. *Statistics and Computing* **17** (to appear).
- 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. *Can. J. Fish. Aquat. Sci.* **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 J.* **40**, 521–531.
- Stoyan, D. & Stoyan, H. (2000). Improving ratio estimators of second order point process characteristics. *Scand. J. Stat.* **27**, 641–656.
- Stoyan, D. & Wälder, O. (2000). On variograms in point process statistics, II Models for markings and ecological interpretation. *Biometrical J.* **42**, 171–187.
- Stoyan, D., Kendall, W. S. & Mecke, J. (1995). *Stochastic geometry and its applications*, 2nd edn. Wiley, Chichester.
- Takacs, R. (1986). Estimator for the pair-potential of a Gibbsian point process. *Statistics* **17**, 429–433.
- Thomas, M. (1949). A generalization of Poisson's binomial limit for use in ecology. *Biometrika* **36**, 18–25.
- Waagepetersen, R. (2005). Discussion of the paper by Baddeley, Turner, Møller & Hazelton (2005). *J. Roy. Statist. Soc. Ser. B* **67**, 662.
- Waagepetersen, R. (2007). An estimating function approach to inference for inhomogeneous Neyman-Scott processes. *Biometrics* **63**, 252–258.
- Waagepetersen, R. & Schweder, T. (2006). Likelihood-based inference for clustered line transect data. *J. Agr. Biol. Envir. Stat.* **11**, 264–279.
- Widom, B. & Rowlinson, J. S. (1970). A new model for the study of liquid-vapor phase transitions. *J. Chem. Phys.* **52**, 1670–1684.
- Wolpert, R. L. & Ickstadt, K. (1998). Poisson/gamma random field models for spatial statistics. *Biometrika* **85**, 251–267.

*Received June 2006, in final form May 2007*

Jesper Møller, Department of Mathematical Sciences, Aalborg University.  
E-mail: jm@math.aau.dk