# Log Gaussian Cox processes

by

Jesper Møller, Anne Randi Syversveen, and Rasmus Plenge Waagepetersen.

# Log Gaussian Cox processes

JESPER MØLLER Aalborg University ANNE RANDI SYVERSVEEN The Norwegian University of Science and Technology RASMUS PLENGE WAAGEPETERSEN University of Aarhus

ABSTRACT. Planar Cox processes directed by a log Gaussian intensity process are investigated in the univariate and multivariate cases. The appealing properties of such models are demonstrated theoretically as well as through data examples and simulations. In particular, the first, second and third-order properties are studied and utilized in the statistical analysis of clustered point patterns. Also empirical Bayesian inference for the underlying intensity surface is considered.

*Key words*: empirical Bayesian inference; ergodicity; Markov chain Monte Carlo; Metropolis-adjusted Langevin algorithm; multivariate Cox processes; Neyman-Scott processes; pair correlation function; parameter estimation; spatial point processes; third-order properties.

AMS 1991 subject classification: Primary 60G55, 62M30. Secondary 60D05.

# **1** Introduction

Cox processes provide useful and frequently applied models for aggregated spatial point patterns where the aggregation is due to a stochastic environmental heterogeneity, see e.g. Diggle (1983), Cressie (1993), Stoyan *et al.* (1995), and the references therein. A Cox process is 'doubly stochastic' as it arises as an inhomogeneous Poisson process with a random intensity measure. The random intensity measure is often specified by a random intensity function or as we prefer to call it an intensity process or surface.

There may indeed be other sources of aggregation in a spatial point pattern than spatial heterogeneity. Cluster processes is a well-known class of models where clusters are generated by an unseen point process, cf. the references mentioned above. The class of nearest-neighbour Markov point processes (Baddeley and Møller, 1989) include many specific models of cluster processes (Baddeley *et al.*, 1996) as well as other types of processes with clustering modelled by 'interaction functions' (Møller, 1994) such as the penetrable sphere model (Widom and Rowlinson, 1970; Baddeley and Van Lieshout, 1995) and the continuum random cluster model (Møller, 1994; Häggström *et al.*, 1996).

In this paper we consider log Gaussian Cox processes, i.e. Cox processes where the logarithm of the intensity surface is a Gaussian process. We show that the class of stationary log Gaussian Cox processes possesses various appealing properties: (i) The distribution is completely characterized by the intensity and the pair correlation function of the Cox process. This makes parametric models easy to interpret and simple methods are available for parameter estimation and model checking. (ii) Theoretical properties are easily derived. Higher-order properties are for instance simply expressed by the intensity and the pair correlation function of the log Gaussian Cox process. Thereby summary statistics based on e.g. the third-order properties can be constructed and estimated. (iii) The underlying Gaussian process and intensity surface can be predicted from a realization of a log Gaussian Cox process observed within a bounded window using Bayesian methods. (vi) There is no problem with edge effects as the distribution of a log Gaussian Cox process restricted to a bounded subset is known.

The properties (i)-(vi) are rather characteristic for Log Gaussian Cox processes. We shall further note that log Gaussian Cox processes are flexible models for clustering, easy to simulate, and that the definition of univariate log Gaussian Cox processes can be extended in a natural way to multivariate log Gaussian Cox processes.

Other transformations than the exponential of the Gaussian process may be considered as well, and in particular  $\chi^2$  Cox processes (as defined in Section 3) may be of interest.

During the final preparation of this paper we realized that a definition of Log Gaussian Cox processes has previously been given in Rathbun and Cressie (1994), but they restrict attention to the case where the intensity is constant within square quadrats and modelled by a conditional autoregression (Besag, 1974). The advantage of these discretized models is mainly that they can easily be explored using Gibbs sampling, but as noticed by Rathbun and Cressie (1994) such models does not converge to anything reasonable as the sides of the quadrats tend to zero. Consequently, it is difficult to investigate the correlation structure of these Gaussian random field models through those summary statistics which are usually estimated for a point pattern such as the pair correlation function. The log Gaussian Cox processes studied in the present paper are in contrast to this specified by such characteristics, and discretized versions of our log Gaussian Cox processes can be simulated exactly without any problem with edge effects. Also the Metropolis-adjusted Langevin algorithm (Besag, 1994; Roberts and Tweedie, 1997) for simulating from the posterior of the intensity surface as studied in Section 8 is both easy to specify and implement. In contrast and even if we ignore the problem with edge effects, Gibbs sampling from the posterior becomes straightforward only if one uses conditional autoregression priors.

The paper is organized as follows. In Section 2 we give a formal definition of univariate log Gaussian Cox processes and inspect some of their properties by simulations. Theoretical results are established in Section 3. In Section 4 we compare log Gaussian Cox processes with the class of Neyman-Scott processes with a Poisson distributed number of offspring. Extensions of log Gaussian Cox processes to the multivariate case are studied in Section 5. In Section 6 we describe different simulation procedures. Section 7 is concerned with parameter estimation and model checking of parametric models for log Gaussian Cox processes. We illustrate this by considering some data sets of univariate and bivariate point patterns. Finally, in Section 8 we discuss how empirical Bayesian methods may be used for the purpose of predicting the unobserved Gaussian process and intensity surface.

# 2 Univariate log Gaussian Cox processes

For specificity and since all the examples presented later on are planar we specify the model in  $\mathbb{R}^2$ , but our model can be completely similar defined in  $\mathbb{R}^d$ , d = 1, 2, ...

Briefly, by a planar spatial point process we shall understand a locally finite random subset X of the plane  $\mathbb{R}^2$ . This is said to be a Cox process directed by a random intensity process  $\Lambda = \{\Lambda(s) : s \in \mathbb{R}^2\}$  if the conditional distribution of X given  $\Lambda$  is a Poisson process with intensity function  $\Lambda(\cdot)$ , i.e. when for bounded Borel sets  $B \subset \mathbb{R}^2$  we have conditional on  $\Lambda$  that  $\operatorname{card}(X \cap B)$  is Poisson distributed with a mean  $\int_B \Lambda(s) ds$  which is assumed to be nonnegative and finite. We restrict attention to the case where  $\Lambda$  and hence X is stationary and sometimes also isotropic, i.e. when the distribution of  $\Lambda$  is invariant under translations and possibly also rotations in  $\mathbb{R}^2$ . The intensity

$$\rho = E\Lambda(s)$$

is henceforth assumed to be strictly positive and finite.

Throughout this paper we model the intensity process by a log Gaussian process:

$$\Lambda(s) = \exp\left\{Y(s)\right\} \tag{1}$$

where  $Y = \{Y(s) : s \in \mathbb{R}^2\}$  is a real-valued Gaussian process (i.e., the joint distribution of any finite vector  $(Y(s_1), \ldots, Y(s_n))$  is Gaussian). It is necessary to impose conditions on Y so that the random mean measure  $\nu$  given by  $\nu(B) = \int_B \Lambda(s) ds$  for bounded Borel sets  $B \subset \mathbb{R}^2$ , becomes well-defined. First it is of course required that the realizations of  $\Lambda$  are integrable almost surely. But further conditions are required in order that  $\nu$  is uniquely determined by the distribution of Y. Here we impose the natural condition that  $\nu$  is given in terms of a continuous modification of Y. Then  $\nu$  is uniquely determined, since all the continuous modifications are indistinguishable (i.e. their realizations are identical with probability one), and it also follows that  $\nu(B) < \infty$  for bounded B.

By stationarity, the distribution of Y and hence X is specified by the mean  $\mu = EY(s)$ , the variance  $\sigma^2 = Var(Y(s))$ , and the correlation function  $r(s_1 - s_2) = Cov(Y(s_1), Y(s_2))/\sigma^2$  of Y. The model is only well-defined for positive semi-definite correlation functions, i.e. when  $\sum_{i,j} a_i a_j r(s_i - s_j) \ge 0$  for all  $a_1, \ldots, a_n \in \mathbb{R}$ ,  $s_1, \ldots, s_n \in \mathbb{R}^2$ ,  $n = 1, 2, \ldots$ . Whether a given function is positive semi-definite may best be answered through a spectral analysis, see e.g. Christakos (1984), Wackernagel (1995), and the references therein. Furthermore, if there exist  $\epsilon > 0$  and K > 0 such that  $1 - r(s) < (-\log ||s||)^{-(1+\epsilon)}$  for all s with ||s|| < 1, then the existence of an almost surely continuous modification is quaranteed (Adler, 1981, page 60). A stronger condition, which in our experience is easier to check, is given by that  $1 - r(s) < \tilde{K} ||s||^{\alpha}$  for some  $\tilde{K} > 0$  and  $\alpha > 0$ .

The parameters  $e^{\mu} > 0$  and  $\sigma > 0$  have a clear interpretation as a scale and shape parameter, respectively, since we can write  $\Lambda_{\mu,\sigma} \stackrel{\mathcal{D}}{=} e^{\mu} \Lambda_{0,1}^{\sigma}$ , where  $\ln \Lambda_{\mu,\sigma}$  is a stationary Gaussian process with mean  $\mu$ , variance  $\sigma^2$ , and correlation function  $r(\cdot)$ . The homogeneous Poisson process may be considered as the limit of a log Gaussian Cox process as  $\sigma$  tends to 0. Another extreme case is  $r(\cdot) = 1$ , whereby we obtain a mixed Poisson process with a randomized intensity  $\Lambda(\cdot) = \lambda$  which is log Gaussian distributed.

If the distribution of Y is invariant under rotations, r(s) = r(||s||) depends only on s through its length ||s||, so the correlation function is invariant under reflections too. Consequently, invariance under translations and rotations implies that the joint distribution of (X, Y) is invariant under rigid motions  $\phi$  in  $\mathbb{R}^2$ :  $(\phi(X), Y(\phi(\cdot))) \stackrel{\mathcal{D}}{=} (X, Y)$ .

Examples of isotropic correlation models are listed in Table 1. The condition for existence of an almost surely continuous modification holds for all of these correlation functions which furthermore all tend to 0 at infinity. Notice that the correlation models are parametrized by a scale parameter  $\beta$  so that the three types of processes (Gaussian, intensity, and Cox) are all parametrized by  $(\mu, \sigma, \beta) \in (-\infty, +\infty) \times (0, +\infty) \times (0, +\infty)$ . The "scale" of the parameter  $\beta$  is with respect to locations: with obvious notation,  $\Lambda_{\mu,\sigma,\beta} \stackrel{\mathcal{D}}{=} \{e^{\mu}\Lambda_{0,1,1}(s/\beta)^{\sigma} : s \in \mathbb{R}^2\}.$ 

The first four models in Table 1 represent well the correlation structures which can be achieved by using the correlation models in this table, so we have restricted attention to these four models in the following Fig.'s 1–3. In Theorem 1, Section 3, it is shown that the corresponding pair correlation functions are given by the exponential to the covariance function  $\sigma^2 r(\cdot)$ . These pair correlation functions are plotted in Fig. 1 a)-d) for various values of  $\beta$  when  $\sigma = 1$ . If one 'standardizes' the pair correlation function  $g(\cdot)$  to g(0) = e or equivalently takes  $\sigma = 1$ , plots of corresponding pair correlation and covariance functions look very similar, cf. Fig. 1, a)-b) and g)-h).

Simulated realizations of Gaussian processes on the unit square with correlation structure given by the four different types of correlation functions are shown in Fig. 2. Fig. 3 shows simulations of the corresponding log Gaussian Cox processes. The parameters in the first row in Fig. 3 are the same as in Fig. 2. In order to facilitate comparison of the four different Cox processes the  $\beta$ -values in Fig. 3 are chosen so that the mean and variance of the number of points are equal for all Cox processes in the same row. By Theorem 1, Section 3, the mean is  $\rho = \exp(\mu + \sigma^2/2)$  and the variance is given by

$$Var\left(\operatorname{card}\left(X \cap [0,1]^2\right)\right) = \rho + \rho^2 \left(\int_{[0,1]^2} \int_{[0,1]^2} e^{\sigma^2 r(s-t)} ds dt - 1\right).$$

The variance thus increases when  $\beta$  and hence the correlation increases. It is difficult to compare unconditional simulations when the variance of the number of points is large and the simulations in Fig.'s 2 and 3 are therefore performed conditional on that the number n of points equals the mean number of points ( $n = \rho = 148$ ).

In the upper row in Fig. 3 a moderate value of  $\sigma$  but large values of  $\beta$  give rise to large but not dense clusters of points. In the lower row moderate values of  $\beta$  but a higher value of  $\sigma$  lead to many but small clusters. In the middle row a high value of  $\sigma$ and intermediate values of  $\beta$  are used, and compared to the lowest row fewer but larger clusters appear. The realizations of the 'Gaussian' and 'cardinal sine' log Gaussian Cox processes are visually quite similar. The 'stable' log Gaussian Cox process is in general less clustered than the other processes. This is not surprising because the Gaussian random field with the stable correlation function is not very peaked except at the small scale, c.f. Fig. 2.

Finally, it should be noted that Cox processes may be extended to models on bounded regions  $W \subset \mathbb{R}^2$ , where the conditional distribution of  $X_W = X \cap W$  given  $\Lambda_W = \{\Lambda(s) : s \in W\}$  is of a Gibbsian type. For instance, consider a conditional distribution of  $X_W$  given  $\Lambda_W = \lambda_W$  with density

$$f(x_W|\lambda_W) = c(\lambda_W) \left\{ \prod_{i=1}^n \lambda(x_i) \right\} \left\{ \prod_{1 \le i < j \le n} \phi(||x_i - x_j||) \right\}$$

with respect to a unit rate Poisson point process and for  $x_W = \{x_1, \ldots, x_n\} \subset W$ , where e.g.  $\lambda_W$  models the large scale properties and the function  $\phi(\cdot) \ge 0$  specifies pairwise interactions terms at the small scale. Although the marginal distribution of X restricted to W becomes analytically intractable, such models may at least be simulated and statistical inference may be performed by Markov chain Monte Carlo methods.

#### **3** Theoretical results

Theoretical properties of Cox processes have been extensively studied, see e.g. Grandell (1976), Daley and Vere-Jones (1988), and Karr (1991). In this section we establish further results for log Gaussian Cox processes. In particular, we discuss the first, second and third-order properties of a univariate log Gaussian Cox process. As in the previous section we consider the planar case, but many of the presented results hold as well in  $\mathbb{R}^d$ , d = 1, 2... (with obvious modifications in a few places).

The most useful characteristics for our purpose are the *n*th order product densities  $\rho^{(n)}$ , n = 1, 2, ..., for the reduced moment measures of the Cox process X. These are given by the moments of the intensity process as

$$\rho^{(n)}(s_1,\ldots,s_n) = E \prod_{1}^n \Lambda(s_i)$$

for pairwise different  $s_1, \ldots, s_n \in \mathbb{R}^2$ . Intuitively speaking  $\rho^{(n)}(s_1, \ldots, s_n)ds_1 \cdots ds_n$  is the probability that X has a point in each of n infinitesimally small disjoint regions of volumes  $ds_1, \ldots, ds_n$ .

**Theorem 1.** A log Gaussian Cox process X is stationary if and only if the corresponding Gaussian field Y is stationary. For a stationary log Gaussian Cox process we have

$$\rho^{(n)}(s_1, \dots, s_n) = \exp\left\{n\mu + \sigma^2 \left[\frac{n}{2} + \sum_{1 \le i < j \le n} r(s_i - s_j)\right]\right\}$$

$$= \rho^n \prod_{1 \le i < j \le n} g(s_i - s_j)$$
(2)

where

$$\rho = \rho^{(1)}(s) = \exp\left\{\mu + \sigma^2/2\right\}$$
(3)

and

$$g(s_1 - s_2) = \rho^{(2)}(s_1, s_2) / \rho^2 = \exp\left\{\sigma^2 r(s_1 - s_2)\right\}$$
(4)

are the intensity and the pair correlation function of the process, respectively.

*Proof.* Let  $c(t) = \exp(\xi + \kappa t^2/2)$  be the Laplace transform of the normal distribution  $N(\xi,\kappa)$  with mean  $\xi$  and variance  $\kappa$ . Let  $\xi = \sum_{1}^{n} \mu(s_i)$  and  $\kappa = \sum_{1}^{n} \sigma^2(s_i) + 2 \sum_{1 \le i < j \le n} \sigma(s_i) \sigma(s_j) r(s_i, s_j)$  where  $\mu(s) = E(Y(s))$ ,  $\sigma^2(s) = Var(Y(s))$ , and  $r(\cdot, \cdot)$  is the correlation function of Y. Then  $\sum_{1}^{n} Y(s_i) \sim N(\xi,\kappa)$ . Hence, by (1),  $\rho^{(n)}(s_1, \ldots, s_n) = E \exp\{\sum_{1}^{n} Y(s_i)\} = \exp(\xi + \kappa/2)$ . The first order product density  $\rho^{(1)}(s)$  and the pair correlation function  $g(s_1, s_2) = \rho^{(2)}(s_1, s_2)/(\rho^{(1)}(s_1)\rho^{(1)}(s_2))$  are in particular given by  $\rho^{(1)}(s) = \exp(\mu(s) + \sigma^2(s))$  and  $g(s_1, s_2) = \exp(\sigma(s_1)\sigma(s_2)r(s_1, s_2))$ , whereby (2)-(4) follow when Y is stationary.

If X is stationary then  $\rho^{(1)}(s) = \rho$ , and we can write  $g(s_1, s_2) = g(s_1 - s_2)$ . By letting  $s_1 = s_2$  it follows that  $\sigma^2(s) = \sigma^2 = g(0)$  is constant, and further that  $\mu(s) = \mu = \log(\rho) - \sigma^2/2$  and  $r(s_1, s_2) = r(s_1 - s_2) = \log(g(s_1 - s_2))/\sigma^2$ , whereby Y is stationary. Finally, by definition of a log Gaussian Cox process, stationarity of Y implies stationarity of X.

Theorem 1 reflects the fact that the distribution of a log Gaussian Cox process is completely determined by  $(\mu, \sigma, r(\cdot))$  or equivalently by  $(\rho, g(\cdot))$  (since r(0) = 1). It follows from (4) and the definition of a log Gaussian Cox process that it is isotropic if and only if the underlying Gaussian process is isotropic or equivalently when  $g(\cdot) = g(||\cdot||)$ . Especially, when  $\Lambda(\cdot) = \lambda$  is a log Gaussian random variable we have a mixed Poisson process with  $r(\cdot) = 1$  and  $g(\cdot) = e^{\sigma^2} > 1$ , whilst for a homogeneous Poisson process  $r(\cdot) = 0$  and  $g(\cdot) = 1$ .

Similar results may be established for other intensity processes being a function of a Gaussian process. Suppose e.g. for the moment that  $\mu = 0$  and  $\Lambda(s) = Y(s)^2$  is  $\sigma^2 \chi^2$ -distributed with 1 degrees of freedom. Then the intensity of the ' $\chi^2$  Cox process' is  $\rho_{\chi^2} = \sigma^2$  and the pair correlation function becomes

$$g_{\chi^2}(s) = 1 + 2r(s)^2.$$

Hence there is not a one-to-one correspondence between  $(\sigma^2, r(\cdot))$  and  $(\rho_{\chi^2}, g_{\chi^2}(\cdot))$  unless the sign of the correlation function is known.

In the statistical analysis of point processes mostly first and second-order properties are investigated (see Section 7), but we shall also explore the following correspondence between the second and third-order properties: For any stationary simple point process X with finite intensity  $\rho > 0$  and well-defined pair correlation function  $g(x_1, x_2) =$  $g(x_1 - x_2) > 0$  and third-order density  $\rho^{(3)}(x_1, x_2, x_3) = \rho^{(3)}(x_2 - x_1, x_3 - x_1)$  define

$$z(t) = \frac{1}{\pi^2 t^4} \int_{||\xi|| \le t} \int_{||\eta|| \le t} \frac{\rho^{(3)}(\xi, \eta)}{\rho^3 g(\xi) g(\eta) g(\xi - \eta)} d\xi d\eta, \quad t > 0.$$
(5)

This has an interpretation as a third-order summary statistic, since

$$\pi^{2} t^{4} \rho^{2} z(t) = E_{0}^{!} \sum_{\xi, \eta \in X: \, ||\xi|| \le t, \, ||\eta|| \le t}^{\neq} 1/\{g(\xi)g(\eta)g(\xi - \eta)\}$$

where  $\neq$  means that the summation is over pairwise distinct points, and where the expectation is with respect to the reduced Palm distribution at the origin (heuristically this means that we have conditioned on that there is a point at the origin and X denotes the collection of the remaining points, cf. e.g. Stoyan *et al.*, 1995). By Theorem 1,

$$z(t) = 1$$
,  $t > 0$ , for a log Gaussian Cox process. (6)

This can be used to check our model assumptions as demonstrated in Section 7.

In the case of rotation invariance we propose an unbiased estimator which uses all triplets of observed points and which takes care of edge effects as follows. For a given 'window'  $W \subset \mathbb{R}^2$  and  $x_1 \in W$ , a > 0, b > 0,  $0 \le \psi < 2\pi$ , let

$$U_{x_1,a,b,\psi} = \{ \varphi \in [0, 2\pi) \, | \, x_1 + a(\cos\varphi, \sin\varphi) \in W, \\ x_1 + b(\cos(\varphi + \psi), \sin(\varphi + \psi)) \in W \}$$

and define the 'edge correction'

$$w_{x_1,a,b,\psi} = 2\pi / \{ length \ of \ U_{x_1,a,b,\psi} \}$$

taking  $2\pi/0 = \infty$ . Then for given  $x_1$ , a, b and  $\psi$ ,  $1/w_{x_1,a,b,\psi}$  is the proportion of triangles which can be observed within W with vertices  $x_1, x_2, x_3 \in W$  such that  $||x_2 - x_1|| = a$ ,  $||x_3 - x_1|| = b$ , and  $\psi$  is the angle (anticlockwise) between the vectors  $x_2 - x_1$  and  $x_3 - x_1$ .

**Theorem 2.** Let  $\psi(x_1, x_2, x_3)$  denote the angle (anticlockwise) between  $x_2 - x_1$  and  $x_3 - x_1$ . For any stationary simple point process X as considered above, assuming that the distribution of X is invariant under rotations about the origin,

$$2\sum_{x_1\in X\cap W}\sum_{\substack{\{x_2,x_3\}\subseteq X\cap W\setminus\{x_1\}:\\||x_1-x_2||\leq t,\ ||x_1-x_3||\leq t}}^{\neq} \frac{w_{x_1,||x_1-x_2||,||x_1-x_3||,\psi(x_1,x_2,x_3)}}{g(||x_1-x_2||)g(||x_1-x_3||)g(||x_2-x_3||)}$$
(7)

is an unbiased estimator of  $A(W)\pi^2 t^4 \rho^3 z(t)$  for all  $t < t^*$ , where A(W) is the area of W and

$$t^* = \inf \left\{ t > 0 | \int_{x_1 \in W} \int_{a \in (0,t]} \int_{b \in (0,t]} \int_{\psi \in [0,2\pi)} \mathbf{1} \left[ w_{x_1,a,b,\psi} = \infty \right] d\psi da db dx_1 > 0 \right\}.$$

*Proof.* Note that factor 2 in (7) appears because the second summation is over unordered pairs of distinct points. Then  $\rho^{(3)}(x_2 - x_1, x_3 - x_1) = \rho^{(3)}_0(a, b, \psi)$  for a function  $\rho^{(3)}_0$  because of the rotation invariance. Moreover,  $||x_2 - x_3|| = f(a, b, \psi)$  is a function of  $(a, b, \psi) = (||x_1 - x_2||, ||x_1 - x_3||, \psi(x_1, x_2, x_3))$  only. Hence, using that the correction factor w is the same for  $(x_1, x_2, x_3)$  as for  $(x_1, x_3, x_2)$  together with the fact that

$$E\sum_{x_1,x_2,x_3\in X}^{\neq} h(x_1,x_2,x_3) = \int \int \int \rho^{(3)}(x_1,x_2,x_3)h(x_1,x_2,x_3)dx_1dx_2dx_3$$

for nonnegative measurable functions h, we find that the mean of (7) equals

$$\begin{split} E & \sum_{x_1,x_2,x_3 \in X \cap W: \, ||x_1 - x_2|| \le t, \, ||x_1 - x_3|| \le t} \frac{w_{x_1,||x_1 - x_2||,||x_1 - x_3||,\psi(x_1,x_2,x_3)}}{g(||x_1 - x_2||)g(||x_1 - x_3||)g(||x_2 - x_3||)} \\ &= \int_{W} \int_{W} \int_{W} \int_{W} \frac{\rho_0^{(3)}(a, b, \psi) w_{x_1,a,b,\psi}}{g(a)g(b)g(f(a, b, \psi))} \mathbf{1}(0 < a \le t, 0 < b \le t) dx_1 dx_2 dx_3 \\ &= \int_{x_1 \in W} \int_{a \in (0,t]} \int_{b \in (0,t]} \int_{\psi \in [0,2\pi)} \int_{\varphi \in U_{x_1,a,b,\psi}} \frac{\rho_0^{(3)}(a, b, \psi) w_{x_1,a,b,\psi}}{g(a)g(b)g(f(a, b, \psi))} abd\varphi d\psi dadb dx_1 \\ &= \int_{x_1 \in W} \int_{a \in (0,t]} \int_{b \in (0,t]} \int_{\psi \in [0,2\pi)} \int_{\varphi \in [0,2\pi)} \frac{\rho_0^{(3)}(a, b, \psi)}{g(a)g(b)g(f(a, b, \psi))} abd\varphi d\psi dadb dx_1 \\ &= A(W) \int_{||\xi|| \le t} \int_{||\eta|| \le t} \int_{\theta \in (0,t]} \frac{\rho^{(3)}(\xi, \eta)}{g(\xi)g(\eta)g(\xi - \eta)} d\xi d\eta \end{split}$$

where we have used that  $t < t^*$  to obtain the third equality. This combined with (5) gives the result.

The estimator (7) is of the same spirit as Ripley's (1977) estimator for the second order reduced moment measure. In fact  $w_{x_1,a,b,\psi}$  agrees with Ripley's edge correction factor when a = b and  $\psi = 0$ . Our edge correction factor is of course also applicable for other third order summary statistics than z.

Applications of z and its estimator are discussed at the end of Section 4 and in Section 7, Example 1. In most applications W will be convex in which case  $t^* \ge I(W)$ , the radius of the maximal inner ball contained in W. We have also considered a naive estimator based on 'minus sampling' and which do not presume rotation invariance, viz. the unbiased estimator of  $A(W_{\ominus t})\pi^2 t^4 \rho^3 z(t)$  given by

$$\sum_{x_1, x_2, x_3 \in X: x_1 \in W_{\ominus t}, a \le t, b \le t}^{\neq} \{g(x_1 - x_2)g(x_1 - x_3)g(x_2 - x_3)\}^{-1}$$

with

$$W_{\ominus t} = \left\{ s \in W \mid \forall u \in \mathbb{R}^2 : ||u|| \le t \implies s + u \in W \right\}.$$

Compared to (7) the variation of this estimator can be very large since not all triplets of points in  $X \cap W$  are used. Another problem may be caused by a clustering of points so that no points are observed within  $W_{\ominus t}$  for even moderate values of t.

Finally, we establish some simple results about ergodicity. Ergodicity may for instance become useful for establishing consistency of nonparametric estimators of  $\rho$  and  $g(\cdot)$ . The log Gaussian Cox processes corresponding to the correlation models in Table 1 are all ergodic as shown in part (b) of Theorem 3 below.

**Theorem 3.** (a) Let  $Z = \{Z(s) : s \in \mathbb{R}^2\}$  be a stationary real-valued stochastic process, let  $h : \mathbb{R} \to [0, \infty)$  be measurable, and suppose that with probability one  $\int_B h(Z(s))ds < \infty$  for bounded Borel sets  $B \subset \mathbb{R}^2$ . Then a Cox process with random intensity function  $\{h(Z(s)) : s \in \mathbb{R}^2\}$  is ergodic if Z is ergodic. Conversely, assuming that the realizations of Z are continuous with probability one and that h is strictly monotone, ergodicity of the Cox process implies that Z is ergodic.

(b) If Z is a stationary Gaussian process where the correlations decay to zero, i.e. when

$$r(s) \to 0 \ as \ ||s|| \to \infty$$
 (8)

then the corresponding log Gaussian Cox process is ergodic. Especially, a stationary log Gaussian Cox process is ergodic if

$$g(s) \to 1 \quad as \quad ||s|| \to \infty.$$
 (9)

*Proof.* We first need some measure theoretical details. Let  $F = \mathbb{R}^{\mathbb{R}^2}$  denote the space of functions  $f : \mathbb{R}^2 \to \mathbb{R}$  equipped with the  $\sigma$ -field  $\sigma_F$  generated by the projections  $p_s : F \to \mathbb{R}, s \in \mathbb{R}^2$ , where  $p_s(f) = f(s)$ . Further, let  $(M, \mathcal{M})$  be the measure space of locally finite measures defined on the Borel  $\sigma$ -field  $\mathcal{B}_2$  in  $\mathbb{R}^2$  where the  $\sigma$ -field  $\mathcal{M}$ is generated by the projections  $\tilde{p}_A : M \to \mathbb{R}, A \in \mathcal{B}_2$ , given by  $\tilde{p}_A(m) = m(A)$ . Furthermore, let  $H : F \to M$  be defined by

$$H(f)(A) = \int_{A} h(f(s))ds, \ A \in \mathcal{B}_2.$$

It is not difficult to show that for any fixed  $A \in \mathcal{B}_2$ , the function  $H_A : F \to \mathbb{R}$  given by  $H_A(f) = \tilde{p}_A(H(f))$  is measurable. Hence H is measurable and so  $\Xi = H(Z)$  is a random measure. Now, consider a stationary Cox process as in (a). This is ergodic if and only if the random measure  $\Xi$  is ergodic, cf. e.g. Proposition 10.3.VII in Daley and Vere-Jones (1988). Ergodicity of  $\Xi$  means that  $P(\Xi \in I) \in \{0, 1\}$  for all events  $I \in \mathcal{M}$  which are invariant under translations in the plane (I is invariant if  $m \in I \Rightarrow m_t \in I$  for all  $t \in \mathbb{R}^2$  where  $m_t(A) = m(\{s : s + t \in A\})$ ). Similarly, ergodicity of Z means that  $P(Z \in J) \in \{0, 1\}$  for all events  $J \in \sigma_F$  which are invariant under translations in the plane (i.e. J is invariant if  $f \in J \Rightarrow f_t \in J$  for all  $t \in \mathbb{R}^2$  where  $f_t(s) = f(s + t)$ ). Using these definitions it is straightforward to show the first implication in (a).

Assuming that h is strictly monotone, then H restricted to  $F_c = \{f \in F : f \text{ continuous}\}\$  becomes injective. Assume further that  $J \in \sigma_F$  is invariant and  $\Xi$  is ergodic. Then it follows that H(J) is invariant so that  $P(Z \in H^{-1}(H(J))) \in \{0, 1\}$ . Under the additional assumption that realizations of Z are continuous a.s., it is no restriction to assume that  $J \subseteq F_c$ . Then, since H is injective on  $F_c$ ,  $H^{-1}(H(J)) \setminus J \subseteq F \setminus F_c$  whereby  $P(Z \in J) \in \{0, 1\}$ , and the second implication in (a) is proved.

According to (a) a stationary log Gaussian Cox process is ergodic if the underlying Gaussian process is ergodic. But ergodicity of the Gaussian process is in fact implied by (8), cf. Theorem 6.5.4 in Adler (1981). Using (4) we get the equivalence between (8) and (9). This completes the proof.  $\Box$ 

Conditions for continuity of random fields may be found in Adler (1981) or Ledoux & Talagrand (1991). Notice that (8) and (9) are equivalent and that (9) implies ergodicity also for a  $\chi^2$  Cox process.

# 4 Comparison with Neyman-Scott processes

We shall now compare our log Gaussian Cox processes with a popular and frequently used class of models which are simultaneously Poisson cluster and Cox processes, namely those Neyman-Scott processes where the number of points per cluster is Poisson distributed (see e.g. Bartlett, 1964; Diggle, 1983; Stoyan and Stoyan, 1994; Stoyan *et al.*, 1995).

Imagine a point process  $\{p_i\} \subset \mathbb{R}^2$  of (unobserved) parents which form a homogeneous Poisson point process of intensity  $\omega > 0$ , and which generate clusters of offspring  $\bigcup_{j=1}^{n_i} \{p_i + x_{ij}\}$ . The counts  $n_i$  are assumed to be *iid* Poisson distributed with mean  $\nu > 0$  and the relative positions  $x_{ij}$  of offspring are *iid* with density f. Further, the  $\{p_i\}, \{n_i\}, \text{and } \{x_{ij}\}$  are mutually independent. The Poisson cluster process of offspring  $\bigcup_{i=1}^{n_i} \{p_i + x_{ij}\}$  is then stochastic equivalent to a Cox process with intensity process

$$\Lambda(s) = \nu \sum_{i} f(s - p_i).$$
(10)

The product densities of such Neyman-Scott processes are known: We have that  $\rho = \nu \omega$ ,

$$g(s) = 1 + \frac{1}{\omega} \int f(p)f(p+s)dp$$

$$\rho^{(3)}(s_1, s_2, s_3) = g(s_1 - s_2) + g(s_1 - s_3) + g(s_2 - s_3) - 2$$
$$+ \frac{1}{\omega^2} \int f(p + s_1) f(p + s_2) f(p + s_3) dp$$

and with similar but longer expressions for  $\rho^{(n)}$ ,  $n \ge 4$ . The higher-order product densities of a log Gaussian Cox process as given by Theorem 1 are in general of a different and much simpler form than for Neyman-Scott processes.

In the following we consider some particular but widely used models of Neyman-Scott processes, viz. a Matérn (1960) cluster process and a (modified) Thomas (1949) process (Bartlett, 1964). For the Thomas process, f is the density of a radially symmetric Normal distribution with variance  $\kappa > 0$ , and the pair correlation function becomes

$$g_{\mathrm{T}}(a) = 1 + \frac{1}{4\pi\omega\kappa} \exp\left(-\frac{a^2}{4\kappa}\right), \ a \ge 0.$$

For the Matérn cluster process, f is the density for a uniform distribution on a disc with radius R > 0 centered at 0, and the pair correlation function becomes

$$g_{\rm M}(a) = \begin{cases} 1 + \frac{2}{\omega \pi^2 R^2} \left\{ \arccos \frac{a}{2R} - \frac{a}{2R} \sqrt{1 - \frac{a^2}{4R^2}} \right\} &, \ 0 \le a \le 2R\\ 1 &, \ a > 2R. \end{cases}$$

In Fig. 1 we have included plots of the pair correlation functions for Thomas and Matérn cluster processes. For comparison we have taken  $g_{\rm M}(0) = g_{\rm T}(0) = e$ . Then for the Thomas process  $\omega = 1/(4\pi\kappa(e-1))$  is determined by the value of  $\kappa$ , whilst for the Matérn cluster process  $\omega$  is determined by the value of R. At least for certain values of  $\beta$  and  $\kappa$  the Gaussian pair correlation function and  $g_{\rm T}(\cdot)$  appear to be very similar, whereas  $g_{\rm M}(\cdot)$  looks different from the other pair correlation functions in Fig. 1. For instance, by taking  $\kappa = .001$  and minimizing  $\int_0^{1/2} (g(a) - g_{\rm T}(a))^2 da$  with respect to  $\beta$ , where  $\log g(a) = \exp(-(a/\beta)^2)$ , we get  $\beta = 1/13.45$ . The left plot in Fig. 4 shows that the logarithm of these pair correlation functions for the Thomas and the log Gaussian Cox process with Gaussian correlation function are nearly identical.

This may suggest that  $c_T(\cdot) = \log g_T(\cdot)$  could be considered as a covariance function. One way to check this is through the Hankel transform of  $c_T(\cdot)$  given by

$$C_{\rm T}(t) = \frac{1}{2\pi} \int_{0}^{\infty} J_0(at) a c_{\rm T}(a) da$$

where

$$J_0(t) = \sum_{k=0}^{\infty} (-1)^k \frac{t^{2k}}{(k!)^2 2^{2k}}$$

is the Bessel function of first kind and order zero. Then  $c_{\rm T}(\cdot)$  is positive semi-definite if and only if  $C_{\rm T}(t) \ge 0$  for all  $t \ge 0$ , cf. e.g. Christakos (1984). The Hankel transforms  $C_{\rm T}$  and  $C_{\rm M}$  for the Thomas and Matérn cluster processes given in Fig. 4 show that neither of the two Neyman-Scott processes can be considered as log Gaussian Cox processes. But the close agreement with respect to the pair correlation functions and the remarks below and at the end of this section suggest that certain Thomas processes may in practice be difficult to distinguish from log Gaussian Cox processes with a Gaussian correlation function.

Fig. 5 shows simulated distribution functions F and G for the distance to the nearest point of a point process X with respect to 0 and a typical point of X, respectively (see e.g. Diggle, 1983). Here we consider two kinds of point processes: a log Gaussian Cox process (the solid lines in Fig. 5) with a Gaussian correlation function and  $\mu = 3.888$ ,  $\sigma = 1$ , i.e.  $\rho = e^5$ , and correspondingly a Thomas process with  $\omega = 1/(4\pi(e-1)\kappa)$  and  $\nu = \rho/\omega$  (dotted lines); as before  $\kappa = .001$  and  $\beta = 1/13.45$ . For each model we simulated 100 realizations and calculated the average and the upper and lower envelopes for nonparametric estimates of F and G. (The upper and lower envelopes for F, G or any another summary statistic depending only on the distance are here and elsewhere in the following given by the maximum and minimum values obtained from the simulations at each distance; see e.g. Diggle, 1983.) The averages are then estimates of the theoretical F and G functions. Further simulations confirmed that the envelopes of F for the Thomas process lie beneath those for the log Gaussian Cox process, while the opposite statement holds for the envelopes of G. We recognized further that the G function distinguishes better between the two processes than the F function, but also that none of these summary statistics are really useful for discriminating between the two models. Another experiment confirmed that it may also be difficult to distinguish between the two models by means of the third-order characteristic z in (5).

In Section 7 plots of F, G, and z raise doubt of the appropriateness of the Matérn cluster process as a model for the data in Example 1, but give no reason to question the use of a log Gaussian Cox process with an exponential correlation function.

#### 5 Multivariate log Gaussian Cox processes

Our model can immediately be extended to the case of multivariate Cox processes as follows.

Let us for simplicity just consider the bivariate case of a Cox process  $X = (X_1, X_2)$  directed by random intensity processes  $\Lambda_j = \{\Lambda_j(s) = \exp(Y_j(s)) : s \in \mathbb{R}^2\}$ , j = 1, 2, where  $Y = \{(Y_1(s), Y_2(s)) : s \in \mathbb{R}^2\}$  is a bivariate stationary and possibly isotropic Gaussian process with mean  $(\mu_1, \mu_2)$  and covariance functions  $c_{ij}(a) = Cov(Y_i(s_1), Y_j(s_2))$  for  $a = ||s_1 - s_2||$ , i, j = 1, 2 (in the isotropic case we have that  $c_{12}(\cdot) = c_{21}(\cdot)$ ). Then conditional on Y,  $X_1$  and  $X_2$  are independent Poisson processes with intensity functions  $\Lambda_1$  and  $\Lambda_2$ , respectively. The covariance function matrix of the multivariate Gaussian process must be positive semi-definite. Restricting attention to absolutely integrable and isotropic covariance functions, this is equivalent to that

$$C_{11}(t) \ge 0, \ C_{22}(t) \ge 0, \ \text{ and } |C_{12}(t)|^2 \le C_{11}(t)C_{22}(t), \ t \ge 0$$
 (11)

where

$$C_{ij}(t) = \frac{1}{2\pi} \int_{0}^{\infty} J_0(at)c_{ij}(a)ada$$

is the spectral density or Hankel transform of  $c_{ij}$  (Yaglom, 1986; Christakos, 1992; Wackernagel, 1995). Moreover, many of the results presented in Section 3 may easily be extended to the multivariate case. For example, by Theorem 1 the intensity and pair correlation function of  $X_j$  become

$$\rho_j = \exp\left\{\mu_j + c_{jj}(0)/2\right\}, \quad g_{jj}(a) = \exp\left\{c_{jj}(a)\right\}$$
(12)

and the mixed pair correlation function is given by

$$g_{12}(a) = E[\Lambda_1(s_1)\Lambda_2(s_2)]/(\rho_1\rho_2) = \exp\{c_{12}(a)\}, \ a = ||s_1 - s_2||.$$
(13)

Especially, if we consider affine transformations  $Y_j(s) = \sum_{i=1}^{k} \alpha_{ij} Z_i(s) + \mu_j$  of k independent one-dimensional Gaussian processes  $Z_i = \{Z_i(s) : s \in \mathbb{R}^2\}, i = 1, ..., k$ , each with mean 0, variance 1, and a positive semi-definite correlation function  $r_i$ , then of course Y is well-defined and

$$c_{jj}(s) = \sum_{i=1}^{k} \alpha_{ij}^2 r_i(s) , \ j = 1, 2 , \ c_{12}(s) = \sum_{i=1}^{k} \alpha_{i1} \alpha_{i2} r_i(s) , \ s \in \mathbb{R}^2$$
(14)

(in this case  $c_{12}(\cdot) = c_{21}(\cdot)$  no matter if isotropy is required or not). For example, if  $Y_j = \sigma_j Z + \mu_j$ , j = 1, 2, where Z is a stationary Gaussian process with mean 0, variance 1, and correlation function  $r(\cdot)$ , then the sign of  $\sigma_1 \sigma_2 r(\cdot)$  determines whether there is a positive or negative dependence structure between the two types of patterns  $X_1$  and  $X_2$ . In the special case  $\sigma_1 = \sigma_2$  we have a linked Cox process (Diggle and Milne, 1983) as  $\rho_2 \Lambda_1(\cdot) = \rho_1 \Lambda_2(\cdot)$ . Fig. 6 shows realizations on the square under the exponential model  $r(a) = \exp(-10a)$  with  $\mu_1 = \mu_2 = 2.5$  and for each of  $\sigma_1 = \sigma_2 = 2$  and  $\sigma_1 = -\sigma_2 = 2$ . The different dependence structures are clearly expressed in the simulations.

#### 6 Simulation algorithms

Some properties of Cox processes are hard to evaluate analytically. Fortunately, log Gaussian Cox processes are easy to simulate so that Monte Carlo methods can be applied. An advantage of log Gaussian Cox processes is that there are no boundary effects since all marginal distributions of a Gaussian field are known.

In practice we represent the finite domain of simulation by a grid and approximate the Gaussian process by the values of the corresponding finite dimensional Gaussian distribution on the grid. If we for example wish to simulate a log Gaussian Cox process on the unit square, we approximate the Gaussian process  $\{Y(s)\}_{s\in[0,1]^2}$  on each cell  $D_{ij} = [i-1/(2M), i+1/(2M)[\times[j-1/(2M), j+1/(2M)]]$  by its value  $\tilde{Y}_{ij} = Y((i, j))$  at the center (i, j) of  $D_{ij}$  where  $(i, j) \in I = \{1/(2M), 1/M + 1/(2M), \dots, (M-1)/M +$  1/(2M)<sup>2</sup>, and M is a suitable value for the discretization. Thus, simulations of the field  $\tilde{Y} = (\tilde{Y}_{ij})_{(i,j) \in I}$  are required. For ease of presentation we shall here mainly focus on univariate log Gaussian Cox processes where the discretization is given by a square lattice I; at the end of Subsection 6.1 we consider briefly the case of a multivariate log Gaussian Cox process and a rectangular lattice.

If the Cox process is moderately clustered and the intensity moderate, the very fine scale properties of the Gaussian field are probably not so important and a rather coarse discretization can be used. The choice of discretization also depends on the smoothness of the realizations of the Gaussian field, see Fig. 2. The error due to discretization is e.g. likely to be small when the Gaussian correlation function is used. For the simulations presented in this paper we found it sufficient to use either  $65 \times 65$  or  $129 \times 129$  grids.

Simulation of a log Gaussian Cox process involves two steps. First the Gaussian field is simulated and secondly, given the Gaussian field  $\tilde{Y} = (\tilde{y}_{ij})_{(i,j)\in I}$ , the inhomogeneous Poisson process can be simulated: either within each cell  $D_{ij}$  where the Poisson process is homogeneous with intensity  $\tilde{\lambda}_{ij} = \exp(\tilde{y}_{ij})$ , or by thinning a homogeneous Poisson process with intensity  $\tilde{\lambda}_{max} = \max_{ij} \tilde{\lambda}_{ij}$  so that a Poisson point situated in the *ij*th cell is retained with probability  $\tilde{\lambda}_{ij}/\tilde{\lambda}_{max}$ .

There are several methods available for simulation of a Gaussian random field, see e.g. Lantuéjoul (1994). The simulation method based on Cholesky decomposition of the covariance matrix is too slow even for moderate grid sizes. We used another method based on decomposition of the covariance matrix (see Subsection 6.1) or alternatively the turning bands method (Matheron, 1973). In Subsection 6.2 we describe how simulations conditional on the number of points can be obtained. Finally, in Subsection 6.3 we briefly discuss how the Thomas and Matérn cluster processes studied in Section 4 are simulated.

#### 6.1 Simulation using diagonalization by the two-dimensional discrete Fourier

**transform:** A detailed description of this method in the univariate case and any lattice dimension d = 1, 2, ... assuming only stationarity can be found in Wood and Chan (1994). Below we summarize this for the two-dimensional case (the notation and the results are also used in Sections 7 and 8). For simplicity we assume isotropy.

Suppose that an isotropic covariance function  $c : \mathbb{R}^2 \to \mathbb{R}$  is given and we wish to simulate a Gaussian field  $\tilde{Y} = (\tilde{Y}_{ij})_{(i,j)\in I}$  with covariance matrix  $\Sigma = (\sigma_{ij,kl})_{(i,j),(k,l)\in I} = (c(||(i,j)-(k,l)||))_{(i,j),(k,l)\in I}$  (here we use a lexicographic ordering of the indices ij). Note that  $\Sigma$  is block Toeplitz and block symmetric. Extend the lattice I to  $I_{ext} = \{1/(2M), 1/M + 1/(2M), \dots, (2(M-1)-1)/M + 1/(2M)\}^2$  wrapped on a torus. Let  $d_{ik} = \min(|i-k|, 2(M-1)/M-|i-k|), (i,k) \in I_{ext}$ , and let  $d((i,j), (k,l)) = \sqrt{d_{ik}^2 + d_{jl}^2}$  denote the shortest distance on the torus between (i,j) and (k,l). The symmetric matrix  $K = (\kappa_{ij,kl})_{(i,j),(k,l)\in I_{ext}}$  defined by  $\kappa_{ij,kl} = c(d((i,j),(k,l)))$  is block circulant with 2(M-1) circulant blocks of dimension  $2(M-1) \times 2(M-1)$ . Hence, by Theorem 5.8.1 in Davis (1979),

$$K = \left(\bar{F}_{2(M-1)} \otimes \bar{F}_{2(M-1)}\right) E\left(F_{2(M-1)} \otimes F_{2(M-1)}\right)$$
(15)

where  $F_{2(M-1)} \otimes F_{2(M-1)}$  is unitary and  $E = \text{diag}(e_{ij}, (i, j) \in I_{ext})$  is a diagonal matrix of the eigenvalues for K. Here  $F_{2(M-1)} = (\sqrt{2(M-1)} \exp(-i2\pi kl/(2(M-1))))_{(k,l)\in I_{ext}}$  is the (normalized)  $2(M-1) \times 2(M-1)$  discrete Fourier transform matrix, "denotes complex conjugate, and  $\otimes$  is the Kronecker product.

Now, suppose that K is positive semi-definite (i.e. K has nonnegative eigenvalues). Then we can extend  $\tilde{Y}$  to a Gaussian field  $\tilde{Y}_{ext} = (\tilde{Y}_{ij})_{(i,j) \in I_{ext}}$  with covariance matrix K. Using the above decomposition of K we find that

$$\tilde{Y}_{ext} \stackrel{\mathcal{D}}{=} \Gamma D^{1/2} A \big( \bar{F}_{2(M-1)} \otimes \bar{F}_{2(M-1)} \big)$$

where  $\Gamma \sim N_d(0, I)$  follows a *d*-dimensional standard normal distribution with *d* equal to the rank of *K*, *D* is a diagonal matrix given by the non-zero eigenvalues of *K*, and *A* is a certain  $d \times (2(M-1))^2$  complex matrix of rank *d*. If M-1 is a power of two (or three or five), the calculation of  $\tilde{Y}_{ext}$  is only a  $O((2(M-1))^2 \log_2((2(M-1))^2)))$ operation as the two-dimensional fast Fourier transform (see e.g. Press et al., 1988) can be applied. Thereby a fast simulation algorithm is obtained.

Notice that the extension of the lattice  $\{1/(2M), 1/M + 1/(2M), \dots, (M-1)/M + 1/(2M)\}^2$  to  $\{1/(2M), 1/M + 1/(2M), \dots, (2(M-1)-1)/M + 1/(2M)\}^2$  is the minimal extension which gives a block circulant matrix K. If K turns out not to be positive semi-definite, it may help to use a larger extension (see Wood and Chan, 1994). Also, if M-1 is not a power of two (or three or five), a larger extension can be applied in order to use the two-dimensional fast Fourier transform.

The algorithm can straightforwardly be generalized to the case of a multivariate Gaussian field  $\tilde{Y} = ((Y_{ij1}, \ldots, Y_{ijn}))_{(i,j)\in I}$ , where I is a  $M \times N$  rectangular lattice and  $n \ge 1$ . In this case K becomes a  $4(M-1)(N-1)n \times 4(M-1)(N-1)n$  block circulant matrix given by 2(M-1) blocks, which in turn are block circulants and of dimension  $2(N-1)n \times 2(N-1)n$ . By combining (5.6.3), Theorem 5.6.4, and (3.2.2) in Davis (1979) one obtains that

$$K = \left(\bar{F}_{2(M-1)} \otimes \bar{F}_{2(N-1)} \otimes \bar{F}_n\right) G\left(F_{2(M-1)} \otimes F_{2(N-1)} \otimes F_n\right)$$

where G is a block diagonal matrix with 4(M - 1)(N - 1) blocks of dimension  $n \times n$ . In the bivariate case, simulation of  $\tilde{Y}$  thus amounts to a linear transformation of 4(M - 1)(N - 1) independent two dimensional Gaussian vectors.

The method is fast and practically applicable. Problems with nonpositive semidefiniteness of K occurred very seldom, and were then due to slowly decaying correlation functions like the stable correlation function (see Figure 1).

**6.2 Conditional simulation:** It may sometimes be desired to simulate the conditional distribution of  $X \cap [0, 1]^2$  given that  $N(X) = \operatorname{card}(X \cap [0, 1]^2) = n$  for  $n \in \mathbb{N}$ . Then we need first to simulate a realization  $\tilde{y}$  from  $\tilde{Y}|N(X) = n$  and secondly simulate from  $X|N(X) = n, \tilde{Y} = \tilde{y}$ . The last step is performed by distributing n independent points in the  $M^2$  grid cells, where a cell  $D_{ij}$  is chosen with a probability proportional to  $\tilde{\lambda}_{ij} = \exp(\tilde{y}_{ij}), (i, j) \in I$ , and the point subsequently placed at a uniformly sampled

location in the chosen cell. Rejection sampling (see e.g. Ripley, 1987) is used for the simulation of  $\tilde{Y}|N(X) = n$  as follows. For the conditional density of  $\tilde{Y}$  given N(X) = n we have that

$$f(\tilde{y} \mid n) \propto f(n \mid \tilde{y}) f(\tilde{y}) \le \frac{n^n}{n!} e^{-n} f(\tilde{y}).$$

The rejection sampling can thus be performed by generating realizations of  $\tilde{Y}$  until a realization  $\tilde{y}$  is accepted with probability  $(\bar{\lambda}/n)^n e^{n-\bar{\lambda}}$ , where  $\bar{\lambda} = \sum_{i=0}^{M-1} \sum_{j=0}^{M-1} \tilde{\lambda}_{ij}/M^2$ .

Considered  $\overline{\lambda}$  as a random variable, the mean of  $\overline{\lambda}$  approximates the intensity  $\rho$  of X. Thus the acceptance rates are reasonably high if n is close to  $\rho$  and the variance of  $\overline{\lambda}$  moderate.

**6.3 Simulation of the Thomas and Matérn cluster processes:** Procedures for simulation of the Thomas and Matérn cluster processes on a bounded region A follow straightforwardly from the definitions of these processes as Poisson cluster processes, see Section 4. In order to avoid boundary effects the parent process is simulated on an extended area B containing A. The area B is chosen so that offspring from a parent outside B falls into A with a negligible probability. An approximate procedure for simulation conditional on the number of points can be obtained by using that the Thomas and Matérn processes are Cox processes with intensity surface given by (10) and then proceed as described in Subsection 6.2 above.

# 7 Parameter estimation and model checking

For simplicity we first restrict attention to the univariate case, but our methods for estimation and model checking can also be used in the multivariate case, see Example 2 at the end of this section.

Suppose we have observed a point pattern  $x = \{x_1, \ldots, x_n\}$  within a bounded planar window W of area A(W). Under a homogeneous log Gaussian Cox model with a correlation function  $r_{\beta}(\cdot)$  the density of  $X_W = X \cap W$  with respect to a planar unit Poisson process is

$$L(\mu, \sigma, \beta) = E_{\mu, \sigma, \beta} \left[ \exp\left\{ \int_{W} (1 - \exp\left(Y(s)\right)) ds \right\} \prod_{1}^{n} \exp\left(Y(x_{i})\right) \right].$$

Except for very special models this likelihood is analytically intractable.

Considering this as a 'missing data problem' the likelihood can be approximated by discretizing W as described in Section 6 and making importance sampling as follows: The density of the Gaussian field  $\tilde{Y}$  is proportional to

$$h_{\theta}(\tilde{y}) = \exp\left(-\frac{1}{2\sigma^2}(\tilde{y} - \tilde{\mu})R(\beta)^{-1}(\tilde{y} - \tilde{\mu})^*\right)$$

where  $\theta = (\mu, \sigma, \beta)$ ,  $R(\beta)$  is the correlation matrix (here assumed to be positive definite), and \* denotes transposition. For a given fixed parameter  $\theta_0 = (\mu_0, \sigma_0, \beta_0)$  suppose that  $\tilde{y}^{(1)}, \ldots, \tilde{y}^{(M)}$  is a sample from the distribution of  $\tilde{Y}$  and  $\tilde{y}^{(1)}(x), \ldots, \tilde{y}^{(M)}(x)$  is a sample from the conditional distribution of  $\tilde{Y}$  given  $X_W = x$  (Section 8 describes how the latter sample can be generated). Since the conditional distribution of  $X_W$  given  $\tilde{Y}$ does not depend on  $\theta$ , it is easily seen from the results in Gelfand and Carlin (1991) and Geyer (1994) that the Monte Carlo approximation of the log likelihood is

$$l(\theta) \approx \text{const} + \log \frac{1}{M} \sum_{m=1}^{M} \frac{h_{\theta}\left(\tilde{y}^{(m)}(x)\right)}{h_{\theta_0}\left(\tilde{y}^{(m)}(x)\right)} - \log \frac{1}{M} \sum_{m=1}^{M} \frac{h_{\theta}\left(\tilde{y}^{(m)}\right)}{h_{\theta_0}\left(\tilde{y}^{(m)}\right)}$$

Actually we may replace  $\tilde{Y}$  with the extended Gaussian field  $\tilde{Y}_{ext}$  (see Section 6.1) for which it is easier to invert the correlation matrix. We have no experience about how this would work in practice, but we expect that multimodality of the likelihood may cause problems for finding the (approximate) maximum likelihood estimate. Since only the Gaussian density (up to scale) appears in the approximation of the log likelihood, there may be some analog here to Ripley's (1988) discussion on the difficulties associated with likelihood analysis for spatial Gaussian processes.

Pseudo-likelihood (Besag 1977; Jensen and Møller, 1991) is not useful since a closed expression of the density is not known even not up to multiplication with a positive constant (so a closed expression of the socalled Papangelou conditional intensity is not known). For the same reason we also doubt the usefulness of the more general method of Takacs-Fiksel estimation (see e.g. Ripley, 1988, and the references therein).

Since the distribution of a log Gaussian Cox process is completely determined by its first and second order properties we suggest instead to base the inference on corresponding summary statistics as described in the following.

As a natural estimate of the intensity we shall use

$$\hat{\rho} = n/A(W). \tag{16}$$

This estimator is unbiased. If  $r_{\beta}(a) \to 0$  as  $a \to \infty$ , then the ergodicity implies that  $\hat{\rho} \to \rho$  almost surely as W extends to  $\mathbb{R}^2$ , cf. Theorem 3.

The parameters  $\sigma^2 > 0$  and  $\beta > 0$  are estimated by a minimum contrast method: Assume henceforth that the correlation function is isotropic. Let  $\hat{c}(\cdot)$  denote a nonparametric estimate of the covariance function. Then  $\hat{\sigma}^2$  and  $\hat{\beta}$  are chosen to minimize

$$\int_{\epsilon}^{a_0} \left\{ \hat{c}(a)^{\alpha} - \left[\sigma^2 r_{\beta}(a)\right]^{\alpha} \right\}^2 da$$
(17)

where  $0 \le \epsilon < a_0$  and  $\alpha > 0$  are user specified parameters; in Examples 1 and 2 we take  $\epsilon = \min_{i \ne j} ||x_i - x_j||$ , while  $a_0$  and  $\alpha$  are determined by the form of  $\hat{c}(\cdot)$  and  $r_{\beta}(\cdot)$ . These parameters must of course be chosen so that the terms in (17) are well-defined.

For fixed  $\beta$  the minimum of (17) is obtained at

$$\hat{\sigma}_{\beta}^{2} = \left[B(\beta)/A(\beta)\right]^{1/\alpha} \text{ with } B(\beta) = \int_{\epsilon}^{a_{0}} \left\{\hat{c}(a)r_{\beta}(a)\right\}^{\alpha} da , A(\beta) = \int_{\epsilon}^{a_{0}} r_{\beta}(a)^{2\alpha} da$$

provided  $B(\beta) > 0$ ; otherwise there exists no minimum. Inserting this into (17) and using that  $\rho = \exp(\mu + \sigma^2/2)$  give the estimates

$$\hat{\beta} = \arg\max B(\beta)^2 / A(\beta) , \quad \hat{\sigma}^2 = \hat{\sigma}_{\hat{\beta}}^2 , \quad \hat{\mu} = \log\left(\hat{\rho}\right) - \hat{\sigma}^2 / 2. \tag{18}$$

Diggle (1983) describes a similar estimation procedure using the K-function

$$K(t) = 2\pi \int_{0}^{t} ag(a)da , \ t > 0$$

instead of the covariance function, but for the data considered later on we found that there may be many local minima, and it may be difficult to find a global minimum. The procedure in (18) is computationally much simpler; we need only to maximize with respect to  $\beta$ , whereas Diggle's procedure involves  $\sigma^2$  as well. In our examples the function  $B(\beta)^2/A(\beta)$  turned out to be unimodal.

As the nonparametric estimate of the covariance function we have used  $\hat{c}(\cdot) = \log \hat{g}(\cdot)$  with

$$\hat{g}(a) = \frac{1}{2\pi a \hat{\rho}^2 A(W)} \sum_{i} \sum_{j \neq i} k_h(a - ||x_i - x_j||) b_{ij} , \quad a < a^*$$
(19)

where  $k_h(\cdot)$  is the Epanecnikov kernel

$$k_h(a) = \frac{3}{4h} (1 - a^2/h^2) \mathbf{1}[-h \le a \le h]$$

with bandwidth h > 0,  $b_{ij}$  is the proportion of the circumference of the circle with center  $x_i$  and radius  $||x_i - x_j||$  lying within W, and  $a^*$  is the circumradius of W. The estimator (19) and other estimators of the pair correlation function are discussed in Stoyan and Stoyan (1994); in particular they discuss how to choose the bandwidth of the kernel.

To study how well our estimation procedure works we performed 20 simulations from the model with an exponential covariance function where  $\sigma^2 = 2.0$  and  $\beta = .05$ . A scatter plot of the estimated values of  $\beta$  and  $\sigma^2$  together with the true values is shown in Fig. 7. There is a large variation in the estimate of  $\beta$ , but the mean values of the 20 estimates are  $\bar{\beta} = .0513$  and  $\bar{\sigma}^2 = 2.08$ , not far from the true values. The other plot in Fig. 7 shows the mean covariance function  $\bar{c}$  (solid line) and upper and lower envelopes for the empirically estimated covariance functions obtained from the 20 simulations. The values of  $\bar{c}$  are close to the exponential covariance function, especially at small distances. Estimating the parameters from  $\bar{c}$  gives  $\hat{\sigma}^2 = 2.145$  and  $\hat{\beta} = .0461$ , which indicates that a good estimate of the covariance function gives good parameter estimates. Having estimated the parameters we may check our model assumptions by comparing nonparametric estimators of various summary statistics with those obtained under the estimated log Gaussian Cox model. We have considered the distribution functions F and G of the distance to the nearest point in X from a fixed point in the plane and a 'typical point' in X, respectively. Under the log Gaussian Cox model  $F = F_{\mu,\sigma^2,\beta}$ and  $G = G_{\mu,\sigma^2,\beta}$  are given by

$$F_{\mu,\sigma^2,\beta}(a) = 1 - E_{\mu,\sigma^2,\beta} \exp\left\{-\int\limits_{||s|| \le a} e^{Y(s)} ds\right\}$$

and

$$G_{\mu,\sigma^{2},\beta}(a) = 1 - e^{-\mu - \sigma^{2}/2} E_{\mu,\sigma^{2},\beta} \left[ e^{Y(0)} \exp\left\{ -\int_{||s|| \le a} e^{Y(s)} ds \right\} \right]$$

where the mean values may be approximated by Monte Carlo.

As in Diggle (1983), Stoyan and Stoyan (1994), and Stoyan et al. (1995) we have in Examples 1 and 2 compared nonparametric estimates of F, G,  $L = \sqrt{K/\pi}$  based on the data with those obtained by simulations under the estimated log Gaussian Cox model. For short we call such nonparametric estimates for empirical F, G, and Lfunctions. Moreover, we have obtained a nonparametric estimate of the third-order characteristic z in (5) by combining (7) with (16) and (19), and considered whether this summary statistic varies around 1 in accordance with the result (6) for log Gaussian Cox processes.

*Example 1:* The first data set consists of the locations of 126 Scots pine saplings in a square plot of  $10 \times 10 m^2$ . The pine forest has grown naturally in the Eastern Finland and the data have previously been analyzed by Penttinen *et al.* (1992) and Stoyan and Stoyan (1994), who both fitted a Matérn cluster process using the *L*-function both for parameter estimation and model checking. The estimation in Penttinen *et al.* (1992) was carried out by trial-and-error, while Stoyan and Stoyan (1994) used a minimum contrast method. The fit in both cases seems quite good, see Fig. 11 in Penttinen *et al.* (1992) and Fig. 131 in Stoyan and Stoyan (1994), but one may object to that the same summary statistics have been used for both estimation and model checking.

Fig. 8 shows several characteristics for the pine data: The data normalized to a unit square are shown in a). The logarithm of the estimated pair correlation function is plotted in b) (solid line), and the shape of the curve suggests to use the exponential covariance function. We estimated the parameters by minimizing (17) with  $a_0 = 0.1$  and  $\alpha = 0.5$ , which are chosen to give more weight to values of a close to zero. The estimates are  $\hat{\beta} = 1/33$  and  $\hat{\sigma}^2 = 1.91$ . The dotted line in b) shows the covariance function for the estimated model. The plot in c) shows the empirical *L*-function for the fitted model based on 19 simulations. This is the same as Stoyan and Stoyan (1994), Fig.

131, and our model shows a better fit with respect to the *L*-function than the Matérn cluster model. The empirical *L*-function falls within the envelopes from the simulations except for very small values of *t*. The plots in d) and e) show the nonparametric estimates  $\hat{F}$  and  $\hat{G}$  based on the data against the mean of these estimates obtained from 99 simulations under the estimated model. The plots show a reasonable good fit to the chosen model and  $\hat{F}$  and  $\hat{G}$  fall within the upper and lower envelopes based on the 99 simulations. For the Matérn cluster model fitted by Stoyan and Stoyan (1994) we have also created plots similar to d) and e) which indicate that our model fits the data better. Finally, f) shows a realization under the estimated log Gaussian Cox process.

We also used the third-order characteristic z to check our model assumptions. The left plot in Fig. 9 shows the estimated z for the data and two sets of envelopes based on 20 unconditional simulations of the estimated log Gaussian Cox process and 20 simulations where we condition on the observed number of points. The plot gives no reason to doubt the model no matter whether the 'unconditional' or 'conditional' envelopes are considered. The two sets of envelopes are not very different in this situation where  $\beta$  is rather small and the correlation therefore not very strong. To check the discriminatory power of z we similarly calculated envelopes for the Matérn cluster process estimated by Stoyan and Stoyan (1994), see the right plot in Fig. 9. The estimated z-function based on the data crosses the envelopes in an interval of t-values and even though the large variability of the estimator for small t makes it difficult to make definitive conclusions, the plot raises serious doubt concerning the appropriateness of the Matérn cluster process as a model for the data.

*Example 2:* In this example we study a bivariate data set consisting of two types of trees, 219 spruces and 114 birches in a square plot of  $50 \times 50 m^2$ . The data has been collected by Kari Leinonen and Markku Nygren as a part of a larger data set where also a very few pines were present and marks consisting of tree length and diameter were included. These data has earlier been studied by Kuuluvainen *et al.* (1996). They found that small trees are clustered, while larger trees are regularly distributed.

Fig. 10 a) shows the data normalized to unit area. The plot indicates clustering and a positive dependence between the two types of trees. In Fig. 10 b) the empirical covariance functions  $\hat{c}_{22}$ ,  $\hat{c}_{11}$  and  $\hat{c}_{12}$  are plotted (solid line, from top to bottom) using the equations (12) and (13) to obtain  $\hat{c}_{ij} = \ln \hat{g}_{ij}$ . Here  $\hat{g}_{jj}(a)$  is the estimate (19) based on the point pattern  $x_j = \{x_{j1}, \ldots, x_{jn_i}\}$  of type j trees. Further,

$$\hat{g}_{12}(a) = \frac{A(W)}{2\pi a n_1 n_2} \left\{ \frac{n_2}{n_1 + n_2} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} k_h(a - ||x_{1i} - x_{2j}||) b_{ij} + \frac{n_1}{n_1 + n_2} \sum_{i=1}^{n_2} \sum_{j=1}^{n_1} k_h(a - ||x_{2i} - x_{1j}||) b_{ij} \right\}$$

with the correction factor  $b_{ij}$  similarly defined as in (19), and where we have combined kernel estimation with the way Lotwick and Silverman (1982) and Diggle (1983) recommend to estimate  $K_{12}(t) = 2\pi \int_0^t ag_{12}(a)da, t > 0$ .

Based on the plot of the empirical covariance functions we specify a model for a bivariate log Gaussian Cox process with exponential covariance functions  $c_{ij}(a) = \sigma_{ij}^2 \exp(-a/\beta_{ij})$  and corresponding spectral density  $C_{ij}(t) = \sigma_{ij}^2/[2\pi\beta_{ij}(t^2 + \beta_{ij}^{-2})^{1.5}]$ . Estimating the parameters from the empirical covariance functions by minimizing (17) gives  $\hat{\sigma}_{11}^2 = 2.18$ ,  $1/\hat{\beta}_{11} = 12.0$ ,  $\hat{\sigma}_{22}^2 = 1.65$ ,  $1/\hat{\beta}_{22} = 16.0$ ,  $\hat{\sigma}_{12}^2 = 0.86$ ,  $1/\hat{\beta}_{12} = 9.0$ . The estimated covariance functions are shown as dotted lines in Fig. 10 b). This indicates a good fit to the empirical covariance functions. We have moreover checked that the condition (11) is fulfilled under the estimated model so that we have a valid covariance matrix function.

However, plots of the function F for each of the two types of trees show a poor fit of the estimated model to the data as there seems to be more 'empty space' in realizations of the estimated model than in the pattern a). As an example Fig. 10 c) shows the nonparametric estimate of  $F_{22}$  based on birch data plotted against the mean of the estimate obtained from 99 simulations under the estimated model. We have also tried to fit models with covariance functions as in (14) with k = 2 terms and various combinations of Gaussian, exponential and stable correlation functions, but again there was too much empty space under the fitted models. This may be caused by the regularity in the pattern of the larger trees, so one suggestion may be to include 'repulsive' pairwise interaction terms into the model as discussed at the end of Section 2. Another possibility is to include a thinning operation, cf. Diggle and Milne (1983) and Diggle (1983).

### 8 Prediction and Bayesian inference

We conclude this paper by considering prediction of the unobserved Gaussian process and intensity process under a given model for a univariate log Gaussian Cox process when this is observed within a bounded window. We use an empirical Bayesian approach, where the a posteriori distribution of the intensity process is obtained by considering the Gaussian distribution as a prior which smoothes the intensity surface, and where the prior may be estimated as described in Section 7. The posterior is not analytically tractable so we use a Markov chain Monte Carlo algorithm to simulate the posterior distribution whereby various posterior characteristics can be estimated. The results are applied on the data set in Example 1 and we compare various Bayesian estimators of the intensity process with a parametric kernel estimator studied in Diggle (1985), Berman and Diggle (1989), and Cressie (1993). Ogata and Katsura (1988) developed another objective Bayesian method for estimating the intensity function of a marked inhomogeneous Poisson point process using spline functions. Other related research but for Poisson (and more general) cluster processes include Lawson (1993), Baddeley and Van Lieshout (1993), and Granville and Smith (1995), who consider Bayesian estimation of cluster centres and cluster membership. Simultaneously with the development of the material of this section, Heikkinen and Arjas (1996) have been working with nonparametric Bayesian estimation of the intensity function of inhomogeneous planar Poisson processes generalizing the method of Arjas and Gasbarra (1994).

Suppose that a realization x of a log Gaussian Cox process is observed within a bounded window  $W^{(1)}$  and we wish to predict the Gaussian process and the intensity surface on the bounded set  $W \supseteq W^{(1)}$ . As in Section 6 we shall without loss of generality assume that W is the unit square and consider a finite subdivision of  $W^{(1)}$  and  $W^{(2)} = W \setminus W^{(1)}$  into cells  $D_{ij}$  of area  $A_{ij} > 0$ ,  $(i, j) \in I$ , where  $I = \{1/(2M), 1/M + 1/(2M), \ldots, (M-1)/M + 1/(2M)\}^2$ . Define the sublattices,  $I^{(a)} = W^{(a)} \cap I$ , a = 1, 2. Further, we approximate the Gaussian field Y restricted to W by a Gaussian field  $\tilde{Y} = (\tilde{Y}_{ij})_{(i,j)\in I}$  with mean vector  $\tilde{\mu} = (\mu)_{(i,j)\in I}$  and a covariance matrix  $\Sigma$  given by the covariance function of Y. As noticed in Subsection 6.1 we can extend  $\tilde{Y}$  to  $\tilde{Y}_{ext} = (\tilde{Y}_{ij})_{(i,j)\in I_{ext}} \stackrel{\mathcal{D}}{=} \Gamma Q + \tilde{\mu}_{ext}$ , where  $I_{ext} = \{1/(2M), 1/M + 1/(2M), \ldots, (2(M-1)-1)/M + 1/(2M)\}^2$ , K given by (15) is assumed to be positive semi-definite and of rank d,  $\Gamma \sim N_d(0, I)$ , Q is a certain  $d \times (2(M-1))^2$  real matrix of rank d, and  $\tilde{\mu}_{ext} = (\mu)_{(i,j)\in I_{ext}}$ . We shall later on explain why it (apart from ease of exposition) may be preferred to use  $\Gamma$  instead of  $\tilde{Y}_{ext}$ .

Now, if  $f(\gamma|x)$  denotes the density of the conditional distribution of  $\Gamma$  given that  $X \cap W^{(1)} = x$ ,

$$\log f(\gamma|x) = \text{const}(x) - \frac{1}{2} ||\gamma||^2 + \sum_{(i,j)\in I_{ext}} \left( \tilde{y}_{ij} n_{ij} - e^{\tilde{y}_{ij}} A_{ij} \right)$$
(20)

where  $n_{ij} = \operatorname{card}(x \cap D_{ij})$  is the number of points of x contained in the ijth cell if  $(i, j) \in I^{(1)}$ , and we set  $n_{ij} = A_{ij} = 0$  if  $(i, j) \notin I^{(1)}$ . Though this conditional distribution is not defined in accordance to the covariance structure of the Gaussian process outside W, we shall refer to this as the a posteriori distribution of  $\Gamma$  given x; the important point is that the marginal distribution of  $\tilde{Y}$  under this posterior agrees with the conditional distribution of  $\tilde{Y}$  given  $X \cap W^{(1)} = x$ . In the following the gradient of the posterior

$$\nabla(\gamma) := \partial \log f(\gamma|x) / \partial \gamma = -\gamma + \left( n_{ij} - e^{\tilde{y}_{ij}} A_{ij} \right)_{(i,j) \in I_{ext}} Q^*$$

plays a keyrole. It is easily seen that  $\partial \nabla(\gamma) / \partial \gamma^*$  is strictly negative definite. Thus the posterior is strictly log-concave.

For simulation of the posterior we use a Metropolis-adjusted Langevin algorithm (MALA) as suggested by Besag (1994) in the discussion of Grenander and Miller (1994) and further studied in Roberts and Tweedie (1997). This is a Metropolis-Hastings type Markov chain Monte Carlo (MCMC) method inspired by the definition of a Langevin diffusion through a stochastic differential equation which in the present context is

$$d\Gamma(t) = (h/2)\nabla(\Gamma(t))dt + \sqrt{h}dB(t)$$

where  $B(\cdot)$  is standard Brownian motion and h > 0 is a user specified parameter (see Example 1 below); the posterior  $\Gamma | x$  is a stationary distribution of this Markov process  $\Gamma(\cdot)$ .

The MALA is given by two steps: First, if  $\gamma^{(m)}$  is the current state of the chain, a 'proposal'  $u^{(m+1)}$  is generated from a multivariate normal distribution with mean

 $\xi(\gamma^{(m)}) = \gamma^{(m)} + (h/2)\nabla(\gamma^{(m)})$  and independent coordinates with common variance h. In general, the use of gradient information in the proposal kernel may lead to much faster convergence than for e.g. a random walk Metropolis chain (Roberts and Rosenthal, 1995). Secondly, with probability

$$1 \wedge \frac{f\left(u^{(m+1)}|x\right) \exp\left(-\left|\left|\gamma^{(m)} - \xi\left(u^{(m+1)}\right)\right|\right|^{2}/(2h)\right)}{f\left(\gamma^{(m)}|x\right) \exp\left(-\left|\left|u^{(m+1)} - \xi\left(\gamma^{(m)}\right)\right|\right|^{2}/(2h)\right)}$$

the next state becomes  $\gamma^{(m+1)} = u^{(m+1)}$ ; otherwise  $\gamma^{(m+1)} = \gamma^{(m)}$ . This gives an irreducible and aperiodic Markov chain with the posterior as the stationary distribution, but it is not geometrically ergodic as the posterior has lighter tails than the Gaussian distribution (this can formally be verified using Theorem 4.2 in Roberts and Tweedie, 1996b).

Briefly, the problem with the light tails is that the Markov chain may leave the center of the posterior for a very long time, since  $||\nabla(\gamma)||$  may become extremely large if  $\gamma$  is far away from the mode of the posterior. As suggested in Roberts and Tweedie (1996b) more robust geometric ergodicity properties may be obtained by truncating the gradient in the mean of the proposal kernel: In the Appendix we show that if  $\nabla(\gamma)$  is replaced by

$$\nabla(\gamma)^{trunc} = -\gamma + \left(n_{ij} - \left(H \wedge e^{\tilde{y}_{ij}}\right)A_{ij}\right)_{(i,j) \in I_{ext}} Q^*$$
(21)

for some constant H > 0, then the 'truncated MALA' becomes geometrically ergodic when 0 < h < 2. However, if a sensible value of h is chosen, the undesirable properties of the (untruncated) MALA may not be a problem. In our examples the chain behaved very nicely and a truncation of the gradient (for a suitably large H) would not have made a difference.

Note that  $\Sigma$  and K do not need to be strictly positive definite. This is one reason for using  $\Gamma$  instead of  $\tilde{Y}$  when the posterior is considered. In the case where K is strictly positive definite we have compared MALA's for simulating the conditional distribution of  $\Gamma$  respective  $\tilde{Y}$  given x, where the gradient in the latter case is given by

$$\nabla(\tilde{y}_{ext}) = -(\tilde{y}_{ext} - \tilde{\mu}_{ext})K^{-1} + \left(n_{ij} - e^{\tilde{y}_{ij}}A_{ij}\right)_{(i,j)\in I_{ext}}$$

For the data in Example 1 considered below we found that in the former case the algorithm mixes much faster (Fig. 11), so this is another reason to prefer the 'parametrization' given by  $\Gamma$ .

By simulating the posterior we can obtain MCMC estimates of the posterior mean, credibility intervals, etc. for the Gaussian process and intensity surface. Conditional simulations of the unobserved part  $X \cap W^{(2)}$  of the point process given  $X \cap W^{(1)} = x$  can also be obtained. To do this one generates first a realization from the posterior distribution of the intensity surface and given this realization,  $X \cap W^{(2)}$  is simulated along the same lines as described in the beginning of Section 6.

Maximum a posteriori (MAP) estimation is also possible. Since  $f(\gamma|x)$  is strictly log-concave and its tails tend to zero at infinity, the MAP-estimate  $\gamma^{MAP}$  is the unique solution to  $\nabla(\gamma) = 0$ . Because of the linear relationship between  $\tilde{Y}_{ext}$  and  $\Gamma$ , the MAP-estimate of  $\tilde{Y}_{ext}$  is simply given by  $y_{ext}^{MAP} = \gamma^{MAP}Q + \tilde{\mu}_{ext}$ . Note that the MAPestimate  $y^{MAP}$  of  $\tilde{Y}$  agrees with  $y_{ext}^{MAP}$  restricted to I. It can be shown that  $y_{ext}^{MAP}$ restricted to  $I^{(2)}$  is the same as the predictor of  $\tilde{Y}^{(2)} = \left(\tilde{Y}_{ij}^{(2)}\right)_{(i,j)\in I^{(2)}}$  obtained from 'data'  $\tilde{Y}^{(1)} = \left(\tilde{y}_{ij}^{MAP}\right)_{(i,j)\in I^{(1)}}$  using kriging (see e.g. Cressie, 1993).

The conditional density of the intensity surface  $\tilde{\Lambda}_{ext} = (\exp(\tilde{Y}_{ij}))_{(i,j)\in I_{ext}}$  (with respect to *d*-dimensional Hausdorff measure with carrier space of dimension  $(2(M-1))^2$ ) is not log-concave and  $(\exp(y_{ij}^{MAP}))_{(i,j)\in I}$  is clearly not the MAP-estimate of the intensity surface. If K is strictly positive definite, then using an obvious notation,  $f_{\tilde{Y}_{ext}}(\tilde{y}_{ext}|x)$  is strictly log-concave, and so

$$h(\tilde{y}_{ext}) = f_{\tilde{\Lambda}_{ext}}((\exp(\tilde{y}_{ij}))_{(i,j)\in I_{ext}}|x) = f_{\tilde{Y}_{ext}}(\tilde{y}_{ext}|x) / \prod_{(i,j)\in I_{ext}}\tilde{y}_{ij}$$

is strictly log-concave. Consequently, in this case the MAP-estimate  $\lambda^{MAP}$  of the intensity surface on W is the same as  $\lambda_{ext}^{MAP} = (\exp(\bar{y}_{ij}))_{(i,j)\in I_{ext}}$  restricted to I, where  $\bar{y}$  is the unique solution to  $\nabla(\tilde{y}_{ext}) = (1, \ldots, 1)$ . Note here that since the log Gaussian distribution is heavy tailed and skewed,  $\lambda^{MAP}$  is not necessarily a sensible estimator of the intensity surface (see also the discussion in Example 1 below).

We have used a discrete gradient ascent algorithm for finding  $\gamma^{MAP}$ , since this algorithm involves only the calculation of the gradient: Given an initial value  $\gamma^{(0)}$  the iteration is given by  $\gamma^{(m+1)} = \gamma^{(m)} + \delta \nabla (\gamma^{(m)})$ ,  $m = 0, 1, \ldots$ , where  $\delta > 0$  is a user specified parameter. The algorithm for finding  $\bar{y}$  is similar, except that in each iteration we replace  $\nabla$  by  $\nabla - (1, \ldots, 1)$ . A too high value of  $\delta$  may cause the algorithm to diverge – we used in Example 1 the modest value  $\delta = .1$ .

*Example 1 (continued):* We now consider estimation of the Gaussian process and the intensity surface on a grid  $I = \{0, ..., 64\}^2$  under the log Gaussian Cox process which was estimated in Example 1, Section 7. In this example  $W = W^{(1)} = [0, 1)^2$ .

After some preliminary runs of the MALA the parameter h was adjusted to be .06 in order to obtain an acceptance rate close to the optimal rate .574 given in Roberts and Rosenthal (1995) (they formally prove their results for target distributions with i.i.d. components, but notice that various generalizations are possible and the optimal rate appear to be quite robust over changes in the model). Then a sample of length 300.000 was generated by the MALA and we used a subsample of this (with spacing equal to 10) for obtaining Monte Carlo estimates of the various characteristics of the posterior. These estimates are reported below.

To study the convergence properties and to compare the different implementations of the MALA we have considered various plots of timeseries and estimated autocorrelations for selected cells on the initial as well as the extended lattice. It appears from these plots that the potential problem related to geometrical ergodicity of the MALA is rather hypothetical. As an illustration Fig. 11 shows timeseries and estimated autocorrelations for a subsample of  $\tilde{Y}_{98,98}$  when the invariant distribution of the MALA is either  $\Gamma | x$  or  $\tilde{Y}_{ext} | x$ . In the former case the autocorrelations die out much faster.

Monte Carlo posterior means of the Gaussian process and the intensity surface are shown in Fig. 12. For comparison we have also included Diggle's (1985) nonparametric kernel estimate of the intensity surface. For the uniform kernel given by the uniform density on a disc, the band width of the kernel can be chosen by minimization of an estimate of the mean square error (see Diggle, 1985, and Berman and Diggle, 1989). Instead of the uniform kernel we actually used a planar Epanecnikov kernel since the estimate obtained with this kernel has a more suitable smooth appearance. The band width 0.089 for the planar Epanecnikov kernel was obtained by calibration of the chosen band width for the uniform kernel as suggested in Diggle (1985). The posterior mean of the intensity surface is quite peaked since the minimum and maximum values are 56.86 and 2724.26. This is not surprising recalling the heavytailedness of the log Gaussian distribution. The kernel estimate is less peaked with a range 0-716.37. Integration of the Monte Carlo posterior mean of the intensity surface and the kernel estimate over the unit square yields 125.66 and 126.01, respectively, so the expected number of points for the inhomogeneous Poisson processes with intensity surfaces given by the posterior mean respective the kernel estimate are practically equal and very near to the observed number of points (126). We have also in Fig. 12 included a plot of the logarithm to the Monte Carlo posterior mean of the intensity surface as this gives a better impression of the variability for intermediate values of the posterior mean.

The application of MCMC also facilitates assessment of posterior uncertainty. The estimated posterior variance of the Gaussian process,  $Var(\tilde{Y}_{ij}|x)$ ,  $(i, j) \in I$ , is shown in the left plot in Fig. 13. The largest variance is 1.76 whilst the smallest is .69. By comparing this plot with the Monte Carlo posterior mean of the intensity surface in Fig. 12, we see that the posterior variance is smallest where the posterior mean is largest and vice versa. For the posterior distribution of the intensity surface we have further for selected cells  $D_{ij}$  estimated the 10% and 90% quantiles. These credibility intervals are shown in Table 2 when (i, j) are given by (13a, 13b),  $a, b = 1, \ldots, 4$ , and (14, 36). The credibility interval for  $D_{14,36}$  is largest as this cell is situated in a peak of  $E(\tilde{\Lambda}_{ij}|x)$ .

As an illustration of the simulation method on the extended lattice and the effect of wrapping the extended Gaussian field on a torus, the right plot in Fig. 13 shows the Monte Carlo posterior mean of  $\tilde{Y}_{ext}$ . Notice that outside the original field and away from the boundaries the estimated posterior mean is constant and equal to the unconditional mean.

Finally, we have considered MAP-estimation of the Gaussian process and the intensity process. The extended matrix K was strictly positive definite, and  $y_{ext}^{MAP}$  and  $\lambda_{ext}^{MAP}$  were obtained by iterating the discrete gradient ascent algorithm until the gradient was practically zero (i.e. until its coordinates were numerically less than  $10^{-5}$ ). The minimum and maximum values of  $y^{MAP}$  are 3.53 and 7.96, while the corresponding values of the estimated posterior means  $E(\tilde{Y}_{i,j}|x), (i,j) \in I$ , are 3.24

and 7.59. Actually,  $y^{MAP}$  is very similar to these posterior means, so we have omitted the plot of  $y^{MAP}$ . Since  $\max_{(i,j)\in I} \lambda_{ij}^{MAP} \approx 10^{-15}$  the MAP-estimate is clearly a totally unreasonable estimate of the intensity surface. This may as noted before be due to the skewness and heavytailedness of the log Gaussian distribution combined with the fact that the intensity surface is a random field of correlated log Gaussian random variates.  $\Box$ 

In Example 1 the posterior mean and the nonparametric kernel estimate gave very different estimates of the intensity surface. To study these estimators under known conditions we simulated a point pattern on the unit square from the log Gaussian Cox process with exponential correlation function and parameters  $\mu = 4$ ,  $\sigma^2 = 2$ ,  $\beta = .1$ . Using the same procedure as in Example 1, Section 7, the estimates of  $\mu$ ,  $\sigma^2$ ,  $\beta$  are 3.78, 2.46, .077, and the procedure for choosing the bandwidth yields .061. Plots of the true intensity surface, the Monte Carlo posterior mean of the intensity surface under the estimated model, and the kernel estimate are shown in Fig. 14. In this case the two estimates look much more similar than in Example 1. The large difference between the intensity surface estimates in Example 1 may be explained by the considerably larger bandwidth which was used in the kernel estimate in Example 1, and which yielded an oversmoothed estimate of the intensity surface. In Fig. 14 the range of the true intensity surface, the Monte Carlo posterior mean, and the kernel estimate are .44-7802.62, 21.27-7653.15, and 0-4898.43, respectively. Integration of the estimates give 150.06 for the posterior mean and 147.69 for the kernel estimate, while the integral of the true surface is 158.32, and the true and estimated intensity are  $\rho = 148.41$  and  $\hat{\rho} = 150$ .

In conclusion, at least for the particular cases of Example 1 and the simulation study, the posterior mean seems to be the better estimate.

#### Acknowledgments

This research will be a part of the second and third authors Ph.D. Dissertations. It has been funded by the Danish Informatics Network in the Agricultural Sciences, the Danish Natural Science Research Council, NORFA and the Research Council of Norway. Antti Penttinen kindly provided the data studied in Examples 1 and 2. We thank Anders Brix, Poul Svante Eriksen, Peter Green, Jørgen Hoffmann-Jørgensen, Steffen L. Lauritzen, Antti Penttinen, Gareth Roberts, Mats Rudemo, Dietrich Stoyan, the editor and two anonymous referees for helpful comments.

# Appendix: Geometrical ergodicity of the truncated MALA

Below we prove that the truncated MALA discussed in Section 8 is geometrically ergodic when 0 < h < 2. For simplicity and without loss of generality we shall assume that  $\mu = 0$ . Letting  $\tilde{\gamma} = \gamma Q$ , then by (20) the logarithm of the posterior density is

$$\log f(\gamma|x) = \operatorname{const}(x) - \frac{1}{2} ||\gamma||^2 + n\tilde{\gamma}^* - \sum_{(i,j)\in I_{ext}} e^{\tilde{\gamma}_{ij}} A_{ij}$$

where  $n = (n_{ij})_{(i,j) \in I_{ext}}$ , and by (21) the truncated gradient is

$$\nabla(\gamma)^{trunc} = -\gamma + (n - R(\tilde{\gamma}))Q^*$$

where

$$R(\tilde{\gamma}) = \left( \left( H \wedge e^{\gamma_{ij}} \right) A_{ij} \right)_{(i,j) \in I_{ext}}$$

In the following g will denote a measurable function mapping  $\mathbb{R}^d$  into  $\mathbb{R}$ . The results concerning geometrical ergodicity are the following:

**Theorem 4.** When 0 < h < 2 the truncated MALA is  $V_s$ -uniformly ergodic for  $V_s(\gamma) = \exp(s \|\gamma\|)$  and any  $s \ge 0$ , i.e., there exist  $0 < R_s < \infty$  and  $0 < \rho_s < 1$  such that for any  $\gamma \in \mathbb{R}^d$ ,

$$\left\| P^{(m)}(\gamma, \cdot) - \int_{\cdot} f(\dot{\gamma}|x) d\dot{\gamma} \right\|_{V_s} \le V_s(\gamma) R_s \rho_s^m, \ m \ge 1.$$

Here  $P^{(m)}$  denotes the *m*-step transition kernel of the truncated MALA and for a signed measure  $\mu$ ,  $\|\mu\|_{V_s} = \sup_{|g| \leq V_s} |\int g(\dot{\gamma})\mu(d\dot{\gamma})|$ .

**Corollary.** Suppose that  $|g(\gamma)| \leq \exp(s ||\gamma||/2)$ ,  $\gamma \in \mathbb{R}^d$ , and that  $(\Gamma_l)_{l\geq 0}$  is generated from the truncated MALA where 0 < h < 2 and the initial distribution of  $\Gamma_0$  is arbitrary. Define the Monte Carlo approximation  $\bar{\xi}_m = \sum_{1}^m g(\Gamma_l)/m$  of the mean  $\xi = Eg(\Gamma_0)$ calculated for the stationary chain. Assuming first that the density of  $\Gamma_0$  is  $f(\cdot|x)$ , then

$$\sigma_g^2 := \lim_{m \to \infty} m Var(\bar{\xi}_m) = Var(g(\Gamma_0)) + 2\sum_{l=1}^m Cov(g(\Gamma_0), g(\Gamma_l)) < \infty$$

Moreover, if  $\sigma_g^2 > 0$ , we have a central limit theorem independently of the chosen initial distribution of  $\Gamma_0$ :

$$(\bar{\xi}_m - \xi) / \sqrt{\sigma_g^2 / m} \xrightarrow{\mathcal{D}} N(0, 1) \text{ as } m \to \infty.$$

*Proof.* Let  $c(\gamma) = \gamma + (h/2)\nabla(\gamma)^{trunc}$  and let  $q(\gamma, \cdot)$  be the density of  $N_d(c(\gamma), hI_d)$ . Then (36) in Roberts and Tweedie (1996b) holds since  $n - R(\tilde{\gamma})Q^*$  is bounded. Since the proof of Theorem 4.1 in Roberts and Tweedie (1996b) is also applicable in our situation, the geometrical ergodicity then follows if we can show that the truncated MALA 'converges inwards'. More precisely let as in Roberts and Tweedie (1996b),  $I(\gamma) = \{\dot{\gamma} : \|\dot{\gamma}\| \le \|\gamma\|\}$  and  $A(\gamma) = \{\dot{\gamma} : f(\gamma|x)q(\gamma,\dot{\gamma}) \le f(\dot{\gamma}|x)q(\dot{\gamma},\gamma)\}$ . Then we need to show that

$$\int_{A(\gamma) \triangle I(\gamma)} q(\gamma, \dot{\gamma}) d\dot{\gamma} \to 0 \ as \ \|\gamma\| \to \infty.$$

Here  $\triangle$  denotes symmetric difference, i.e.  $A \triangle B = [A \setminus B] \cup [B \setminus A]$ .

Note that for any  $\epsilon > 0$  we can choose  $S_{\epsilon}$  such that  $\int_{\mathbb{R}^d \setminus B_{\epsilon}(\gamma)} q(\gamma, \dot{\gamma}) d\dot{\gamma} < \epsilon$  where  $B_{\epsilon}(\gamma) = \{\dot{\gamma} : \|\dot{\gamma} - c(\gamma)\| < S_{\epsilon}\}$ . Thus

$$\int_{A(\gamma) \triangle I(\gamma)} q(\gamma, \dot{\gamma}) d\dot{\gamma} \le \int_{(A(\gamma) \triangle I(\gamma)) \cap B_{\epsilon}(\gamma)} q(\gamma, \dot{\gamma}) d\dot{\gamma} + \epsilon.$$

Since

$$\dot{\gamma} \in B_{\epsilon}(\gamma) \Rightarrow \|\dot{\gamma}\| \le \|\dot{\gamma} - c(\gamma)\| + \|c(\gamma)\| \le |1 - h/2| \|\gamma\| + S_{\epsilon} + constant$$

we have for  $\|\gamma\|$  sufficiently large that  $B_{\epsilon}(\gamma) \subseteq I(\gamma)$  so that  $(A(\gamma) \triangle I(\gamma)) \cap B_{\epsilon}(\gamma) = B_{\epsilon}(\gamma) \setminus A(\gamma)$ . It is therefore enough to show that when  $\|\gamma\|$  is sufficiently large then  $\dot{\gamma} \in B_{\epsilon}(\gamma)$  implies that  $\dot{\gamma} \in A(\gamma)$ .

It is straightforwardly seen that the inequality which defines the set  $A(\gamma)$  is equivalent to  $J_1 + J_2 + J_3 + J_4 + J_5 \ge 0$  where

$$J_{1} = (h/8) \left( ||\gamma||^{2} - ||\dot{\gamma}||^{2} \right), J_{2} = (h/8) \left( ||(n - R(\tilde{\gamma}))Q^{*}||^{2} - \left| \left| \left( n - R\left(\tilde{\dot{\gamma}}\right) \right)Q^{*} \right| \right|^{2} \right), J_{3} = \sum_{(i,j)\in I_{ext}} A_{ij} \left( e^{\tilde{\gamma}_{ij}} - e^{\tilde{\dot{\gamma}}_{ij}} \right), J_{4} = \left( \tilde{\gamma} - \tilde{\dot{\gamma}} \right) \left( R(\tilde{\gamma}) + R\left(\tilde{\dot{\gamma}}\right) \right)^{*} / 2, J_{5} = (h/4) \left[ \tilde{\dot{\gamma}} \left( n - R\left(\tilde{\dot{\gamma}}\right) \right)^{*} - \tilde{\gamma} (n - R(\tilde{\gamma}))^{*} \right]$$

and  $\tilde{\dot{\gamma}} = \dot{\gamma}Q$ . If  $\dot{\gamma} \in B_{\epsilon}(\gamma)$  and  $\|\gamma\|$  is sufficiently large, then  $J_1 + J_2 \ge 0$  since  $J_2$  is bounded and 0 < h < 2. We therefore just need to show that  $J_3 + J_4 + J_5 \ge 0$  for  $\dot{\gamma} \in B_{\epsilon}(\gamma)$  and  $\|\gamma\|$  sufficiently large.

If  $\|\gamma\| \to \infty$  then also  $\|\tilde{\gamma}\| \to \infty$  because Q is of full rank. Furthermore, if  $\dot{\gamma} \in B_{\epsilon}(\gamma)$ , then  $\tilde{\dot{\gamma}}_{ij} = (1 - h/2)\tilde{\gamma}_{ij} + (v(\dot{\gamma})Q)_{ij}$  where  $v(\dot{\gamma})$  is a uniformly bounded vector. Since 0 < h < 2, we have therefore that  $\tilde{\gamma}_{ij} \to \infty$  implies that  $\tilde{\dot{\gamma}}_{ij} \to \infty$ , while  $\tilde{\gamma}_{ij} \to -\infty$  implies that  $\tilde{\dot{\gamma}}_{ij} \to -\infty$ , where in both cases  $|\tilde{\gamma}_{ij}| \to \infty$  at a rate faster than  $|\tilde{\dot{\gamma}}_{ij}|$ , since  $\tilde{\gamma}_{ij} - \tilde{\dot{\gamma}}_{ij}$  is of the order  $(h/2)\tilde{\gamma}_{ij}$  asymptotically. Let now  $B = \{(i, j) \in I_{ext} : \|\tilde{\gamma}_{ij}\| \neq \infty\}$ . Then  $J_3 + J_4 + J_5$  can be written as

$$\sum_{\substack{(i,j)\in I_{ext}\setminus B}} \{A_{ij}\left(e^{\tilde{\gamma}_{ij}}-e^{\tilde{\gamma}_{ij}}\right) + \left(\tilde{\gamma}_{ij}-\tilde{\gamma}_{ij}\right)[HA_{ij}e^{\tilde{\gamma}_{ij}}/\left(H\vee e^{\tilde{\gamma}_{ij}}\right) + HA_{ij}e^{\tilde{\gamma}_{ij}}/\left(H\vee e^{\tilde{\gamma}_{ij}}\right)]/2 + (h/4)[\tilde{\gamma}_{ij}\left(n_{ij}-HA_{ij}e^{\tilde{\gamma}_{ij}}/\left(H\vee e^{\tilde{\gamma}_{ij}}\right)\right) - \tilde{\gamma}_{ij}\left(n_{ij}-HA_{ij}e^{\tilde{\gamma}_{ij}}/\left(H\vee e^{\tilde{\gamma}_{ij}}\right)\right)]\} + H\left(B,\tilde{\gamma},\tilde{\gamma}\right)$$

where  $H(B, \tilde{\gamma}, \tilde{\dot{\gamma}})$  is a finite sum of bounded terms. Since for each  $(i, j) \in I_{ext} \setminus B$  the corresponding term  $\{\ldots\}$  in the sum converges to  $\infty$  when  $\|\tilde{\gamma}\| \to \infty$ , the proof of the theorem is completed. The corollary is a direct consequence of Theorem 4.1 in Roberts and Tweedie (1996a).

# References

Adler, R. (1981). The Geometry of Random Fields. Wiley, New York.

Arjas, E. and Gasbarra, D. (1994). Nonparametric Bayesian inference from right censored survival data, using the Gibbs sampler. *Statistica Sinica* **4**, 505–524.

- Baddeley, A.J. and Van Lieshout, M.N.M. (1993). Stochastic geometry models in highlevel vision. In K. Mardia and G.K. Kanji (eds.) *Statistics and images, Advances in Applied Statistics, a supplement to J. Appl. Statist.* **20**, 231–256.
- Baddeley, A.J. and Van Lieshout, M.N.M. (1995). Area-interaction point processes. Ann. Inst. Statist. Math. 47, 601–619.
- Baddeley, A.J., Van Lieshout, M.N.M. and Møller, J. (1996). Markov properties of cluster processes. *Adv. Appl. Prob. (SGSA)* **28**, 346–355.
- Baddeley, A. J. and Møller, J. (1989). Nearest-neighbour Markov point processes and random sets. *Int. Statist. Rev.* 57, 89–121.
- Bartlett, M.S. (1964). Spectral analysis of two-dimensional point processes. *Biometrika* 44, 299–311.
- Berman, M. and Diggle, P.J. (1989). Estimating weighted integrals of the second-order intensity of a spatial point process. J. R. Statist. Soc. B 51, 81–92.
- Besag, J.E. (1974). Spatial interaction and the statistical analysis of lattice systems. J. Roy. Statist. Soc. B 36, 192–225.
- Besag, J.E. (1977). Some methods of statistical analysis for spatial data. *Bull. Internat. Statist. Inst.* **47**, 77–92.
- Besag, J.E. (1994). Discussion of the paper by Grenander and Miller. J. R. Statist. Soc. B 56, 591–592.
- Christakos, G. (1984). On the problem of permissible covariance and covariogram models. *Water Resources Research* **20**, 251–265.
- Christakos, G. (1992). *Random Field Models in Earth Sciences*. Academic Press, San Diego.
- Cressie, N. (1991). Statistics for Spatial Data. Wiley, New York.
- Daley, D.J. and Vere-Jones, D. (1988). *An Introduction to the Theory of Point Processes*. Springer-Verlag, New York.
- Davis, P.J. (1979). Circulant Matrices. Wiley, 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. Statist.* **34**, 138–147.
- Diggle, P.J. and Milne, R.K. (1983). Bivariate Cox processes: Some models for bivariate spatial point processes. J. R. Statist. Soc. B 45, 11–21.

- Gelfand, A.E. and Carlin, B.P. (1991). Maximum likelihood estimation for constrained or missing data models. Research Report 91–002, Division of Biostatistics, University of Minnesota.
- Georgii, H.-O. (1988). *Gibbs Measures and Phase Transitions*. Walter de Gruyter, Berlin.
- Geyer, C.J. (1994). On the convergence of Monte Carlo maximum likelihood calculations. J. R. Statist. Soc. B 56, 261–274.
- Grandell, J. (1976). *Doubly Stochastic Poisson Processes*. Lecture Notes in Mathematics, 529. Springer-Verlag, Berlin.
- Granville, V. and Smith, R.L. (1995). Clustering and Neyman-Scott process parameter simulation via Gibbs sampling. (Manuscript.) Statistical Laboratory, University of Cambridge.
- Grenander, U. and Miller, M.I. (1994). Representations of knowledge in complex systems (with discussion). J. R. Statist. Soc. B 56, 549–603.
- Häggström, O., Van Lieshout, M.N.M. and Møller, J. (1996). Characterisation results and Markov chain Monte Carlo algorithms including exact simulation for some spatial point processes. Research Report R-96–2040, Department of Mathematics, Aalborg University. (Submitted for publication.)
- Heikkinen, J. and Arjas, E. (1996). Nonparametric Bayesian estimation of a spatial Poisson intensity. Preprint 20, Department of Statistics, University of Jyvaskyla. (Submitted for publication.)
- Jensen, J.L. and Møller, J. (1991). Pseudolikelihood for exponential family models of spatial point processes. *Annals of Applied Probability* **1**, 445–461.
- Karr, A.F. (1991). *Point Processes and Their Statistical Inference*. (2nd ed.) Marcel Dekker, New York.
- Kuuluvainen, T., Penttinen, A., Leinonen, K, and Nygren, M. (1996). Statistical opportunities for comparing stand structural heterogeneity in managed and primeval forests: an example from boreal spruce forest in southern Finland. *Silvia Fennica* 30, 315–328
- Lantuéjoul, C. (1994). Nonconditional simulation of stationary isotropic multigaussian random functions. In *Geostatistical Simulations*, (eds. M. Armstrong and P. Dowd) Kluwer Academic Publishers, Dordrecht.
- Lawson, A.B. (1993). Discussion contribution to the The Gibbs sampler and other Markov chain Monte Carlo methods. J. R. Statist. Soc. B 55, 61–62.

- Ledoux, M. and Talagrand, M. (1991). *Probability in Banach spaces*. Springer-Verlag, Berlin.
- Lotwick, H.W. and Silverman, B.W. (1982). Methods for analysing spatial processes of several types of points. J. R. Statist. Soc. B 44, 406-413.
- Mantouglou, A. and Wilson, J. L. (1982). The turning bands method for simulation of random fields using line generation by a spectral method. *Water Resources Research* **18** (5), 1379–1394.
- Matérn, B. (1960). *Spatial Variation*. Meddelanden från Statens Skogsforskningsinstitut, Vol. 49 (5). Statens Skogsforskningsinstitut, Stockholm.
- Matheron, G. (1973). The intrinsic random functions and their applications. *Adv. Appl. Prob.* **5**, 439–468.
- Møller, J. (1994). Markov chain Monte Carlo and spatial point processes. Research Report 293, Department of Theoretical Statistics, University of Aarhus. To appear in O.E. Barndorff-Nielsen *et al.* (eds.): *Proc. Seminaire Européen Statistique Toulouse 1996 "Current trends in stochastic geometry with applications"*, Chapman and Hall.
- Ogata, Y. and Katsura, K. (1988). Likelihood analysis of spatial inhomogeneity for marked point patterns. J. Am. Statist. Ass. 40, 29–39.
- Penttinen, A., Stoyan, D. and Henttonen H.M (1992). Marked point processes in forest statistics. *Forest Science* **38** (4), 806–824.
- Preston, C. (1976). *Random Fields*. Lecture Notes in Mathematics, 534. Springer-Verlag, Berlin.
- Rathbun, S.L. and Cressie, N. (1994). A space-time survival point process for a longleaf pine forest in Southern Georgia. J. Am. Statist. Ass. 89, 1164–1174.
- Ripley, B.D. (1977). Modelling spatial patterns (with discussion). J. R. Statist. Soc. B 39, 172–212.
- Ripley, B.D. (1987). Stochastic Simulation. Wiley, New York.
- Ripley, B.D. (1988). *Statistical Inference for Spatial Processes*. Cambridge University Press, Cambridge.
- Roberts, G.O. and Rosenthal, J.S. (1995). Optimal scaling of discrete approximations to Langevin diffusions. Research Report no 95–11, Statistical Laboratory, Cambridge University.

- Roberts, G.O. and Tweedie, R.L. (1996a). Geometric convergence and central limit theorems for multidimensional Hastings and Metropolis algorithms. *Biometrika* 83, 95-110.
- Roberts, G.O. and Tweedie, R.L. (1996b). Exponential convergence of Langevin diffusions and their discrete approximations. *Bernoulli* **2**, 341–363.
- Stoyan, D., Kendall, W.S. and Mecke, J. (1995). *Stochastic Geometry and Its Applications.* 2nd ed. Wiley, Chichester.
- Stoyan, D. and Stoyan, H. (1994). *Fractals, Random Shapes and Point Fields*. Wiley, Chichester.
- Thomas, M. (1949). A generalization of Poisson's binomial limit for use in ecology. *Biometrika* **36**, 18–25.
- Wackernagel, H. (1995). Multivariate Geostatistics. Springer-Verlag, Berlin.
- Widom, B. and Rowlinson, J.S. (1970). New models for the study of liquid-vapor phase transitions. J. Chem. Physics 52, 1670–1684.
- Wood, A.T.A. and Chan, G. (1994). Simulation of stationary Gaussian processes in  $[0,1]^d$ . J. Computational and Graphical Statistics **3**, 409–432.
- Yaglom, A.M. (1986). Correlation Theory of Stationary and Related Random Functions I. Springer, Berlin.

Jesper Møller, Department of Mathematics, Aalborg University, Fredrik Bajers Vej 7E, DK- 9220 Aalborg Ø, Denmark. Email: jm@math.auc.dk

Anne Randi Syversveen, Department of Mathematical Sciences, The Norwegian University of Science and Technology, N-7034 Trondheim, Norway. Email: annerand@math.ntnu.no

Rasmus Plenge Waagepetersen, Department of Theoretical Statistics and Operations Research, Departments of Mathematical Sciences, University of Aarhus, DK-8000 Århus C, Denmark. Email: rasmus@mi.aau.dk 1. Gaussian: $\exp(-(a/\beta)^2)$ 5. Hyperbolic: $(1 + a/\beta)^{-1}$ 2. Exponential: $\exp(-a/\beta)$ 6. Bessel: $\exp(-(a/\beta)^2)J_0(a/\beta)$ 3. Cardinal sine: $\sin(a/\beta)/(a/\beta)$ 7. Spherical: $1(a/\beta < 1) \cdot [(a/\beta)^3/2 + 1 - (3a/(2\beta))]$ 4. Stable: $\exp(-\sqrt{a/\beta})$  $1 - (3a/(2\beta))]$ 

Table 1. Correlation functions.  $J_0$  is the Bessel function of the first kind of order zero.

	i = 13	i = 26	i = 39	i = 52
j = 52	7.0-181.3	26.6-546.2	7.4-190.4	28.6-541.3
	78.1	235.0	83.0	234.5
j = 39	7.9-207.5	8.8-218.5	8.0-215.3	8.0-207.3
	89.9	95.7	93.0	89.0
j = 26	6.5-173.3	11.5-282.0	11.6-282.9	6.7-170.4
	74.3	119.2	119.8	73.2
j = 13	17.2-380.3	12.9-311.4	10.5-248.4	5.0-138.2
	163.2	130.6	105.2	60.7
(i, j) =	363.3-3733.3			
(36, 14)	1734.1			

Table 2. Example 1. Estimated 80%-credibility intervals and posterior means of the intensity surface at selected cells  $C_{ij}$  organized in accordance with Figure 11, where (i, j) = (0, 0), (0, 64), (64, 0), (64, 64) correspond to the lower left, upper left, lower right, and upper right cells.



Fig. 1. Upper row, a)-d): Various pair correlation functions with varying values of  $\beta$  (solid line = smallest value of  $\beta$ ) when  $\sigma = 1$ . Lower row: e), f) pair correlation functions for the Thomas and Matérn cluster processes. g), h) Gaussian and exponential correlation functions with  $\beta$  as in the upper row.



Fig. 2. Simulated realizations of Gaussian random fields with  $\sigma = 1$ . Left to right: Gaussian  $\beta = 0.172$ , exponential  $\beta = 0.143$ , cardinal sine  $\beta = 0.094$ , stable  $\beta = 0.071$ .



Fig. 3. Simulations of log Gaussian Cox processes conditional on that the number of points is 148. First column: Gaussian correlation function. Second column: Exponential. Third column: Cardinal sine. Fourth column: Stable. First row: Same values of parameters as in Figure 2, i.e.  $\sigma^2 = 1$  and  $\beta = 0.172$ , 0.143, 0.094, 0.071 (left to right). Second row:  $\sigma^2 = 2.4$ ,  $\beta = 0.110$ , 0.100, 0.053, 0.049. Third row:  $\sigma^2 = 2.4$ ,  $\beta = 0.057$ , 0.050, 0.027, 0.020. Mean and variance of the number of points are equal in each row.



Fig. 4. Left: Plot of Gaussian correlation function (solid) and  $\ln g_{\rm T}(\cdot)$  (dotted line) for  $\kappa = 0.001$ . Middle: Hankel transform of  $\ln g_{\rm T}(\cdot)$  for  $\kappa = 0.001$ . Right: Hankel transform of  $\ln g_{\rm M}(\cdot)$  for R = 0.1.



Fig. 5. Left: Dotted lines: Average and envelopes for the nonparametric estimator of F based on 100 simulations of the Thomas process. Solid lines: The same but for the log Gaussian Cox process with Gaussian correlation function. Right: The same as the left plot but with F substituted by G.



Fig. 6. Left: Gaussian random field with exponential correlation function,  $\sigma^2 = 2$  and  $\beta = 0.1$ . Middle: bivariate log Gaussian Cox process,  $\mu_1 = \mu_2 = 2.5$ ,  $\sigma_1 = \sigma_2 = 2$ . Right: Bivariate log Gaussian Cox process,  $\mu_1 = \mu_2 = 2.5$ ,  $\sigma_1 = -\sigma_2 = 2$ .



Fig. 7. Left: Estimated parameters  $\beta$  and  $\sigma^2$  from 20 simulations under the log Gaussian Cox process with exponential covariance function  $c(t) = 2.0 \exp(-20t)$ . The true parameter value is marked with a square. Right: The true covariance function (dotted line), the mean and upper and lower envelopes for the estimated covariance functions (solid lines).



Fig. 8. Example 1. Several characteristics for the pine data (see the text for explanations).



Fig. 9. Example 1. Estimate of z based on the data (solid line) and 'conditional' envelopes (— — — ) and 'unconditional' envelopes (- - - -) based on 20 simulations. Left: Log Gaussian Cox process. Right: Matérn cluster process.



Fig. 10. Example 2. a) Plot of data, spruces marked with '·' and birches marked with 'x'. b) Empirical covariance functions (solid line), from top  $\hat{c}_{22}$ ,  $\hat{c}_{11}$ ,  $\hat{c}_{12}$ , and covariance functions for the fitted model (dotted line). c) Empirical  $\hat{F}_{22}$  (solid line) together with lower and upper envelopes (dotted line) plotted against the mean of 99 simulations from the fitted model.



Fig. 11. Example 1. Upper plots: Timeseries (left) and estimated autocorrelations (right) for  $\tilde{Y}_{98,98}|x$  obtained by transforming a subsample of  $\Gamma_{98,98}|x$  (spacing = 10) generated by MALA. Lower row: Same as upper row but no transformation is used, i.e.  $\tilde{Y}_{98,98}|x$  is generated directly by MALA.



Fig. 12. Example 1. Upper left plot: Monte Carlo posterior mean of the Gaussian field. Upper right plot: Monte Carlo posterior mean of the intensity surface. Lower left plot: Logarithm to the upper right plot. Lower right plot: Diggle's nonparametric kernel estimate of the intensity surface.



Fig. 13. Example 1. Left: Monte Carlo posterior variance of the Gaussian field on the original lattice. Right: Monte Carlo posterior mean of the Gaussian field on the extended lattice.



Fig. 14. Simulation study. Upper left plot: True Gaussian surface. Upper right plot: True intensity surface. Lower left plot: Monte Carlo posterior mean of the intensity surface. Lower right plot: Diggle's nonparametric kernel estimate.