Geometric anisotropic spatial point pattern analysis and Cox processes

January 27, 2012

Jesper Møller
Department of Mathematical Sciences, Aalborg University

Håkon Toftaker
Department of Mathematical Sciences, Norwegian University of Science and Technology

Abstract

We consider spatial point processes with a pair correlation function $g(u)$ which depends only on the lag vector u between a pair of points. Our interest is in statistical models with a special kind of 'structured' anisotropy: g is geometric anisotropy if it is elliptical but not spherical. In particular we study Cox process models with an elliptical pair correlation function, including shot noise Cox processes and log Gaussian Cox processes, and we develop estimation procedures using summary statistics and Bayesian methods. Our methodology is illustrated on real and synthetic datasets of spatial point patterns.

Key words: Bayesian inference; K-function; log Gaussian Cox process; minimum contrast estimation; pair correlation function; second-order intensity-reweighted stationarity; shot noise Cox process; spectral density; Whittle-Matérn covariance function.

1 Introduction

For statistical analysis of spatial point patterns, considering an underlying spatial point process model, stationarity is often assumed—or at least second-order stationarity—meaning that the spatial point process has a constant intensity function ρ and a pair correlation function $g(u)$ which depends only on the lag vector u between pairs of points (definitions of ρ and g are given in Section 2.1). Moreover, due to simpler interpretation and ease of analysis, $g(u) = g_0(\|u\|)$ is often assumed to be isotropic. However, these assumptions are known not to be satisfied in many applications, and failure to account for spatial and directional inhomogeneity can result in erroneous inferences.

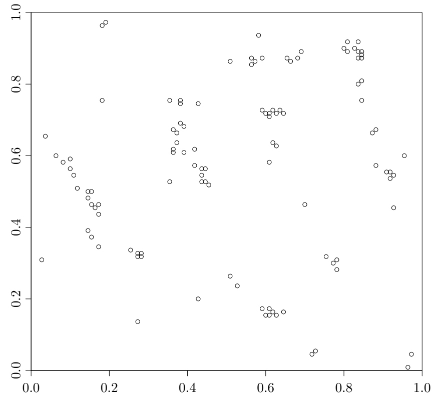


Figure 1: Locations of 110 chapels observed in a square region (rescaled to the unit square).

For instance, the distribution of plant seedling locations often exhibits spatial inhomogeneity due to different ground conditions and anisotropy due to factors such as prevailing wind direction. This paper considers two other examples. Figure 1 shows the locations of 110 chapels in a square region near Ebbw Vale in Wales. These data were analyzed in Mugglestone & Renshaw (1996) assuming that ρ is constant and using a spectral analysis for $\rho\delta_0+\rho^2(g-1)$ (Bartlett (1964) called this the complete covariance density function; here δ_0 denotes Dirac's delta function). The chapels clearly exhibit directional features at the larger scale, caused by four more or less parallel valleys, and clustering at a smaller scale. Furthermore, Figure 2 shows the epicentral locations of 6922 earthquakes registered in a rectangular area around Los Angeles, California, in the period from 1st January 1984 to 17th June 2004; these data were analyzed in Veen & Schoenberg (2006). There seems to be both inhomogeneity and a prevailing direction in this point pattern. Veen & Schoenberg (2006) estimated ρ by a mixture of a non-parametric anisotropic normal kernel estimator and the usual non-parametric estimator for a homogeneous Poisson process (i.e., the number of observed points divided by the area of the observation window), where different choices of the mixing probability were investigated by plots of the estimated inhomogeneous K-function introduced in Baddeley, Møller & Waagepetersen (2000). Since $K(r) = \int_{\|u\|\leq r} g(u)\,\mathrm{d}u$ for $r \geq 0$, this function is not informative about a possible anisotropy of g.

To the best of our knowledge, in the literature on anisotropic spatial point processes, second-order stationarity is always assumed. This literature can be summarized as follows. Ohser & Stoyan (1981), Stoyan & Stoyan (1994), and Illian, Penttinen, Stoyan & Stoyan (2008) discussed functional summary statistics for detecting anisotropy in spatial point patterns. Mugglestone &

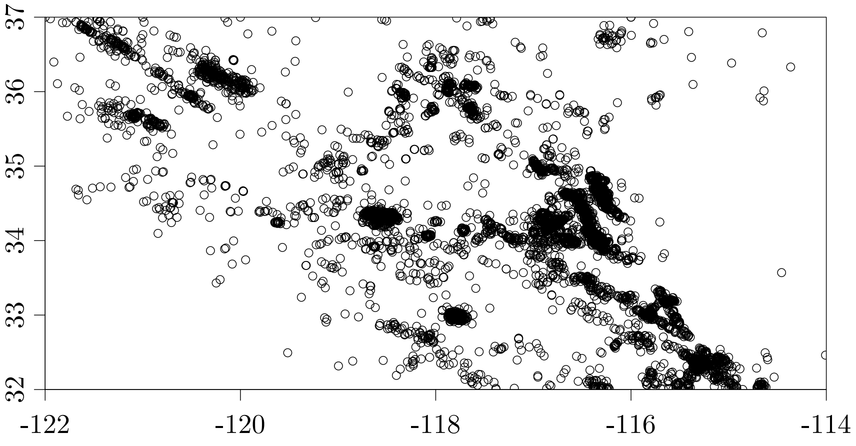


Figure 2: Registered earthquakes in a part of California during the period 1st January 1984 to 17th June 2004. The observation window is a rectangle between latitudes 32° and 37° and longitudes $-122°$ and $-114°$.

Renshaw (1996) and other papers by these authors consider a spectral analysis for the second-order characteristics of an anisotropic spatial point process (Bartlett 1964), while Rosenberg (2004) used an alternative approach based on wavelets. Castelloe (1998) considered a Bayesian approach for an anisotropic Poisson cluster process obtained by extending the well-known Thomas process (Thomas 1949, Cox & Isham 1980) with an isotropic bivariate normal offspring distribution to the case of a general bivariate normal offspring distribution. Guan, Sherman & Calvin (2006) proposed a formal approach to test for isotropy based on the asymptotic joint normality of the sample second-order intensity function.

The focus in this paper is on geometric anisotropy, meaning that g is elliptical but not spherical. This property is useful for descriptive purposes of anisotropy, in particular for Cox processes (Cox 1955) as shown later on. Our paper relates to but substantially extends the work by Castelloe (1998). Throughout the paper we assume second-order intensity-reweighted stationarity (Baddeley et al. 2000), i.e. when second-order stationarity is weakened to the case where ρ is not necessarily constant. Thereby it is possible to model spatial inhomogeneity. For Cox processes driven by a random intensity function λ, second-order intensity-reweighted stationarity is satisfied when $\lambda = \rho S$, where S is a so-called residual process with unit mean and a stationary covariance function c. Then $g = c + 1$, and so anisotropy of g corresponds to anisotropy of the covariance function of the residual process.

The remainder of this paper is organized as follows. Section 2 provides some background material on spatial point processes and discusses the meaning of geometric anisotropy. Section 3 considers Cox processes as mentioned above and studies in particular the meaning of geometric anisotropy for log Gaussian Cox processes (LGCPs) and shot noise Cox processes (SNCPs). Section 4 deals with various informal but simple and fast estimation procedures for such Cox processes based on ρ, g, the K-function (Ripley 1976, Baddeley et al. 2000), the spectral density (Mugglestone & Renshaw 1996), and other functional summary statistics closely related to g. Section 5 discusses simulation-based Bayesian inference in relation to SNCPs. Furthermore, in Sections 4–5, the usefulness of the methods is investigated using synthetic datasets for LGCPs and SNCPs. Finally, Section 6 illustrates how our methodology applies for analyzing the real datasets in Figures 1–2.

2 Geometric anisotropy

2.1 Assumptions

Throughout this paper we consider the following general setting of a spatial point process X which is second-order reweighted stationary and has an elliptical pair correlation function. For specificity and simplicity, we view X as a random locally finite subset of $\mathbb{R}^2$, but most ideas and results in this paper extend to spatial point processes defined on $\mathbb{R}^d$, $d = 2, 3, \ldots$ (for $d = 1$, any pair correlation function is isotropic, so this case is not of our interest). For background material on spatial point processes, including measure theoretical details, see Møller & Waagepetersen (2004) and the references therein.

We assume that X is second-order intensity-reweighted stationary (Baddeley et al. 2000). This means that X has an intensity function $\rho : \mathbb{R}^2 \mapsto [0, \infty)$ and a pair correlation function $g : \mathbb{R}^d \mapsto [0, \infty)(u)$ such that for any Borel functions $h_1 : \mathbb{R}^2 \mapsto [0, \infty)$ and $h_2 : \mathbb{R}^2 \times \mathbb{R}^2 \mapsto [0, \infty)$,

$$\mathrm{E} \sum_{u \in X} h_1(u) = \int h_1(u)\rho(u)\,\mathrm{d}u \tag{1}$$

and

$$\mathrm{E} \sum_{u,v \in X}^{\neq} h_2(u, v) = \int h_2(u, v)\rho(u)\rho(v)g(u - v)\,\mathrm{d}u\,\mathrm{d}v \tag{2}$$

where the integrals are finite if h_1 and h_2 are bounded and have compact support, and where $\neq$ over the summation sign means that $u \neq v$. Note that $g = 1$ if X is a Poisson process.

We also assume that g is elliptical, i.e.

$$g(u) = g_0\left(\sqrt{u\Sigma^{-1}u^t}\right), \quad u \in \mathbb{R}^2, \tag{3}$$

where u is treated as a row vector with transpose u^t, Σ is a 2×2 symmetric positive definite matrix with determinant $|\Sigma| > 0$, and $g_0 : \mathbb{R} \mapsto [0, \infty)$ is a Borel function such that g becomes locally integrable, i.e. $\int_0^r sg(s)\,\mathrm{d}s < \infty$ whenever $0 < r < \infty$. This may or may not be a reasonable assumption in applications, and it has the following interpretation. Recall that for any $u \in \mathbb{R}^2$, $\rho_u(v) = \rho(u)g(v-u)$ is the intensity function of the conditional distribution of $X \setminus \{u\}$ given that $u \in X$ (formally, this conditional distribution is the reduced Palm distribution of X at the event u, see e.g. Stoyan, Kendall & Mecke (1995)). So if $\rho(u) > 0$ for some $u \in \mathbb{R}^2$ (which is the case if X is non-empty with a positive probability), then g is elliptical if and only if ρ_u is elliptical.

We say that g is *isotropic* if it is spherical, and *geometric anisotropic* if it is elliptical but not spherical.

2.2 Geometry

For later use we notice the following properties of the elliptical pair correlation function (3).

By the spectral theorem, Σ is of the form

$$\Sigma = \omega^2 U_\theta \mathrm{diag}\left(1, \zeta^2\right) U_\theta^t, \quad 0 \le \theta < \pi,\ \omega > 0,\ 0 < \zeta \le 1,$$

where U_θ^t is the orthonormal matrix with rows $(\cos\theta, \sin\theta)$ and $(-\sin\theta, \cos\theta)$. The ellipse $E = \{u : u\Sigma^{-1}u^t = 1\}$ has semi-major axis ω corresponding to the angle θ, and semi-minor axis $\omega\zeta$ corresponding to the angle $\theta + \pi/2$. Thus ζ is the ratio of the minor axis and the major axis; we call it the *anisotropy factor*. Note the periodicity: Σ is unchanged if we replace θ by $\theta + \pi$.

In the isotropic case of g, we have that $\zeta = 1$, $\Sigma = \omega^2 I$, E is a circle of radius ω, and the value of θ plays no role. Furthermore, Ripley's K-function (Ripley 1976) or the inhomogeneous K-function (Baddeley et al. 2000) is then given by

$$K(r) = 2\pi \int_0^r s g_0(s/\omega)\,\mathrm{d}s, \quad r \ge 0. \tag{4}$$

Define the 'square roots'

$$\Sigma^{1/2} = \omega U_\theta \mathrm{diag}\left(1, \zeta\right), \quad \Sigma^{-1/2} = U_\theta \mathrm{diag}\left(1, 1/\zeta\right)/\omega.$$

Thereby $\Sigma = \Sigma^{1/2}(\Sigma^{1/2})^t$, $\Sigma^{-1} = \Sigma^{-1/2}(\Sigma^{-1/2})^t$, $(\Sigma^{1/2})^{-1} = (\Sigma^{-1/2})^t$, and $g(u) = g_0\left(\|u\Sigma^{-1/2}\|\right)$. Thus g can be seen as an isotropic pair correlation function in a new system of coordinates obtained by a rotation of θ radians followed by rescaling the abscissa of the original vector coordinates by ζ.

Define, in terms of polar coordinates (r, ϕ), the pair correlation function $g_1(r, \phi) = g(r\cos\phi, r\sin\phi)$, i.e.

$$g_1(r, \phi) = g_0\left(\frac{r}{\omega}\sqrt{\frac{1 - (1-\zeta^2)\cos^2(\phi - \theta)}{\zeta^2}}\right), \quad r > 0,\ 0 \le \phi < 2\pi. \tag{5}$$

For any fixed $r > 0$, the function $g_1(r, \cdot)$ is periodic with period π. We shall later consider cases where g_0 is strictly decreasing. Then $g_1(r, \cdot)$ has global maximum points at $\phi = \theta$ and $\phi = \theta + \pi$, global minimum points at $\phi = \theta + \pi/2$ and $\phi = \theta + 3\pi/2$ if $0 \le \theta \le \pi/2$, and global minimum points at $\phi = \theta - \pi/2$ and $\phi = \theta + \pi/2$ if $\pi/2 \le \theta < \pi$.

2.3 Spectral analysis

Mugglestone & Renshaw (1996) demonstrated how Bartlett's spectral density (Bartlett 1964) can be used to analyze both clustering properties and directionality in point patterns. For a formal mathematical treatment of Bartlett's spectral density, see Daley & Vere-Jones (2003).

Briefly, following Mugglestone & Renshaw (1996), we assume X to be second-order stationary (so $\rho(u) = \rho$ is constant), and its spectral density is defined by

$$\gamma(u) = \rho + \rho^2 \tilde{c}(u), \quad u \in \mathbb{R}^2, \tag{6}$$

where $\tilde{c}$ is the two-dimensional inverse Fourier transform of $c = g - 1$ (provided this transform exists), i.e.

$$\tilde{c}(u) = \int c(v) \exp\left(-\mathrm{i} u v^t\right) \, \mathrm{d}v, \quad u \in \mathbb{R}^2, \tag{7}$$

where $\mathrm{i} = \sqrt{-1}$ (as in Mugglestone & Renshaw (1996) the factor $(2\pi)^{-1}$, which appears in Bartlett (1964), is omitted in (7); thereby, since c is symmetric, $\tilde{c}$ is the same as the Fourier transform of c). The key ingredient in (6) is the term $\tilde{c}(u)$, which is also well-defined in the more general setting of second-order intensity-reweighted spatial point processes.

Defining $c_0 = g_0 - 1$ and $c_\Sigma = c$, as g is elliptical,

$$c_\Sigma(u) = c_0\left(\sqrt{u \Sigma^{-1} u^t}\right)$$

is elliptical too. Note that $c_I(u) = c_0(\|u\|)$ is isotropic, and

$$\tilde{c}_I(u) = 2\pi \int_0^\infty r c_0(r) J_0(r\|u\|) \, \mathrm{d}r \tag{8}$$

is essentially the Hankel transform of order zero (here J_0 is the Bessel function of the first kind, of order zero). Therefore,

$$\gamma(u) = \rho + \rho^2 \tilde{c}_\Sigma(u) = \rho + \rho^2 |\Sigma|^{1/2} \tilde{c}_0\left(\sqrt{u \Sigma u^t}\right) \tag{9}$$

is elliptical, where $\tilde{c}_0(r) = \tilde{c}_I(u)$ for $\|u\| = r$. It follows that the elliptical contours of γ are similar to those of c (or g) except for a scaling factor and a rotation of $\pi/2$ radians.

Finally, for later use, define the following functions. By (9), the spectral density $\gamma_1(r, \phi) = \gamma(r\cos\phi, r\sin\phi)$ defined in terms of polar coordinates $r > 0$ and $0 \le \phi < 2\pi$ is given by

$$\gamma_1(r, \phi) = \rho + \rho^2 \omega\zeta \, \tilde{c}_0\left(\omega r \sqrt{1 - (1 - \zeta)\sin^2(\phi - \theta)}\right). \tag{10}$$

For any fixed $r > 0$, the function $\gamma_1(r, \cdot)$ is periodic with period π. When $\tilde{c}_0$ is strictly decreasing, $\gamma_1(r, \cdot)$ has global minimum points at $\phi = \theta$ and $\phi = \theta + \pi$, global maximum points at $\phi = \theta + \pi/2$ and $\phi = \theta + 3\pi/2$ if $0 \leq \theta \leq \pi/2$, and global maximum points at $\phi = \theta - \pi/2$ and $\phi = \theta + \pi/2$ if $\pi/2 \leq \theta < \pi$. Furthermore, define the ϕ-spectrum by

$$\gamma_2(\phi) = \int_0^\infty \gamma_1(r, \phi)\,\mathrm{d}r, \quad 0 \leq \phi < 2\pi. \tag{11}$$

This is also periodic with period π. If $\tilde{c}_0$ is strictly decreasing, then its maximum and minimum points agree with those of $\gamma_1(r, \cdot)$ (for an arbitrary $r > 0$). It is used to investigate directional features. (Similarly, we can define the r-spectrum, which can be used to investigate scales of pattern, see Mugglestone & Renshaw (1996).)

3 Cox processes

3.1 General definition and remarks

Assume that $S = \{S(u) : u \in \mathbb{R}^2\}$ is a non-negative second-order stationary random field with mean one and covariance function

$$c(u) = \mathrm{E}[S(0)S(u)] - 1, \quad u \in \mathbb{R}^2,$$

and that X conditional on S is a Poisson process with an intensity function of the multiplicative form

$$\lambda(u) = \rho(u)S(u), \quad u \in \mathbb{R}^2. \tag{12}$$

Then X is a Cox process driven by λ (Cox 1955). It follows from (12) that X has intensity function ρ, and X is second-order intensity-reweighted stationary with pair correlation function

$$g(u) = 1 + c(u) = \mathrm{E}[S(0)S(u)], \quad u \in \mathbb{R}^2.$$

We therefore call S the residual process, and assume that c (or equivalently g) is elliptical.

In general it is easy to simulate a Poisson process, so simulation of the Cox process within a bounded region $W \subset \mathbb{R}^2$ boils down to simulation of the residual process restricted to W. See e.g. Møller & Waagepetersen (2004) and Appendix A.

In the sequel, when specifying examples of residual processes, we use the function

$$k_\nu(r) = \frac{r^\nu K_\nu(r)}{\pi 2^{\nu+1}\Gamma(\nu+1)}, \quad r \geq 0, \tag{13}$$

where $\nu > -1/2$ and K_ν is the modified Bessel function of the second kind (also called Basset function or MacDonald function). The functions $u \mapsto k_n(\|u\|/\omega)/\kappa$

with $\nu > 0$, $\omega > 0$, and $\kappa > 0$ provide a flexible class of isotropic covariance functions known as Whittle-Matérn covariance function, see e.g. Guttorp & Gneiting (2006, 2010) . Note that $k_\nu(\|u\|)$ considered as function of $u \in \mathbb{R}^2$ is a density.

3.2 Log Gaussian Cox processes

Assuming $\log S$ is a stationary Gaussian random field, then X is a log Gaussian Cox process (LGCP), see Møller, Syversveen & Waagepetersen (1998), Diggle (2003), and Møller & Waagepetersen (2004). Note that X is stationary if and only if ρ is constant.

Denoting C the covariance function of $\log S$, the assumption that $\mathrm{E}S(u) = 1$ means that $\mathrm{E}\log S = -C(0)/2$. Since

$$g(u) = \exp(C(u)), \quad u \in \mathbb{R}^2, \tag{14}$$

we also assume that C is elliptical, i.e.

$$C(u) = C_0\left(\sqrt{u\Sigma^{-1}u^t}\right), \quad u \in \mathbb{R}^2.$$

Taking

$$C_0(r) = k_\nu(r)/\kappa, \quad r \geq 0,$$

where $\nu > 0$ and $\kappa > 0$ are parameters, we obtain

$$g(u) = \exp\left(k_\nu\left(\sqrt{u\Sigma^{-1}u^t}\right)/\kappa\right), \quad u \in \mathbb{R}^2. \tag{15}$$

We refer to this model for X as the Whittle-Matérn LGCP.

Simulation of Gaussian random fields and LGCPs is discussed in Møller & Waagepetersen (2004) and the references therein. Figure 3 show simulated realizations within a square region W of various Cox process models which are used in Sections 4–5. The top panels in Figure 3 concern stationary Whittle-Matérn LGCPs within $W = [0,1]^2$ and with parameters $\rho = 200$ (top left panel) or $\rho = 500$ (top right panel), $\theta = \pi/4$ (both top panels), $\zeta = 0.6$ (top left panel) or $\zeta = 0.2$ (top right panel), and where the remaining parameter values are specified in Table 1 (see 'dataset 3' and 'dataset 4'). The anisotropy is clearly seen in the top right panel, and also to some extent in the top left panel.

3.3 Shot noise Cox processes

3.3.1 Definition

Let Φ be a stationary Poisson process on $\mathbb{R}^2$ with intensity $\kappa > 0$, and f be a quadratically integrable density function on $\mathbb{R}^2$. Define the residual process by

$$S(u) = \frac{1}{\kappa}\sum_{v\in\Phi} f(u - v), \quad u \in \mathbb{R}^2. \tag{16}$$

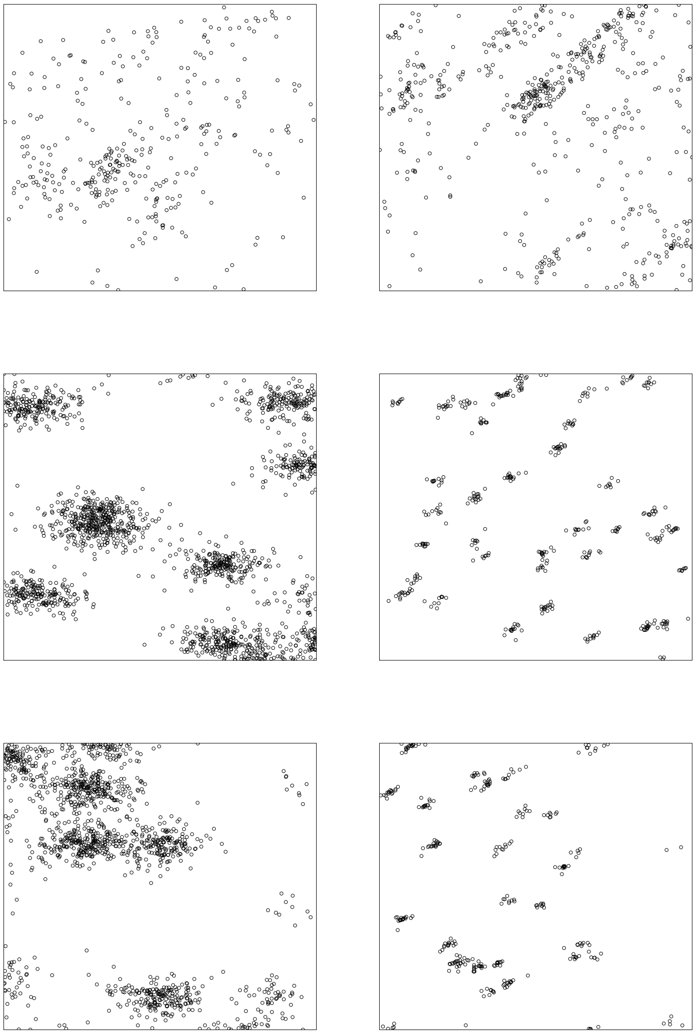


Figure 3: Simulated realizations within a square region W of various Cox process models. Top left panel: stationary Whittle-Matérn LGCP, with $\rho = 200$ and $W = [0,1]^2$. Top right panel: stationary Whittle-Matérn LGCP, with $\rho = 500$ and $W = [0,1]^2$. Mid left panel: stationary Whittle-Matérn SNCP, with $\rho = 200$ and $W = [0,3]^2$. Mid right panel: stationary Whittle-Matérn SNCP, with $\rho = 100$ and $W = [0,2]^2$. Bottom left panel: inhomogeneous Whittle-Matérn SNCP, with $\rho(x,y) = 400(3-x)/3$ on $W = [0,3]^2$. Bottom right panel: inhomogeneous Whittle-Matérn SNCP, with $\rho(x,y) = 100(2-x)$ on $W = [0,2]^2$.

Then X is a shot noise Cox process (SNCP), see Møller (2003), Møller & Waagepetersen (2004, 2007), and the references therein.

Clearly, S is stationary and $\mathrm{E}S(u) = 1$, while X is stationary if and only if ρ is constant. Assume that $f = f_\Sigma$ is elliptical, i.e.

$$f_\Sigma(u) = f_0\left(\sqrt{u\Sigma^{-1}u^t}\right)|\Sigma|^{-1/2}, \quad u \in \mathrm{R}^2. \tag{17}$$

Then

$$c(u) = f_\Sigma * f_\Sigma(u)/\kappa = f_I * f_I(u\Sigma^{-1/2})/\left[\kappa|\Sigma|^{1/2}\right], \quad u \in \mathbb{R}^2, \tag{18}$$

is elliptical, where $*$ denotes convolution.

Taking $f_0 = k_\nu$ with $\nu > -1/2$, then

$$g(u) = 1 + k_{2\nu+1}\left(\sqrt{u\Sigma^{-1}u^t}\right)/\left[\kappa|\Sigma|^{1/2}\right], \quad u \in \mathbb{R}^2. \tag{19}$$

We refer to this model for X as the Whittle-Matérn SNCP.

3.3.2 Remarks

Cluster process: The SNCP process has an interpretation as a Poisson cluster process, since X is distributed as a superposition $\cup_{v\in\Phi}X_v$, where conditional on Φ, the cluster X_v is a Poisson process with intensity function $\lambda_v(u) = \rho(u)f(u-v)/\kappa$, and the clusters are independent. We therefore refer to the events of Φ as cluster centres, and to λ_v as the 'offspring intensity' (associated to the cluster centre v). If X is stationary, then λ_v is elliptical, and so the clusters tend to have an elliptical shape.

Simulation: Appendix A discusses how to simulate X within a bounded region $W \subset \mathbb{R}^2$, considering another bounded region $W_{\mathrm{ext}} \supseteq W$ which is so large that $X_{\max} \cap W$ is well approximated by replacing the infinite Poisson process Φ by the finite Poisson process $\tilde{\Phi} = \Phi \cap W_{\mathrm{ext}}$ in the simulation. The approximation may be evaluated by calculating the probability q_W that some cluster of $X_{\max}$ with its centre outside W_{ext} intersects W. Equation (44) in Appendix A establishes an upper bound on q_W. Furthermore, Appendix B discusses conditional simulation of Φ given a realization $X_{\max} \cap W = x$.

The mid panels in Figure 3 show simulated realizations of stationary SNCPs, while the bottom panels show simulated realizations of inhomogeneous SNCPs. In the panels to the left, the realizations are restricted to $W = [0,3]^2$, all parameter values are the same and given in Table 1 (see 'dataset 1') except that $\rho = 200$ in the mid left panel and $\rho(x,y) = 400(3-x)/3$ (for $(x,y) \in W$) in the bottom left panel; thus the expected number of points in W is the same in the two cases, namely $\int_W \rho(u)\,\mathrm{d}u = 1800$. In the panels to the right, the realizations are restricted to $W = [0,2]^2$, all parameter values are the same and given in Table 1 (see 'dataset 2') except that $\rho = 100$ in the mid right panel and $\rho(x,y) = 100(2-x)$ (for $(x,y) \in W$) in the bottom right panel; thus the expected number of points in W is 400 in the two cases. The region W_{ext} is rectangular, with sides in the directions of θ and $\theta + \pi/2$, and determined by requiring that $q_W \leq 0.01$ (in fact q_W is much smaller than 0.01, as this value is used for the upper bound (44)). The sides of W_{ext} are about seven and four times larger than the sides of $W = [0,3]^2$ in case of the mid left panel (dataset 1), and about seven and three times larger than the sides of $W = [0,2]^2$ in case of the mid right panel (dataset 2).

Spectral density and K-function: By (6) and (18), the spectral density is

$$\gamma(u) = \rho + \rho^2 |\Sigma|^{1/2} \tilde{f}_I(u\Sigma^{1/2})^2/\kappa, \quad u \in \mathbb{R}^2. \tag{20}$$

In the geometric isotropic case where $\Sigma = \omega^2 I$, this reduces by the Hankel transform, cf. (8).

For example, consider the Whittle-Matérn SNCP. Combining (8), (13), and (20) give

$$\gamma(u) = \rho + \rho^2 |\Sigma|^{1/2} \left(1 + (2\pi)^2 u\Sigma u^t\right)^{-2(\nu+1)} / \kappa$$

or in terms of polar coordinates

$$\gamma_1(r,\phi) = \rho + \rho^2 \omega\zeta \, \left(1 + (2\pi)^2 (\omega r)^2 \left(1 - (1-\zeta\,) \sin^2(\phi - \theta)\right)\right)^{-2(\nu+1)} / \kappa. \tag{21}$$

Furthermore, combining (4) and (19) with

$$\int_0^r t^{\nu+1} K_\nu(t)\, \mathrm{d}t = -r^{\nu+1} K_{\nu+1}(r) + 2^\nu \Gamma(\nu+1)$$

(Abramowitz & Stegun 1964), we obtain in the isotropic case where $\Sigma = \omega^2 I$,

$$K(r) = \pi r^2 + \left[1 - 8\pi(\nu+1) k_{2\nu+2}(r/\omega)\right]/\kappa, \quad r \geq 0. \tag{22}$$

Thomas process: If f_Σ is the density of $N_2(0,\Sigma)$ (the bivariate zero-mean normal distribution with covariance matrix Σ), $\Sigma = \omega^2 I$, and $\rho(u) = \rho$ is constant, then X is the well-known (modified) Thomas process (Thomas 1949, Cox & Isham 1980), and

$$c(u) = \exp\left(-\|u\|^2/(4\omega^2)\right) / \left[4\pi\omega^2\kappa\right]$$

is an isotropic Gaussian covariance function. For the extension of the Thomas process to the case where Σ can be any 2×2 symmetric positive definite matrix, κc becomes the density of $N_2(0, 2\Sigma)$, and the model is a limiting case of the Whittle-Matérn SNCP. This extension with a constant intensity $\rho(u) = \rho$ was studied in Castelloe (1998).

4 Estimation procedures based on first and second order properties

Let $W \subset \mathbb{R}^2$ denote a bounded observation window of area $|W| > 0$, and set $X_W = X \cap W$. Suppose that a realization x of X_W is observed and a parametric model for X is specified. Quick non-likelihood approaches to parameter estimation when X is second-order intensity-reweighted stationary with an isotropic pair correlation function, using various estimating functions based on first and second order properties, are reviewed in Møller & Waagepetersen (2007). In much the same spirit, Tanaka, Ogata & Stoyan (2008) suggested a method for parameter estimation based on the intensity function of the difference process $\{u - v : u, v \in X\}$, assuming stationarity and isotropy of the spatial point process.

This section considers instead parameter estimation procedures for our situation where X is a spatial point process with an elliptical pair correlation function. The methods are in particular useful for Cox processes. Non-parametric kernel estimators of various functional summary statistics will be used, where (approximate) unbiasedness follows from (1)–(2), and where we use the generic notation k_h for a kernel function with bandwidth $h > 0$. As usual the results are sensitive to the choice of bandwidth, which is estimated by inspection of plots of the functional summary statistic.

The methods are illustrated by using the simulations in Figure 3 as synthetic data, where we refer to the realizations of the two stationary Whittle-Matérn SNCPs in the mid panels as datasets 1 and 2, respectively, to the realizations of the two stationary Whittle-Matérn point LGCP's in the top panels as datasets 3 and 4, respectively, and to the realizations of the two inhomogeneous Whittle-Matérn SNCPs in the bottom panels as datasets 5 and 6, respectively. Datasets 5–6 are only considered in Section 4.4.1.

4.1 Estimation of the intensity function

When ρ is assumed to be constant on W, the usual non-parametric estimate is $\hat{\rho} = n(x)/|W|$, where $n(x)$ denotes the observed number of points. For non-parametric estimation in the inhomogeneous case of ρ, we follow Diggle (1985, 2010) in using

$$\hat{\rho}(u) = \sum_{v \in x} k_h(u - v)/c_{h,W}(v) \tag{23}$$

where k_h is an isotropic normal kernel and $c_{h,W}(v) = \int_W k_h(u - v)\,\mathrm{d}u$ is an edge correction factor. In both cases, the estimated mean number of points $\int_W \hat{\rho}(u)\,\mathrm{d}u$ agrees with $n(x)$.

If a parametric model $\rho(u) = \rho_\psi$ is assumed, e.g., a parametric model involving covariate information, then the parameter ψ may be estimated from a composite likelihood which coincides with the likelihood for a Poisson process with intensity function ρ_ψ, see Section 8.1 in Møller & Waagepetersen (2007).

	θ	ζ	ω	ν	κ	ρ	W
Dataset 1	0	0.5	0.1	2	1	200	$[3,3]^2$
Method A	176	0.48	0.081	2	1.3	188	
Method B	164	0.48	0.085	2	1.21	188	
Dataset 2	30	0.43	0.02	0.5	10	100	$[2,2]^2$
Method A	30.6	0.42	0.021	0.5	11.4	103	
Method B	17.27	0.43	0.019	0.5	12	103	
Dataset 3	45	0.6	0.05	5	0.015	200	$[1,1]^2$
Method A	37.2	0.55	0.08	5	0.03		
Dataset 4	45	0.2	0.2	0.5	0.08	500	$[1,1]^2$
Method A	43	0.19	0.29	0.5	0.1	541	

Table 1: Specification of W and the true and estimated parameter values for the stationary Whittle-Matérn SNCPs (datasets 1 and 2) and the stationary Whittle-Matérn point LGCP's (datasets 3 and 4). The values of the angle θ are given in degrees between 0 and 180. Method A refers to the estimation method based on the procedures in Sections 4.2 and 4.3.1, and Method B to the estimation method based on the procedures in Sections 4.2.2 and 4.3.2.

4.2 Estimation of θ and ζ

Sections 4.2.1–4.2.2 discuss two ways of estimating θ and ζ. If a parametric model $\rho(u) = \rho_\psi$ is assumed (see Section 4.1), we assume that the range of (θ, ζ) is not depending on ψ, i.e. (θ, ζ) has range $[0, 2\pi) \times (0, 1]$ no matter the value of ψ.

4.2.1 Estimation procedures based on the pair correlation function

Method: Assume that g_0 is strictly decreasing. For instance, this is the case in (15) and (19) and for most other Cox process models as well. It follows from the monotonicity and periodicity properties of $g_1(r, \cdot)$ (see Section 2.2) that given user-specified parameters $b_1 > a_1 \geq 0$, the function

$$D(\phi) = \int_{a_1}^{b_1} [g_1(r,\phi) - g_1(r,\phi+\pi/2)]\,\mathrm{d}r, \quad 0 \leq \phi < \pi, \tag{24}$$

has a unique maximum at $\phi = \theta$.

First, we estimate θ by an approximation of this maximum as follows.

We use a non-parametric estimate suggested in Stoyan & Stoyan (1994) for the pair correlation function g_1 given by (5) and modified to our case of second-order intensity-reweighted stationarity: for $r > 0$ and $0 \leq \phi < \pi$,

$$\hat{g}_1(r,\phi) = \hat{g}_1(r,\phi+\pi) = \frac{1}{2} \sum_{u,v \in x:\, u \neq v} \frac{K(u-v,(r,\phi)) + K(u-v,(r,\phi+\pi))}{\hat{\rho}(u)\hat{\rho}(v)|W \cap W_{u-v}|} \tag{25}$$

where $\hat{\rho}$ is an estimate obtained as discussed in Section 4.1, W_{u-v} denotes W translated by $u-v$, $|W \cap W_{u-v}|$ is an edge correction factor, and

$$K(u-v,(r,\phi)) = \frac{1}{r} k_{h_r}(\|u-v\| - r) k_{h_\phi}(\alpha(u,v) - \phi) \tag{26}$$

where $\alpha(u,v)$ is the angle between the directed line from u to v and the abscissa-axis. The general problem of finding a non-parametric estimate of the pair correlation function was discussed by Stoyan & Stoyan (2000).

Now, compute $\hat{g}_1(r,\phi)$ on a rectangular grid $\{(r_i,\phi_j) : i = 1,...,n_r, j = 1,...,n_\phi\}$, where $r_i = a_1 + (i-1/2)(b_1-a_1)/n_r$ and $\phi_j = (j-1/2)\pi/n_\phi$, and so that the function D (up to proportionality) is well represented by

$$\hat{D}(\phi_j) \propto \sum_{i=1}^{n_r} [\hat{g}_1(r_i,\phi_j) - \hat{g}_1(r_i,\phi_j+\pi/2)]/n_r, \quad j = 1,\ldots,n_\phi. \tag{27}$$

Then we estimate θ by

$$\hat{\theta} = \arg_{\phi_j} \max \hat{D}(\phi_j). \tag{28}$$

Second, we obtain an estimate of ζ based on the following considerations.

Let $B(\theta,\zeta) = U_\theta \mathrm{diag}(1, 1/\zeta)$ and consider the transformed point process

$$Y = XB(\theta,\zeta) = \omega X \Sigma^{-1/2}. \tag{29}$$

Note that Y has intensity function

$$\rho_{\theta,\zeta}(u) = \rho\left(uU_\theta \mathrm{diag}\left(1,\zeta\right)\right)\zeta \tag{30}$$

and an isotropic pair correlation function which with respect to polar coordinates is given by

$$g_{\theta,\zeta,1}(r,\phi) = g_0(r/\omega). \tag{31}$$

Replacing θ by $\hat{\theta}$, then for any $\zeta \in (0,1]$, the functions $\rho_{\hat{\theta},\zeta}$ and $g_{\hat{\theta},\zeta,1}$ are estimated in the same way as above but based on the transformed data and transformed window,

$$y = xB(\hat{\theta},\zeta), \quad V = WB(\hat{\theta},\zeta),$$

i.e. in (25) we replace x by y, $\hat{\rho}$ by $\hat{\rho}_{\hat{\theta},\zeta}$, and the edge correction factor by

$$|V \cap V_{u-v}| = |W \cap W_{u'-v'}|/\zeta$$

for any distinct points $u, v \in y$, with corresponding points $u' = uB(\hat{\theta},\zeta)^{-1}$ and $v' = vB(\hat{\theta},\zeta)^{-1}$ from x. Then, for different values of $\zeta \in (0,1]$ (e.g. $\zeta = 0.01, 0.02, ..., 1$), we consider plots of $\hat{c}_{\hat{\theta},\zeta,1}(r,\phi) = \hat{g}_{\hat{\theta},\zeta,1}(r,\phi) - 1$. By (31), $c_{\theta,\zeta,1}(r,\phi) = g_0(r/\omega) - 1$ does not depend on ϕ, and we let therefore $\hat{\zeta}$ be the value of ζ where the contours of $\hat{c}_{\hat{\theta},\zeta,1}(r,\phi)$ are closest to vertical lines.

Example: We find suitable values for the bandwidths h_ϕ and h_r in (26) by looking at contour plots of $\hat{g}_1(r,\phi)$ for several different bandwidths, where we choose the smallest bandwidths such that $\hat{g}_1(r,\phi)$ still appears to be smooth. The chosen values for datasets 1–2 and the corresponding contour plot of $\hat{g}_1(r,\phi)$ are shown in the top panels of Figure 4. The dashed vertical lines in these panels mark the values of a_1 and b_1 used in (24). Here a_1 is chosen such that for $r < a_1$, the function $\hat{g}(r,\phi)$ rises rapidly as r decreases. The value of b_1 is chosen such that for $r > b_1$, $\hat{g}(r,\phi)$ is approximately 0 for all ϕ. For dataset 1, $(a_1, b_1) = (0.05, 0.6)$, while for dataset 2, $(a_1, b_1) = (0.01, 0.25)$, reflecting the difference in the size of the clusters in the two datasets. In a similar way we have chosen the values of h_ϕ, h_r, a_1, and b_1 when considering datasets 3–4. Then from (28) we find the estimates of the angle θ for datasets 1–4 given in Table 1 (see the rows named Method A). These estimates are in rather close agreement with the true values of θ (noticing that for dataset 1, $\hat{\theta} = 176^\circ$ is close to 180° which by periodicity corresponds to the true value of $\theta = 0^\circ$, cf. Section 2.2). Finally, the estimates of ζ for datasets 1–4 are obtained as described above, where the estimates and the true values in Table 1 are in good agreement. The mid panels in Figure 4 show the (approximately) vertical contours of $\hat{g}_{\hat{\theta},\hat{\zeta},1}(r,\phi)$ for datasets 1–2, where the contours fluctuate less for dataset 1 than for dataset 2, possibly because dataset 1 contains about twice as many points as dataset 2 and the anisotropy factors are nearly the same for the two datasets.

4.2.2 Estimation procedures based on the spectral density

Method: Assume that X is stationary, the spectral density exists, and the function $\tilde{c}_0$ from (10) is strictly decreasing (e.g. this is the case in (21)). By the monotonicity and periodicity properties of $g_1(r,\cdot)$,

$$\theta = \arg_{\phi\in[0,\pi)} \max\left[\gamma_2(\phi+\pi/2) - \gamma_2(\phi)\right] \tag{32}$$

where γ_2 is the ϕ-spectrum, cf. (11).

We follow Mugglestone & Renshaw (1996) in estimating $\gamma_1(r,\phi) = \gamma_1(r,\phi+\pi)$ by

$$\hat{\gamma}_1(r,\phi) = \hat{\gamma}_1(r,\phi+\pi) = \frac{1}{|W|}\sum_{u,v\in x} \mathrm{e}^{-\mathrm{i}w(u-v)^t}, \quad r>0,\ 0\le\phi<\pi, \tag{33}$$

where $w = (r\cos\phi, r\sin\phi)$. Now, let $b_2 > a_2 \geq 0$ be user-specified parameters, and let

$$D_\gamma(\phi) = \int_{a_2}^{b_2} \gamma_1(r,\phi+\pi/2) - \gamma_1(r,\phi)\,\mathrm{d}r. \tag{34}$$

Then $D_\gamma(\phi)$ has a maximum at $\phi = \theta$. As in Section 4.2.1, using a fine rectangular grid $\{(r_i,\phi_j) : i = 1,...,n_r, j = 1,...,n_\phi\}$ where now $r_i = a_2 + (i - 1/2)(b_2 - a_2)/n_r$, we estimate $D_\gamma(\phi_j)$ up to proportionality by

$$\hat{D}_\gamma(\phi_j) \propto \frac{1}{n_r}\sum_{i=1}^{n_r}\left[\hat{\gamma}_1(r_i,\phi_j+\pi/2) - \hat{\gamma}_1(r_i,\phi_j)\right], \quad j = 1,...,n_\phi. \tag{35}$$

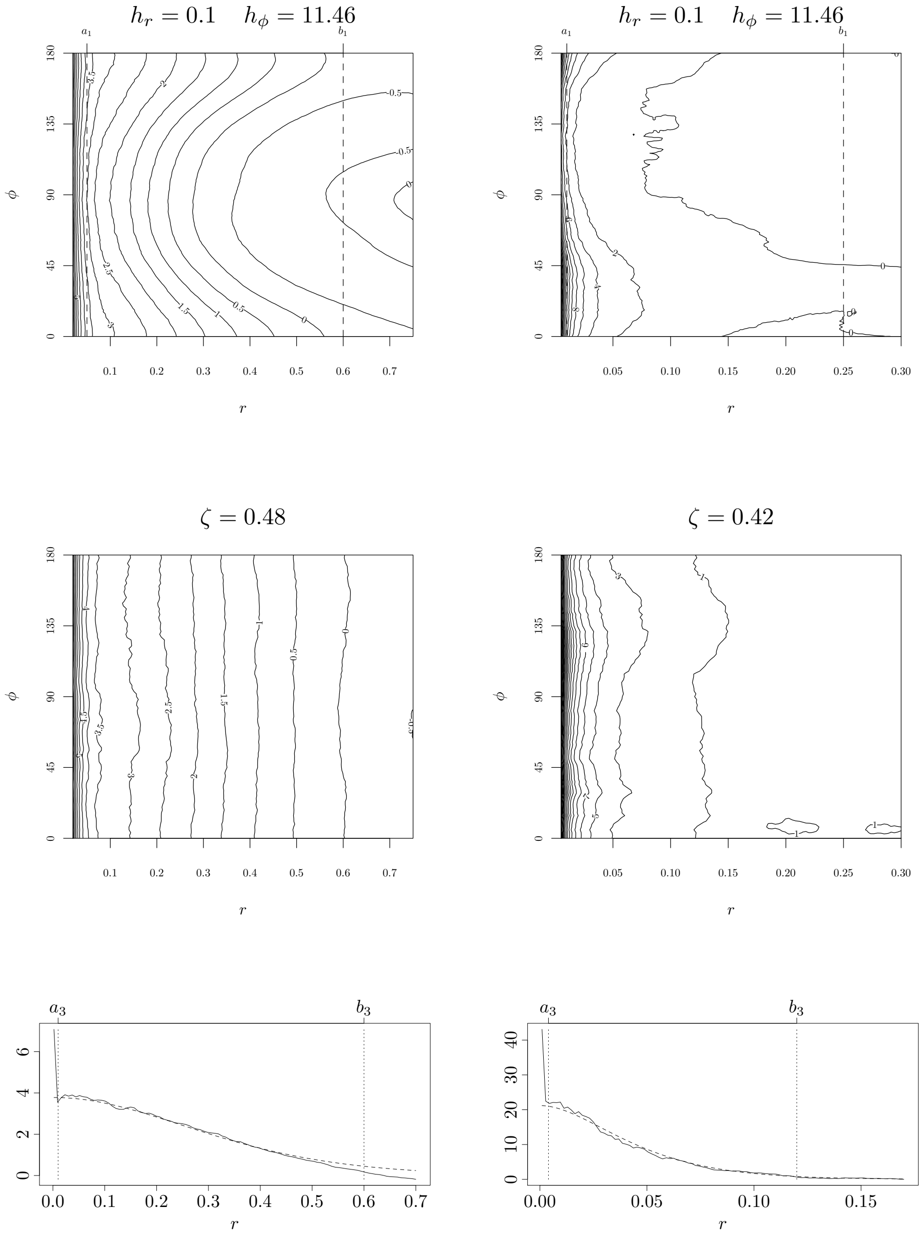


Figure 4: Plots for the procedure of estimating θ and ζ based on datasets 1 (left panels) and 2 (right panels). The top panels show contour plots of $\hat{g}_1(r, \phi)$ and the limits of integration a_1 and b_1 (dashed vertical lines), and the bandwidths are given above each plot. The mid panels show contour plots of $\hat{g}_{\hat{\theta},\hat{\zeta},1}(r, \phi)$ where the values of $\hat{\zeta}$ are given above each plot. The bottom panels show $\hat{g}_{Y,1}(r)$ and the limits of integration a_3 and b_3 (dashed vertical lines).

First, in accordance with (32), we estimate θ by

$$\hat{\theta} = \arg_{\phi_j} \max \left[\hat{D}_\gamma(\phi_j) \right].$$

Second, similar to the way ζ was estimated in Section 4.2.1, for different values of $\zeta \in (0, 1]$, we consider transformed data $y = xB(\hat{\theta}, \zeta)$ and corresponding non-parametric estimates $\hat{\gamma}_{\hat{\theta},\zeta,1}$ and $\hat{D}_{\hat{\theta},\zeta,\gamma}$ obtained in a similar way as in (33)–(35). As opposed to the true spectral density $\gamma_{\theta,\zeta,1}$ which does not depend on ϕ, $\hat{\gamma}_{\hat{\theta},\zeta,1}(r, \phi)$ exhibits an oscillatory behaviour, and we therefore do not look at the contours of $\hat{\gamma}_{\hat{\theta},\zeta,1}(r, \phi)$ directly. Instead we estimate ζ by the value where $\hat{D}_{\hat{\theta},\zeta,\gamma}$ is closest to a constant.

Example: The parameter estimates of θ and ζ for datasets 1–2 and obtained by the spectral approach are given in Table 1 (see the rows named Method B). As expected, the results are much like those obtained in Section 4.2.1 (see the rows named Method A).

It is helpful to look at a contour plot of $\hat{\gamma}_1(r, \phi)$ when choosing a_2 and b_2 used in (34)–(35). Contour plots based on datasets 1–2 are found in the top panels of Figure 5, where also the choices of a_2 and b_2 are shown. Considering each dataset, for $r < a_2$, the function $\hat{\gamma}_1(r, \phi)$ rises rapidly as r decreases, and for $r > b_2$, $\hat{\gamma}(r, \phi)$ is almost constant. Plots of the function $\hat{D}_\gamma$ (or more precisely the right hand side of (35)) based on datasets 1–2 are shown in the mid panels of Figure 5.

4.3 Estimation of remaining parameters

In the following we replace (θ, ζ) by the estimate $(\hat{\theta}, \hat{\zeta})$ obtained as discussed above. Estimation of the remaining parameters given by ω and a parametric model for g_0 may be based on the transformed point process $Y = XB(\theta, \zeta)$, with corresponding data $y = xB(\theta, \zeta)$, estimated intensity function $\hat{\rho}_Y(u)$ given by (30) when ρ is replaced by an estimate $\hat{\rho}$ (Section 4.1), and isotropic pair correlation function $g_{Y,1}(r) = g_0(r/\omega)$, cf. (31). Since Y is second-order intensity-reweighted stationary with an isotropic pair correlation function, the various methods in Møller & Waagepetersen (2007) and Tanaka et al. (2008) (discussed at the very beginning of Section 4) can be applied. Below we concentrate on two minimum contrast procedures based on respective the pair correlation function for Y (Section 4.3.1) and the K-function for Y (Section 4.3.2). If a parametric model $\rho(u) = \rho_\psi$ is assumed (see Section 4.1), we assume that the range of ω and the parameters for g_0 is not depending on ψ.

Below, for specificity, we consider the Whittle-Matérn SNCP model (Section 3.3). Then Y is also a Whittle-Matérn SNCP, where the intensity parameter of the centre process is $\kappa\zeta$, and by (19) and (31), $g_{Y,1}(r) = 1 + c_{\zeta,\omega,\nu,\kappa}(r)$ where

$$c_{\zeta,\omega,\nu,\kappa}(r) = k_{2\nu+1}(r/\omega) / \left[\kappa\zeta\omega^2\right]. \tag{36}$$

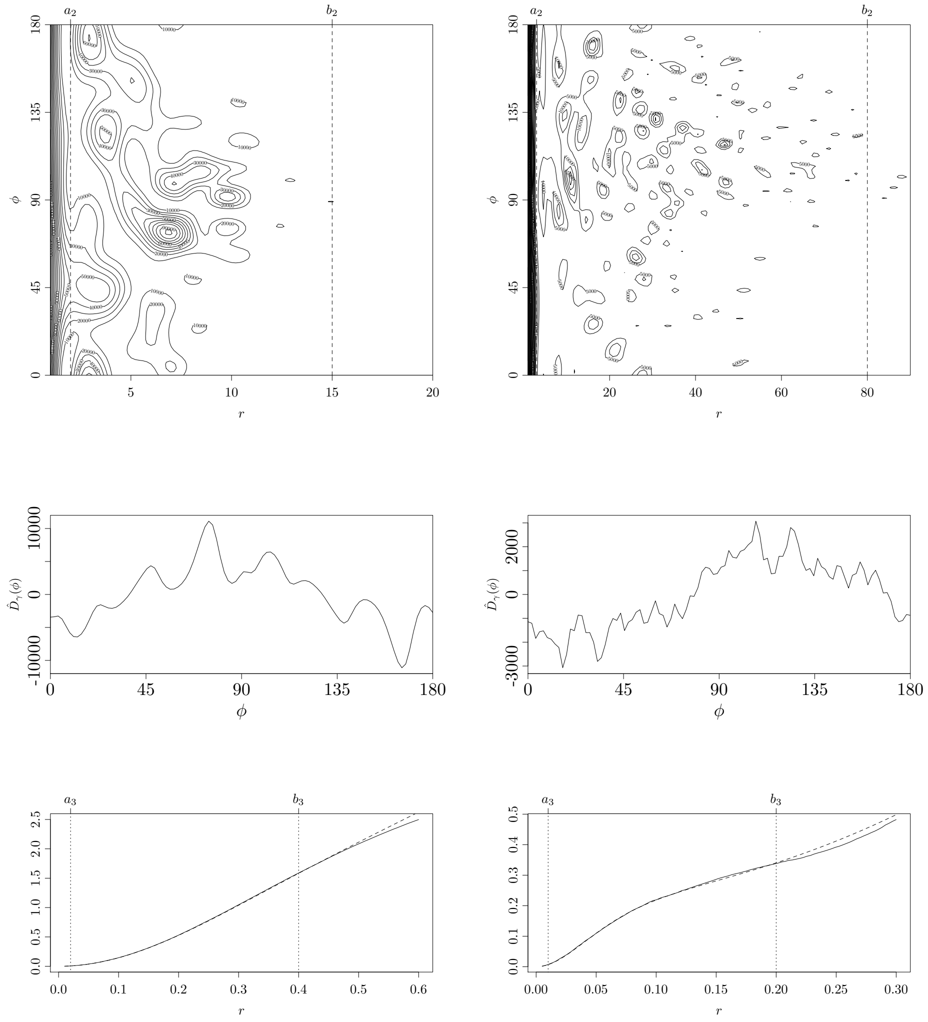


Figure 5: Plots associated to method B and based on dataset 1 (left panels) and dataset 2 (right panels). Top panels: contour plots of $\hat{\gamma}_1(r, \phi)$, with the values of a_2 and b_2 indicated (dashed lines). Mid panels: plots of $\hat{D}_\gamma$. Bottom panels: plots of $\hat{K}_Y$ (solid line) and the fitted K-function (dashed line), with the values of a_3 and b_3 indicated (dotted lines).

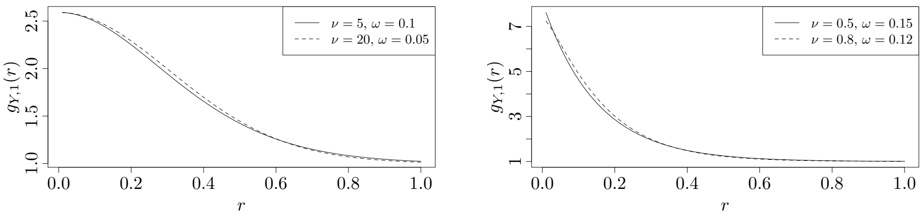


Figure 6: Plots of $g_{Y,1}(r)$ for different values of (ω, ν).

As we are substituting ζ by its estimate $\hat{\zeta}$, the parameters to be estimated are $\omega > 0$, $\nu > -1/2$, and $\kappa > 0$. In the sequel, we restrict $2\nu + 1$ to a few values, i.e. the set $\{0.1, 0.5, 1, 2, 5\}$. In fact, if we let $(\omega, 2\nu+1, \kappa)$ vary freely in $(0, \infty)^3$, it becomes difficult to estimate the parameters, since $g_{Y,1}$ for rather different values of (ω, ν) may have a similar behaviour; see Figure 6.

4.3.1 Minimum contrast estimation based on the pair correlation function

Method: A non-parametric estimate for $g_{Y,1}(r)$ is given by

$$\hat{g}_{Y,1}(r) = \frac{1}{2\pi r} \sum_{u,v \in y:\, u \neq v} \frac{k_h(r - \|u - v\|)}{\hat{\rho}_Y(u)\hat{\rho}_Y(v)|V \cap V_{u-v}|}, \quad r > 0, \tag{37}$$

see e.g. Illian et al. (2008). We obtain a minimum contrast estimate by minimizing the L_2-distance between $\hat{g}_{Y,1}(r)$ and $g_{Y,1}(r)$ restricted to a finite interval:

$$(\hat{\omega}, \hat{\nu}, \hat{\kappa}) = \arg\min_{\omega,\nu,\kappa} \int_{a_3}^{b_3} \left[\hat{g}_{Y,1}(r) - 1 - c_{\hat{\zeta},\omega,\nu,\kappa}(r)\right]^2 \mathrm{d}r \tag{38}$$

where $b_3 > a_3 > 0$ are user-specified parameters. The solution for κ can be expressed in terms of ω and ν,

$$\hat{\kappa}(\omega, \nu) = A(\omega, \nu)/B(\omega, \nu)$$

where

$$A(\omega, \nu) = \frac{1}{1 - \hat{e}^2} \int_{a_3}^{b_3} k_{\omega,2\nu+1}(r)^2 \,\mathrm{d}r$$

and

$$B(\omega, \nu) = \frac{1}{\sqrt{1 - \hat{e}^2}} \int_{a_3}^{b_3} \hat{c}(r) k_{\omega,2\nu+1}(r) \,\mathrm{d}r.$$

Therefore, the problem is reduced to an optimization over (ω, ν), i.e.

$$(\hat{\omega}, \hat{\nu}) = \arg\max_{\omega,\nu} \frac{B(\omega, \nu)^2}{A(\omega, \nu)}$$

and $\hat{\kappa} = A(\hat{\omega}, \hat{\nu})/B(\hat{\omega}, \hat{\nu})$.

Remark: If instead the model is a Whittle Matérn LGCP, the minimum contrast estimation is done in a similar way, but considering a L_2-distance between the covariance function $k_\nu(r/\omega)/\kappa$ (of the underlying Gaussian random field corresponding to Y) and its non-parametric estimate $\log(\hat{g}_{Y,1}(r))$, cf. (14))–(15)), and now restricting ν to the set $\{0.1, 0.5, 1, 2, 5\}$.

Example: We choose a value for the bandwidth h in (37) in a similar way as we chose h_r and h_ϕ in Section 4.2.1, resulting in a reasonably smooth function $\hat{g}_{Y,1}(r)$. In the bottom panels of Figure 4, $\hat{g}_{Y,1}(r)$ for datasets 1 and 2 are plotted. The shape of $\hat{g}_{Y,1}(r)$ in these plots is very typical, and it is shown how we choose a_3 and b_3 from (38). As r approaches 0, $\hat{g}_{Y,1}(r)$ suddenly rises rapidly. This is because for very small r-values, there is a lack of data and so it is hard to obtain a reliable estimate of $g_{Y,1}(r)$. Thus a_3 should be larger than the point where this rapid raise starts. The end point b_3 is chosen where $\hat{g}_{Y,1}(r)$ is close to 0. The results of the estimation for datasets 1–4 are found in Table 1. Overall the agreement between true and estimated parameter values is reasonable good.

4.3.2 Minimum contrast estimation based on the K-function

Method: The minimum contrast estimation procedure above relies on the kernel estimator $\hat{g}_Y$, and the choice of bandwidth is essential to the quality of the results. By instead basing the estimation on the K-function it is possible to avoid kernel estimators.

The K-function of Y, which we denote K_Y, is given by the closed form expression (22) if κ is replaced by $\kappa\zeta$. Here, as above, we replace ζ by its estimate. Further, in the minimum contrast estimation procedure (38), we can replace $1 + c_{\hat{e},\omega,\nu,\kappa}$ by K_Y, and $\hat{g}_Y(r)$ by

$$\hat{K}_Y(r) = \sum_{u,v \in y:\, u \neq v} \frac{\mathbf{1}(\|u - v\| \leq r)}{\hat{\rho}_Y(u)\hat{\rho}_Y(v)|V \cap V_{u-v}|}$$

where $1(\cdot)$ denotes the indicator function. We still use the notation a_3 and b_3 for the range of integration, though their values may be rather different from the values in Section 4.3.1 as demonstrated below. As before the problem reduces to an optimization over (ω, ν).

Example: In order to limit the number of procedures to compare, let us restrict attention to method B, i.e. when first θ and ζ are estimated using the spectral density (Section 4.2.2) and second the remaining parameters are estimated by the method above using the K-function. As an analytic expression for $K(r)$ is not available for the Whittle-Matérn LGCP model, we only apply method B to the SNCP datasets 1–2.

When estimating ν, ω, and κ based on $\hat{K}_Y(r)$, we have the advantage that we do not need to choose a bandwidth. However we still have to choose a_3 and b_3. The choice of b_3 can have a quite large impact on the estimators. In our experience a good strategy is to use a relatively large value of b_3. If b_3 is too large, the difference between the theoretical and empirical K_Y-functions will be large at large distances, and we choose b_3 as the largest value such that this problem does not seem to appear. The value of a_3 should be small, but its value does not seem to significantly affect the results. The fitted K_Y function together with its non-parametric estimate and the values of a_3 and b_3 are shown in the bottom panel of Figure 5 for each of the datasets 1–2.

4.4 Simulation and automatic estimation procedures

The behaviour of the estimators introduced above can be investigated by simulations from a Whittle Matérn SNCP or LGCP model. However, such a simulation study can be rather elaborate, since the estimation procedures involve many subjective choices of user-specified parameters. We now consider three modifications of our estimation procedures with some automatic steps.

Method: First, we decide on the values of the bandwidths and the integration limits in advance by looking at estimates from a few simulated dataset. Second, for any further simulated dataset x, when finding $\hat{\zeta}$ we only search in the set $\{0.1, 0.2, \ldots, 1\}$. Third, when using method A to find first $\hat{\theta}$ and next $\hat{\zeta}$, we consider (27) but with respect to the transformed data y, i.e.

$$\hat{D}_Y(\phi_j) \propto \sum_i [\hat{g}_{\hat{\theta},\zeta,1}(r_i, \phi_j) - \hat{g}_{\hat{\theta},\zeta,1}(r_i, \phi_j + \pi/2)]/n_r$$

and then let $\hat{\zeta}$ be the value of ζ that minimizes

$$\sum_{j=1}^{n_\phi} \left(\hat{D}_Y(\phi_j) - \frac{1}{n_\phi} \sum_{i=1}^{n_\phi} \hat{D}_Y(\phi_j) \right)^2 . \tag{39}$$

The reasoning behind this is that, in the isotropic case of Y, $\hat{g}_{\hat{\theta},\zeta,1}(r, \phi)$ estimates a function that is constant with respect to ϕ, and so $\hat{D}(\phi)$ is expected to deviate less from its mean than it would in the anisotropic case. It is our experience that most of the time this 'automatic estimator' of ζ will be very close to the estimator in Section 4.2.1 resulting from a manual inspection.

When using method B we make a similar modification. We define $\hat{D}_{\gamma,Y}$ in the same way as $\hat{D}_Y$ but with $\hat{g}_{\hat{\theta},\zeta,1}$ replaced by $\hat{\gamma}_{\hat{\theta},\zeta,1}$. Then $\hat{\zeta}$ is the value that minimizes (39) with D_Y replaced by $D_{\gamma,Y}$.

Example: For each true model corresponding to datasets 1–4, we have simulated 200 datasets, and used the values of h_r, h_ϕ, a_1, b_1, h, a_3, and b_3 found above when estimating the model parameters by the automated version of method A. Figure 7 shows the empirical distributions of the estimators. The main impression is that the true parameter values lie within the high probability areas of these distributions.

For the 200 datasets simulated from the model corresponding to datasets 1 and 2, we have also applied the automated version of method B. The empirical

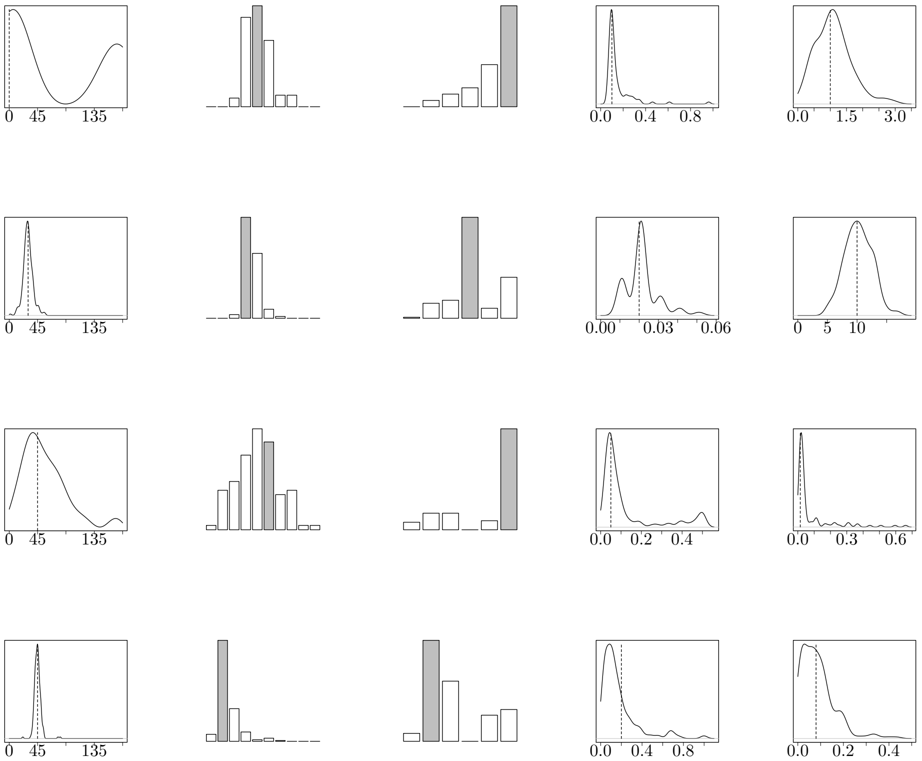


Figure 7: From top to bottom: empirical distributions of the estimators $\hat{\theta}$, $\hat{\zeta}$, $\hat{\nu}$, $\hat{\omega}$, and $\hat{\kappa}$ (from left to right) based on 200 simulations from each of the two SNCP models (corresponding to datasets 1–2) and the two LGCP models (corresponding to datasets 3–4) and using method A. The true parameter values are indicated by a dashed line in the continuous case or a gray shade in the bar-plots. In the bar-plot for ζ, the bars correspond to $0.1, 0.2, ..., 1$. In the bar-plot for ν, the bars correspond to $2\nu + 1 = 0.1, 0.5, 1, 2, 5$ for the SNCP models, and to $\nu = 0.1, 0.5, 1, 2, 5$ for LGCP models.

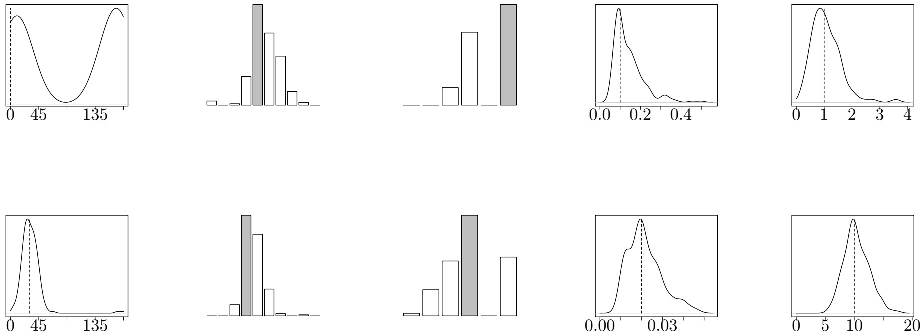


Figure 8: Top and bottom panels: empirical distributions of the estimators $\hat{\theta}$, $\hat{\zeta}$, $\hat{\nu}$, $\hat{\omega}$, and $\hat{\kappa}$ (from left to right) based on 200 simulations from each of the two SNCP models (corresponding to datasets 1–2) using method B. The true parameter values are indicated by a dashed line in the continuous case or a gray shade in the bar-plots. In the bar-plot for ζ, the bars correspond to $0.1, 0.2, ..., 1$. In the bar-plot for ν, the bars correspond to $2\nu + 1 = 0.1, 0.5, 1, 2, 5$.

distributions of the estimators are shown in Figure 8. The main impression is the same as for method A.

4.4.1 The inhomogeneous case

Consider now datasets 5 and 6, i.e. the simulated realizations of the two inhomogeneous Whittle-Matérn SNCPs in the bottom panels of Figure 3. Since it is difficult from only one realization of X_W to distinguish between the residual process and the intensity function, $\rho = \rho_\psi$ is assumed to depend on a covariate and a parameter ψ: in this case study we let the covariate be the x-coordinate, and

$$\rho_\psi(u) = \exp\left(\psi_0 + \psi_1 \log(a - u_1)\right), \quad u = (u_1, u_2) \in W,$$

where $\psi = (\psi_0, \psi_1)$ and a is the upper right corner of the square $W = [0, a]^2$ (i.e. $a = 3$ for dataset 5 and $a = 2$ for dataset 6). Further, the true parameter values are $\exp(\psi_0) = 400/3$ (dataset 5) or 100 (dataset 6), $\psi = 1$, and where the values of $\theta, \zeta, \omega, \nu, \kappa$ are the same as for dataset 1 in case of dataset 5 and for dataset 2 in case of dataset 6 (see Table 1). Recall that for comparison, the expected number of points in W is the same as in the homogeneous case, namely 1800 for datasets 1 and 5, and 400 for datasets 2 and 6.

For each true model corresponding to datasets 5–6, we have simulated 200 datasets, and estimated the parameters by the automated version of method A. For estimation of ψ we use the maximum composite likelihood estimate based on the intensity function for the SNCP, see Section 4.1. This estimate is derived using the `R` function `ppm` which can be found in the package `spatstat`.

Figure 9 shows the empirical distributions of the estimators of $\theta, \zeta, \omega, \nu, \kappa$. It is not easy to detect any difference between this and the results in Figure 7

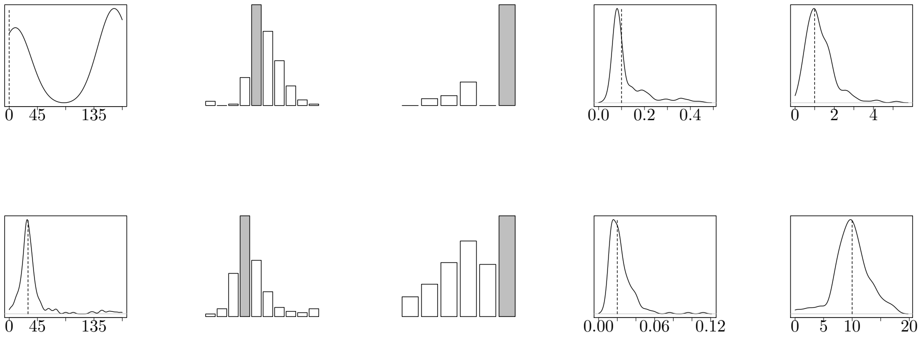


Figure 9: Top and bottom panels: empirical distributions of the estimators $\hat{\theta}$, $\hat{\zeta}$, $\hat{\nu}$, $\hat{\omega}$, and $\hat{\kappa}$ (from left to right) based on 200 simulations from each of the two inhomogeneous SNCP models (corresponding to datasets 5–6) using method A. The true parameter values are indicated by a dashed line in the continuous case or a gray shade in the bar-plots. In the bar-plot for ζ, the bars correspond to $0.1, 0.2, ..., 1$. In the bar-plot for ν, the bars correspond to $2\nu + 1 = 0.1, 0.5, 1, 2, 5$.

for the homogeneous case, although we might expect that the variance of the estimators of $\theta, \zeta, \omega, \nu, \kappa$ would increase, since they now rely on the estimation of ψ.

5 Bayesian inference

Method: Let the setting be as at the beginning of Section 4, and suppose we we will analyze the data $X_W = x$ by a Whittle-Matérn SNCP in a Bayesian setting, with a prior distribution specified as follows.

We assume independent densities $\pi(\theta)$, $\pi(\zeta)$, $\pi(\omega)$, $\pi(\nu)$, and $\pi(\kappa)$. For the intensity function ρ, we consider one of the following cases:

- in the homogeneous case, an independent prior density $\pi(\rho)$;
- in the inhomogeneous case with a non-parametric estimate $\hat{\rho}$ (see Section 4.1), ρ is replaced by $\hat{\rho}$.

(The discussion in the sequel can easily be extended to the inhomogeneous case with a parametric model $\rho = \rho_\psi$ and an independent prior density $\pi(\psi)$.) Specifically,

- θ, ζ, and $2\nu+1$ are uniformly distributed on $[0, \pi)$, $(0, 1]$, $\{0.1, 0.5, 1, 2, 5\}$, respectively;
- ω is inverse gamma distributed with shape parameter α_ω and scale parameter β_ω; if W is rectangular with sides w_1 and w_2, we let $\alpha_\omega = 0.1$ and $\beta_\omega = 0.1 \max\{w_1, w_2\}$ (this implies that the probability that the size of a cluster is larger than the size of W is small);

- κ is gamma distributed with shape parameter α_κ and scale parameter β_κ;
- in the homogeneous case, ρ is gamma distributed with shape parameter α_ρ and scale parameter β_ρ.

Denote, in the homogeneous case, $\Theta = (\theta, \zeta, \omega, \nu, \kappa, \rho)$ with density

$$\pi(\Theta) = \pi(\theta)\pi(\zeta)\pi(\omega)\pi(\nu)\pi(\kappa)\pi(\rho),$$

and in the inhomogeneous case, $\Theta = (\theta, \zeta, \omega, \nu, \kappa)$ with density

$$\pi(\Theta) = \pi(\theta)\pi(\zeta)\pi(\omega)\pi(\nu)\pi(\kappa).$$

We also assume that conditional on κ, Φ is independent of the remaining parameters in Θ. As discussed in Appendix A, the conditional distribution of X_W given Φ can be well approximated by replacing the infinite process Φ with a finite subprocess $\Phi_{W_{\text{ext}}}$. Here W_{ext} is a bounded subset of $\mathbb{R}^2$ containing W, where using the notation and Poisson cluster interpretation from Section 3.3.2, the probability q_W that some cluster X_v with $v \in \Phi \setminus W_{\text{ext}}$ intersects W is very small; e.g. we may require that $q_W < 0.01$. The observation model is then approximated by the density for a Poisson process on W,

$$\pi(x|\Phi_{W_{\text{ext}}}, \Theta) = \exp\left(-\int_W \lambda_{\text{approx}}(u)\,\mathrm{d}u\right) \prod_{u \in x} \lambda_{\text{approx}}(u)$$

where the intensity function λ_{approx} is given by (12) when Φ in (16) is replaced by $\Phi_{W_{\text{ext}}}$, and $f_0 = k_v$ in (17) is given by (13). Moreover, $\Phi_{W_{\text{ext}}}$ has density

$$\pi(\Phi_{W_{\text{ext}}}|\kappa) = \exp\left(-\kappa|W_{\text{ext}}|\right) \kappa^{\#\Phi_{W_{\text{ext}}}}.$$

Thus our target distribution is the approximate posterior distribution for $(\Theta, \Phi_{W_{\text{ext}}})$ with density

$$\pi(\Theta, \Phi_{W_{\text{ext}}}|x) \propto \pi(x|\Phi_{W_{\text{ext}}}, \Theta)\pi(\Phi_{W_{\text{ext}}}|\kappa)\pi(\Theta).$$

We use a Metropolis within Gibbs algorithm for simulating from this target distribution (see e.g. Gilks, Richardson & Spiegelhalter (1996)), with standard Metropolis random walk updates for each of the components in Θ, and using the algorithm in Appendix B to simulate the 'latent data' $\Phi_{W_{\text{ext}}}$ with 'full conditional'

$$\pi(\Phi_{W_{\text{ext}}}|x, \Theta) \propto \pi(x|\Phi_{W_{\text{ext}}}, \Theta)\pi(\Phi_{W_{\text{ext}}}|\kappa). \tag{40}$$

Remarks: For the anisotropic Thomas process studied in Castelloe (1998) a different MCMC algorithm was based on the Poisson cluster process interpretation (Section 3.3.1), i.e. the algorithm incorporated not only the cluster centers but also the cluster relationships as latent data. The amount of latent data is much smaller in our algorithm above, making it a faster algorithm when interest is in estimating the parameters of the model, the cluster centers, and thereby the

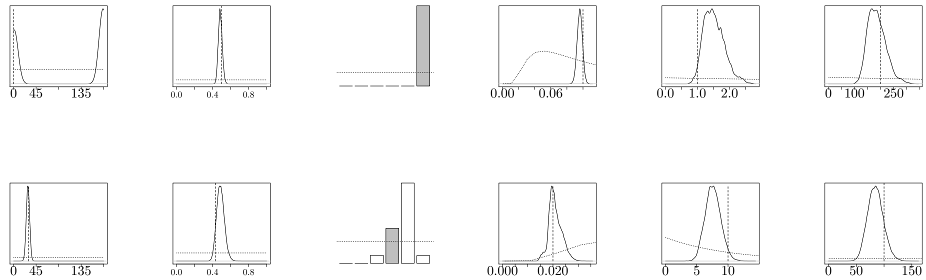


Figure 10: Top and bottom panels: MCMC estimates of the posterior densities (solid lines) and prior densities (dashed lines) of the parameters for the stationary Whittle-Matérn SNCPs corresponding to datasets 1–2. From left to right: θ, ζ, ν, ω, κ, ρ. The true value of the parameter is indicated by a vertical dotted line in the continuous case or a gray shade in the bar-plots.

residual process and the random intensity function $\lambda \approx \lambda_{\text{approx}}$ on W. However, still rather long runs are needed for the data examples considered later in this paper, where we have monitored the mixing properties of our algorithm mainly by trace plots of the parameters and characteristics of $\Phi_{W_{\text{ext}}}$, using different starting points (due to space limitations such plots are omitted).

In the LGCP case, the model fits the framework of integrated nested Laplace approximations (INLA; Rue, Martino & Chopin (2009) and Simpson, Illian, Lindgren, Sørbye & Rue (2011)). However, at the present time it is not possible to do the Bayesian computations for our model using functions available in the R-INLA package (it would require a large amount of time and effort to implement it).

Example: We now consider our Bayesian setting for datasets 1 and 2, modelled by the stationary Whittle-Matérn SNCPs and with prior parameters $\alpha_\kappa = 1$, $\beta_\kappa = 10$, $\alpha_\rho = 1$, and $\beta_\rho = 1000$. For posterior simulation using the Metropolis with Gibbs algorithm, a burn in period of 10000 steps is used, and the inference is based on 90000 consecutive steps of two independent Markov chains (the two chains were also used to determine the burn in period).

Figure 10 shows density estimates of the posteriors of the parameters corresponding to dataset 1 and 2, respectively. For comparison we have also included the prior densities. The priors are very flat and they do not seem to affect the posteriors significantly.

The MCMC estimate of the posterior mean of the random intensity function $\lambda(u) = \rho S(u)$ evaluted at a fine grid of points $u \in W$ is shown in a gray scale plot in Figure 11. As expected the high intensity areas correspond to high consentration of points in the data, cf. the mid panels in Figure 3.

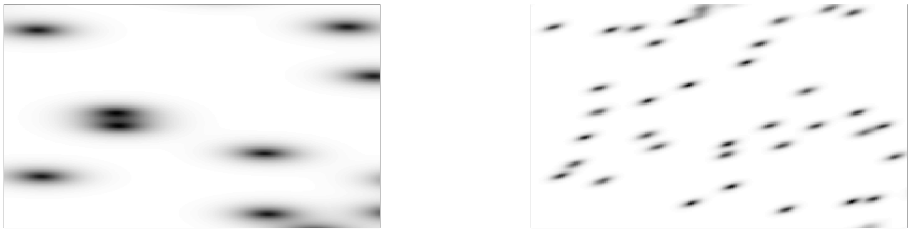


Figure 11: Estimated posterior mean of $\lambda(u) = \rho S(u),\ u \in W$ corresponding to dataset 1 (left panel) and to dataset 2 (right panel).

6 Analysis of real data examples

In this section we model the anisotropic spatial point pattern datasets in Figures 1–2 by a Whittle-Matérn SNCP or LGCP in order to illustrate the application of our methods in Sections 4–5. A challenge is to decide how much the anisotropy is considered to be a result of inhomogeneity as modelled by ρ and how much is due to the residual process S.

6.1 Chapels in Welsh valleys

The locations of the 110 chapels in Figure 1 were extracted by the authors of Mugglestone & Renshaw (1996) from an Ordnance Survey map and then rescaled to the unit square, so in our notation $W = [0, 1]^2$. Mugglestone & Renshaw (1996) analyzed this dataset by a spectral analysis and by Ripley's K-function, and they noticed a large-scale regularity due to four regularly spaced parallel valleys, a clustering at a smaller scale, and a very strong directional effect corresponding to the direction of the valleys. They rejected a stationary Poisson process model but did not study an alternative parametric spatial point process model.

If we had been provided the information about population density or elevation in W, the population density or elevation could had been incorporated as a covariate in a parametric model for ρ, and possibly this could had 'explained (most of) the anisotropy'. Lacking this information, we follow Mugglestone & Renshaw (1996) in assuming stationarity, i.e. ρ is constant. We furthermore assume a Whittle-Matérn SNCP model, where the cluster centres could correspond to places with a high population density, e.g. town centres.

First, considering parameter estimation based on methods A and B from Section 4, the results are found in Table 2. The two methods produce similar parameter estimates, where the estimate of $\theta = 124°$ or $113°$ corresponds to the orientation of the valleys, and the low estimated value of $\zeta = 0.3$ reflects the strong anisotropy. Figure 12 shows a realization from the fitted model using method A. The realizations suggest that the model has captured much of the structure in the dataset. We have performed more formal model checking by use of the empty space function F and the nearest neighbour distribution

	θ	ζ	ω	ν	κ	W
Chapels						$[1,1]^2$
Method A (SNCP)	124	0.3	0.04	-0.25	21	
Method B (SNCP)	113	0.3	0.06	-0.45	18	
Earthquakes						$[8,5]^2$
Method A (SNCP)	158	0.6	0.13	-0.45	1.59	
Method B (SNCP)	152	0.6	0.12	-0.45	0.31	
Method A (LGCP)	158	0.6	0.7	0.25	0.06	

Table 2: Estimated parameters for the chapels dataset modelled by a stationary Whittle-Matérn SNCP model and the earthquakes datasets modelled by an inhomogeneous Whittle-Matérn SNCP or LGCP model.

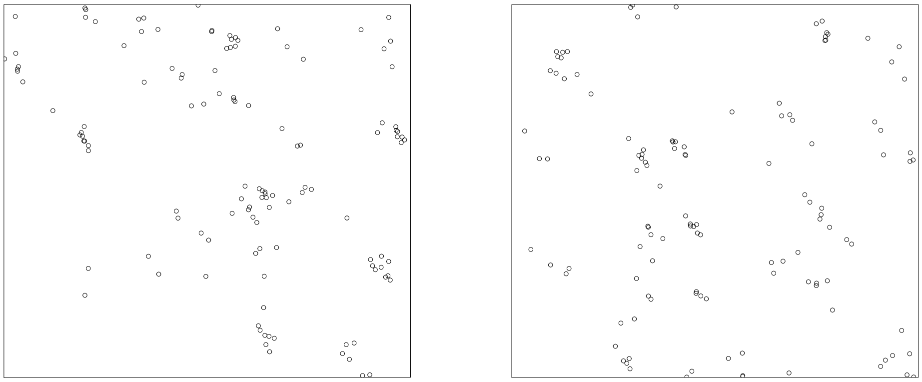


Figure 12: Two realizations from the Whittle-Matérn SNCP model estimated by method A and based on the chapels dataset.

function G (see e.g. Møller & Waagepetersen (2004)). Note that F and G have not been used for estimating the parameters, and non-parametric estimates are found using the `R` functions `Fest` and `Gest` available in the package `spatstat`. Figure 13 shows the estimates $\hat{F}$ and $\hat{G}$ together with pointwise 95% envelopes based on 39 realizations of the estimated model, i.e. the curves given by the minimum respective the maximum of the 39 simulated estimates at each distance r. We see that $\hat{G}(r)$ is within the 95% envelopes, while $\hat{F}(r)$ falls slightly outside the 95% envelopes at a few and mostly very small distances r. Overall the model checking suggests that the Whittle-Matérn SNCP model estimated by method A fits reasonably well.

Next, using the Bayesian setting in Section 5, we let $\alpha_\kappa = 1$, $\beta_\kappa = 10$, $\alpha_\rho = 1$, and $\beta_\rho = 1000$. Figure 14 shows MCMC estimates of posterior densities and corresponding prior densities for each parameter (as the prior densities for ω and ρ are very flat, they are not so easy to see in the plots). For most of the parameters the posterior distributions agree well with the estimated values that we obtained above, in the sense that the estimated values lie in the high

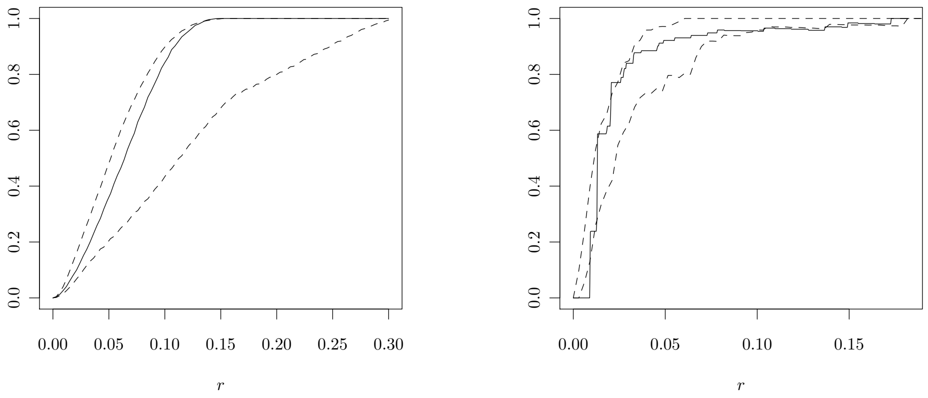


Figure 13: Non-parametric estimates (solid lines) of the nearest-neighbour function (left panel) and the empty space function (right panel) for the chapels dataset. The dashed lines are estimated pointwise 95% envelopes obtained by simulations from the fitted Whittle-Matérn SNCP model using method A.

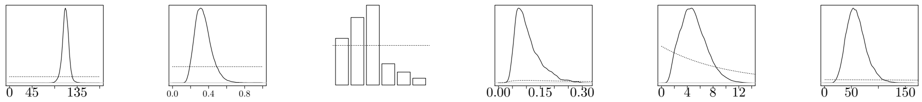


Figure 14: MCMC estimates of the posterior densities (solid lines) and prior densities (dashed lines) of the parameters for the stationary Whittle-Matérn SNCP used for modelling the chapels dataset. From left to right: θ, ζ, ν, ω, κ, ρ.

probability area of the posterior distributions. The prior specification does not seem to affect these posterior distributions significantly. Moreover, Figure 15 shows the estimated posterior mean of the random intensity function $\lambda(u) = \rho S(u)$. Most of the data points lie in areas where $\lambda(u)$ is high, but a few points lie in areas where $\lambda(u)$ is close to 0.

6.2 Earthquake dataset

We now turn to the epicentral locations in Figure 2, previously analyzed in Veen & Schoenberg (2006). Lacking any covariate information for the epicentral locations, we estimate the intensity ρ of this point pattern by the non-parametric kernel estimator given by (23). The subjective choice of bandwidth is a challenge, and depends on how much of the inhomogeneity we consider to be a result of ρ and how much is a result of the residual process S. We decided on a bandwidth of 1.5 after some experimentation. The result is shown in Figure 16.

Veen & Schoenberg (2006) did not attempt to model the anisotropy seen in Figure 2. Parameter estimates obtained by methods A and B using an inhomogeneous Whittle-Matérn SNCP and by method A using an inhomogeneous

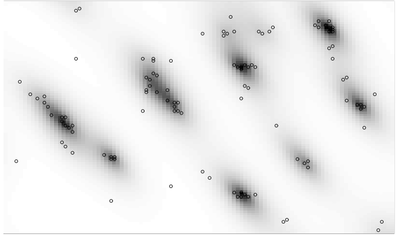


Figure 15: Gray scale plot of the estimated posterior mean of $\lambda(u)$ together with the locations of chapels in the Welsh valleys dataset.

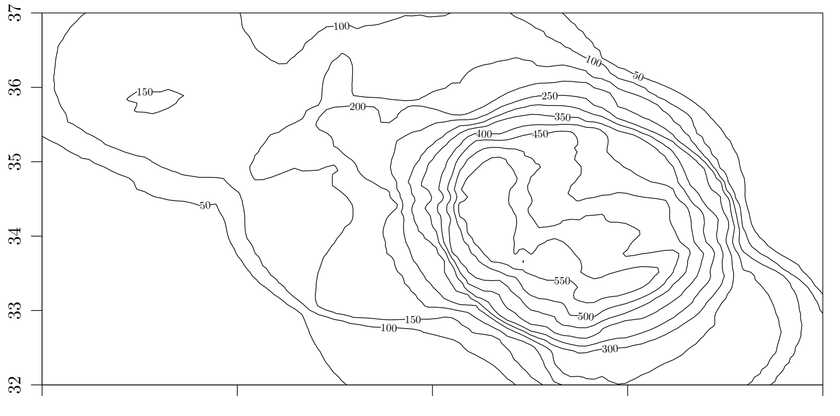


Figure 16: A contour plot of the non-parametric intensity estimate for the earthquake dataset.

Whittle-Matérn LGCP are found in Table 2. For the SNCP, the two methods produce similar estimates for all parameters except κ. The estimate of ζ is equal to 0.6 in all three cases, and the direction of the point pattern is estimated to be 152° or 158°.

Finally, we have considered a number of realizations from each of the estimated models obtained by method A. Figure 17 shows one example. We see that the SNCP model is able to reproduce the shape and size of many of the clusters, but there are too few points far away from the clusters. This is to a large degree not a problem with the LGCP model, which seems to provide a better fit.

Appendix A

Suppose that X is a SNCP as in Section 3.3, and we want to simulate $X_W = X \cap W$, where $W \subset \mathbb{R}^2$ is a bounded Borel set. A simulation procedure in the stationary case with an isotropic pair correlation function g was discussed in Brix & Kendall (2002), Møller (2003), and Møller & Waagepetersen (2004). Below we briefly extends this to first the inhomogeneous case of the intensity function ρ and second the elliptical case of g.

First, assume f is isotropic and ρ is bounded on W by a positive constant $\rho_{\max}$. We can obtain a simulation of X_W by first simulating a stationary SNCP $X_{\max}$ within W and driven by $\lambda_{\max} = \rho_{\max} S$, and second make an independent thinning of $X_{\max,W}$ where the retention probability of a point $u \in X_{\max,W} \cap W$ is given by $p(u) = \rho(u)/\rho_{\max}$. This can be modified to a more efficient simulation procedure if ρ will be far away from $\rho_{\max}$ on W. Then we consider a finite subdivision $W = \cup_{i \in I} C_i$ so that ρ on each set C_i is close to an upper bound ρ_i, and then exploit the fact that conditional on Φ, the Poisson processes $X \cap C_i$, $i \in I$, are independent.

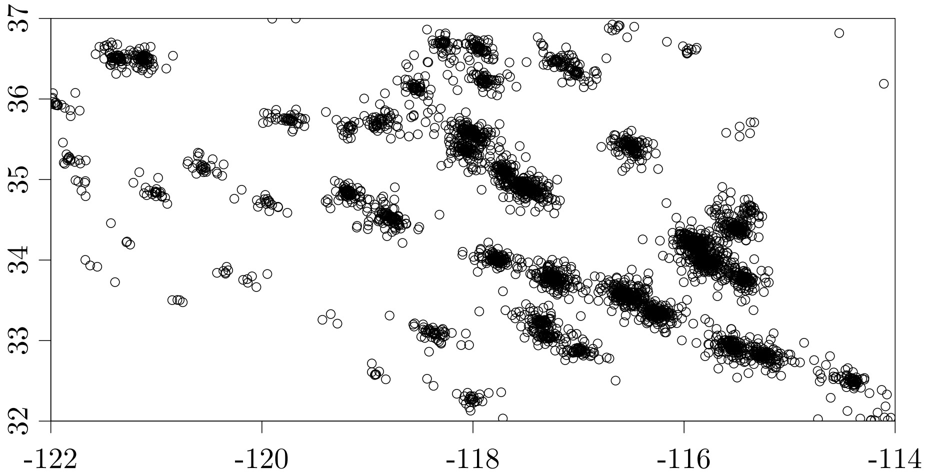
How to make simulation therefore boils down to how to simulate $X_{\max,W}$. Consider a bounded Borel set $W_{\text{ext}} \supseteq W$ which is supposed to be so large that $X_{\max} \cap W$ is well approximated by replacing the infinite Poisson process Φ by the finite Poisson process $\Phi_{W_{\text{ext}}}$ in the simulation. The approximation may be evaluated by calculating the probability q_W that some cluster of $X_{\max}$ with its centre outside W_{ext} intersects W. Let

$$p_W(v) = 1 - \exp\left(-(\rho_{\max}/\kappa) \int_W f(u - v)\, \mathrm{d}u \right)$$

be the probability that a cluster of $X_{\max}$ with center v intersects W. As shown in Møller (2003),

$$q_W = 1 - \exp\left(-\kappa \int_{\mathbb{R}^2 \setminus W_{\text{ext}}} p_W(v)\, \mathrm{d}v \right).$$

If W_{ext} is sufficiently large, it may be possible to obtain a useful upper bound on q_W by finding a sufficiently small Borel set $E \subset \mathbb{R}^2$ with $W \subseteq E \subseteq W_{\text{ext}}$ and



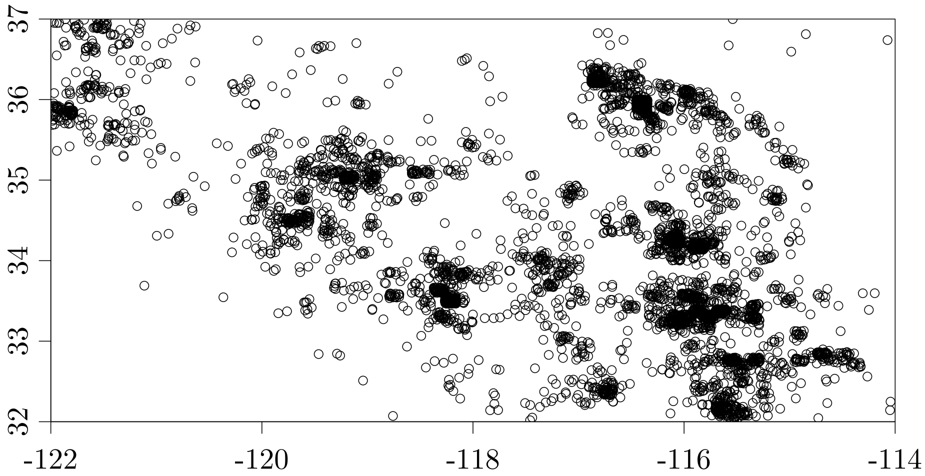


Figure 17: One realization from the inhomogeneous Whittle-Matérn SNCP model and one realization from the inhomogeneous Whittle-Matérn LGCP model estimated by method A and based on the earthquake dataset.

a sufficiently small non-negative Borel function j such that

$$j(v) \geq f(u-v) \quad \text{whenever } v \in \mathbb{R}^2 \setminus W_{\mathrm{ext}} \text{ and } u \in E \tag{41}$$

and both the integral

$$i(v) = j(v)|E| \tag{42}$$

and the upper bound

$$q_W \leq 1 - \exp\left(-\kappa \int_{\mathbb{R}^2 \setminus W_{\mathrm{ext}}} [1 - \exp(-(\rho_{\max}/\kappa) i(v))] \, \mathrm{d}v\right) \tag{43}$$

can easily be evaluated.

Next, let $f = f_\Sigma$ be elliptical, see (17). For specificity, suppose that $f_0(r)$ is a decreasing function for $r \geq 0$ (e.g. this is the case when $f_0 = k_\nu$, i.e. when X is a Whittle-Matérn SNCP), and W is a rectangle with centre at the origin and sides $a > 0$ and $b > 0$. We choose $R > 0$ such that $E = \{u : u\Sigma^{-1}u^t \leq R^2\}$ is the smallest elliptical region containing W and agreeing with the level curves of f_Σ, i.e. $R^2 = (-a/2, b/2)\Sigma^{-1}(-a/2, b/2)$ if $\theta \leq \pi/2$, and $R^2 = (a/2, b/2)\Sigma^{-1}(a/2, b/2)$ if $\theta > \pi/2$. Further, let $W_{\mathrm{ext}} = \{u : u\Sigma^{-1}u^t \leq (R+r)^2\}$, where $r > 0$ determines the error of the approximate simulation procedure, and let $A = \Sigma^{-1/2}$, so $1/|A| = \zeta\omega^2$. For any $u \in E$ and $v \in \mathbb{R}^2 \setminus W_{\mathrm{ext}}$, we have $\|uA\| \leq R$ and $\|vA\| > R + r$, so $\|(u-v)A\| \leq \|vA\| - R$ and hence (41) is satisfied when

$$j(v) = f_0((\|vA\| - R)^2)|A|.$$

Finally, (42) becomes

$$i(v) = f_0((\|vA\| - R)^2)$$

and so (43) becomes

$$q_W \leq 1 - \exp\left(-2\pi\kappa\zeta\omega^2 \int_{R+r}^{\infty} \left[1 - \exp\left(-(\rho_{\max}/\kappa)) f_0((s-R)^2)\right)\right] s \, \mathrm{d}s\right) \tag{44}$$

where the integral may be computed by numerical methods.

Appendix B

An MCMC algorithm very similar to the Metropolis-Hastings algorithm described in Møller (2003) will be used for simulation from (40). Suppose $\phi = \{v_1, ..., v_n\} \subset W_{\mathrm{ext}}$ is the current step of the Markov chain. Then with equal probabilities p we make either a birth or a death proposal, and with probability $1 - 2p$ we propose to move one of the points. The moving alternative is where this algorithm is different from the one described in Møller (2003). Define

$$r(\phi, v) = \frac{\pi(\phi \cup \{v\}|x)|W_{\mathrm{ext}}|}{\pi(\phi|x)(n+1)}, \quad u \in W_{\mathrm{ext}}.$$

If a birth is proposed, a point v is generated from a uniform distribution on W_{ext}, and with probability $\min\{1, r(\phi, v)\}$ the next state of the chain is $\phi \cup \{v\}$, and otherwise ϕ is retained. If a death is proposed, when $n > 0$ the next step of the chain is $\phi \setminus \{v_i\}$ with probability $\min\{1, 1/r(\phi \setminus \{v_i\}, v_i)\}/n$, $i = 1, \ldots, n$, while if $n = 0$ we do nothing. If a move is proposed, a point v is generated from a uniform distribution on W_{ext}, and the next state of the chain is $\Phi \setminus \{v_i\} \cup \{v_i\}$ with probability $\min\{1, r(\Phi \setminus \{v_i\}, v_i)/r(\phi \setminus \{v_i\}, v)\}/n$, $i = 1, ..., n$, and otherwise Φ is retained. Note that

$$\frac{\pi(\phi \cup \{v\}|x)}{\pi(\phi|x)} = \kappa \exp\left(-(1/\kappa)\int_W \rho(u) f(u-v)\,\mathrm{d}u\right) \prod_{u \in x}\left[1 + \frac{f(u-v)}{\sum_{i=1}^n f(u-v_i)}\right].$$

This expression contains an integral, which is computed by numerical methods.

Acknowledgment

Supported by the Danish Council for Independent Research — Natural Sciences, grant 09-072331, "Point process modelling and statistical inference", and by the Centre for Stochastic Geometry and Advanced Bioimaging, funded by a grant from the Villum Foundation. We thank Moira Mugglestone for providing the dataset in Figure 1, and The Southern California Earthquake Center for making the dataset in Figure 2 freely available on www.data.scec.org.

References

Abramowitz, M. & Stegun, I. (1964). *Handbook of Mathematical Functions with Formulas, Graphs and Mathematical Tables*, Dover.

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

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

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

Castelloe, J. (1998). Issues in reversible jump Markov chain Monte Carlo and composite EM analysis, applied to spatial Poisson cluster processes. Ph.D. thesis, University of Iowa.

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

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

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

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

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

Diggle, P. J. (2010). Nonparametric methods, *in* A. E. Gelfand, P. J. Diggle, P. Guttorp & M. Fuentes (eds), *Handbook of Spatial Statistics*, CRC Press, Boca Raton, pp. 299–337.

Gilks, W. R., Richardson, S. & Spiegelhalter, D. J. (1996). *Markov Chain Monte Carlo in Practice*, Chapman & Hall, London.

Guan, Y., Sherman, M. & Calvin, J. (2006). Assessing isotropy for spatial point processes, *Biometrics* **62**: 119–125.

Guttorp, P. & Gneiting, T. (2006). On the Matérn correlation family, *Biometrika* **93**: 989–995.

Guttorp, P. & Gneiting, T. (2010). Continuous parameter stochastic process theory, *in* A. E. Gelfand, P. Diggle, M. Fuentes & P. Guttorp (eds), *A Handbook of Spatial Statistics*, Chapmand and Hall/CRC Press, pp. 17–28.

Illian, J., Penttinen, A., Stoyan, H. & Stoyan, D. (2008). *Statistical Analysis and Modelling of Spatial Point Patterns*, John Wiley and Sons, Chichester.

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

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

Møller, J. & Waagepetersen, R. P. (2004). *Statistical Inference and Simulation for Spatial Point Processes*, Chapman and Hall/CRC, Boca Raton.

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

Mugglestone, M. A. & Renshaw, E. (1996). A practical guide to the spectral analysis of spatial point processes, *Computational Statistics and Data Analysis* **21**: 43–65.

Ohser, J. & Stoyan, D. (1981). On the second-order and orientation analysis of planar stationary point processes, *Biometrical Journal* **23**: 523–533.

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

Rosenberg, M. (2004). Wavelet analysis for detecting anisotropy in point patterns, *Journal of Vegetation Science* **15**: 277–284.

Rue, H., Martino, S. & Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations (with discussion), *Journal of Royal Statistical Society: Series B (Statistical Methodology)* **71**: 319–392.

Simpson, D., Illian, J., Lindgren, F., Sørbye, S. & Rue, H. (2011). Going off grid: Computationally efficient inference for log-Gaussian Cox processes. Preprint Statistics No. 10/2011, Norwegian University of Science and Technology.

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

Stoyan, D. & Stoyan, H. (1994). *Fractals, Random Shapes and Point Fields*, Wiley, Chichester.

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

Tanaka, U., Ogata, Y. & Stoyan, D. (2008). Parameter estimation and model selection for Neyman-Scott point processes, *Biometrical Journal* **50**: 43–57.

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

Veen, A. & Schoenberg, F. P. (2006). Assessing Spatial Point Process Models Using Weighted K-functions: Analysis of California Earthquakes, *Case Studies in Spatial Point Process Modeling* pp. 293–306.