# A Marked Point Process Perspective in Fitting Spatial Point Process Models

Yongtao Guan\*

May 24, 2007

# ABSTRACT

This paper discusses a new perspective in fitting spatial point process models. Specifically the spatial point process of interest is treated as a marked point process where at each observed event  $\mathbf{x}$  a stochastic process  $M(\mathbf{x};t)$ , 0 < t < r, is defined. Each mark process  $M(\mathbf{x};t)$  is compared with its expected value, say  $F(t;\theta)$ , to produce a discrepancy measure at  $\mathbf{x}$ , where  $\theta$  is a set of unknown parameters. All individual discrepancy measures are combined to define an overall measure which will then be minimized to estimate the unknown parameters. The proposed approach can be easily applied to data with sample size commonly encountered in practice. Simulations and an application to a real data example demonstrate the efficacy of the proposed approach.

KEY WORDS: K-function, Marked Point Process, Spatial Point Process.

<sup>\*</sup>Yongtao Guan is Assistant Professor, Division of Biostatistics, Yale School of Public Health, Yale University, New Haven, CT 06520-8034, e-mail: yongtao.guan@yale.edu, phone: (203) 785-6125, fax: (203) 785-2838. The author thanks the Editor, the Associate Editor and a referee for their helpful comments that have improved the paper. This research was partially supported by National Science Foundation grant DMS-0603673

## 1. INTRODUCTION

In spatial point pattern analysis, a common model fitting procedure is the "minimum contrast estimation" (MCE) procedure (e.g., Møller and Waagepetersen 2003), where some theoretical and empirical summary descriptions of the process are obtained and compared. Among these summary descriptions, the K-function (see Section 2 for the definition) is perhaps the most frequently used since its closed-form theoretical expression is available for many processes (see, e.g., Diggle 2003). Throughout this article, let  $K(t; \theta)$  and  $\hat{K}(t)$ denote the theoretical and empirical K-functions, respectively, where t is the inter-point distance (i.e., the lag) and  $\theta$  is the unknown parameter vector determining the distribution of the underlying spatial point process. The estimate of  $\theta$ ,  $\hat{\theta}$ , is defined as the minimizer of the following discrepancy measure:

$$U(\theta) = \int_0^r [\hat{K}^c(t) - K^c(t;\theta)]^2 dt,$$
(1)

for some constants c and r which are often referred to as tuning parameters. The constant r in (1) defines the lag set over which the theoretical and empirical K-functions are compared, where the constant c is used to stabilize the sampling fluctuations in  $\hat{K}(t)$ .

The main attractiveness of using (1) is its computational simplicity which allows one to quickly fit and explore a range of possible models. Despite this computational attractiveness, however, it has been long criticized due to the arbitrariness associated with the selection of the tuning parameters r and c (see, e.g., Cressie 1993). The optimal choice of these parameters, in particular c, is highly dependent on the underlying spatial point process. For example, Diggle (2003) recommended using c = 0.25 for aggregated (i.e., clustered) point patterns but c = 0.5 for inhibitive (i.e., regular) point patterns. Guan and Sherman (2007) found that a good choice of c also depends on the value of r. In particular, they found that the recommended value c = 0.25 was not "optimal" for some cluster processes since  $c \approx 1$ , when combined with an appropriately selected r value, often led to much improved results than c = .25. Clearly the lack of a model-free choice for the tuning parameters is undesirable.

To avoid the selection of c, Diggle (2003) proposed the following alternative to (1):

$$U(\theta) = \int_{0}^{r} w(t) [\hat{K}(t) - K(t;\theta)]^{2} dt,$$
(2)

where w(t) is a weight function that is inversely proportional to the variance of  $\hat{K}(t)$ . For Poisson processes it is suggested to use  $w(t) = t^{-2}$ . In general, however, the variance of  $\hat{K}(t)$  is difficult to obtain which often makes (2) even more difficult to work with than (1). Guan and Sherman (2007) proposed a subsampling approach to estimate the variance of  $\hat{K}(t)$  and to subsequently obtain w(t). However, their approach requires a relatively large sample size and thus may not be appropriate for many data commonly encountered in practice with relatively small sample size. In addition, the use of subsampling necessarily involves the choice of a new tuning parameter, that is, the subblock size. Although it is expected that the choice of the subblock size does not influence the estimation greatly, in general an "optimal" value of it is often difficult to obtain.

This article looks at the use of (2) from a new perspective. Specifically the spatial point process is treated as a marked point process where at each observed event  $\mathbf{x}$  a stochastic process  $M(\mathbf{x};t)$ , 0 < t < r, is also observed. This is related to the recent work of Loh and Stein (2004) where they assigned one single "mark" variable (but not a stochastic process as being considered here) to each event. The mark process  $M(\mathbf{x};t)$  will be defined in Section 2. In what follows let N denote the spatial point process of interest and  $\lambda$  be the first-order intensity of the process. It will be shown that  $E[M(\mathbf{x};t)] = \lambda K(t;\theta)$ . This new perspective allows one to form individual discrepancy measures that are analogous to (2) by comparing each mark process  $M(\mathbf{x};t)$  with the target function  $\lambda K(t;\theta)$ . All the individual discrepancy measures can then be combined together to define an overall measure. A simple expression for the weight function w(t) will be given. In contrast to Guan and Sherman's approach, the calculation of w(t) does not require any additional

tuning parameters and thus is more objective. Furthermore, the proposed approach can be applied without difficulty to both small and large data sets.

The rest of the article is organized as follows. Section 2 formally introduces the proposed method and Section 3 studies the asymptotic properties of the resulting estimator. Section 4 contains a simulation study whereas Section 5 presents an application to some real data.

## 2. THE PROPOSED METHOD

Throughout this section, assume that N is a stationary and isotropic spatial point process. For an arbitrary Borel set in  $\mathbb{R}^2$ , let |B| denote the area of B and N(B) denote the number of points from N in B. Let  $d\mathbf{x}$  denote an infinitesimal region which contains  $\mathbf{x} \in \mathbb{R}^2$  and  $||\cdot||$  denote the Euclidean norm. The second-order intensity function (SIF) of N is defined as (e.g., Diggle 2003)

$$\lambda_2(\mathbf{x}, \mathbf{y}) \equiv \lim_{|d\mathbf{x}|, |d\mathbf{y}| \to 0} \left\{ \frac{\mathrm{E}[N(d\mathbf{x})N(d\mathbf{y})]}{|d\mathbf{x}||d\mathbf{y}|} \right\}$$

Clearly  $\lambda_2(\mathbf{x}, \mathbf{y}) = \lambda_2(||\mathbf{x} - \mathbf{y}||)$  due to isotropy, where  $\lambda_2(\cdot)$  now is a function defined on  $\mathbb{R}$ . Let  $\lambda$  be the first-order intensity of the process. In relation to the SIF, the *K*-function can be expressed as:

$$\lambda^2 K(t) = 2\pi \int_0^t u \lambda_2(u) du$$

In what follows, let  $J(t) \equiv \lambda^2 K(t)$  and let D denote the domain of interest. The empirical versions of J(t), ignoring edge effect, can be defined as:

$$\hat{J}(t) = \frac{1}{|D|} \sum_{\mathbf{x}, \mathbf{y} \in N \cap D, \mathbf{y} \neq \mathbf{x}} I(||\mathbf{y} - \mathbf{x}|| \le t),$$
(3)

where  $I(\cdot)$  is an indicator function. Guan and Sherman (2007) showed that (3) is consistent for J(t) and is asymptotically normal under suitable conditions.

From (3) it can be seen that each event  $\mathbf{x} \in N \cap D$  contributes to the calculation of  $\hat{J}(t)$  with the number of neighboring events in N that are within t distance of itself. In

view of this observation, the point process N can be treated as a marked point process. Specifically, let the events of N be the points of the new process. For each point  $\mathbf{x} \in N$ , define the mark value at  $\mathbf{x}$  as:

$$M(\mathbf{x};t) = \sum_{\mathbf{y}\in N\cap D, \mathbf{y}\neq \mathbf{x}} I(||\mathbf{y}-\mathbf{x}|| \le t).$$

Note that if |D| is replaced by N(D), then (3) is simply the sample average of marks observed at points within D. Ignoring edge effect, a simple application of Campbell's theorem (Stoyan and Stoyan, 1994) and conditional probability yields that

$$E[M(\mathbf{x};t)|\mathbf{x} \in N] = 2\pi \int_0^t u\lambda_2(u)du/\lambda = \lambda K(t)$$

Let  $\sigma^2(\mathbf{x};t) \equiv Var[M(\mathbf{x};t)|\mathbf{x} \in N]$ . Note that  $\sigma^2(\mathbf{x};t)$  is independent of  $\mathbf{x}$  due to stationarity. For simplicity, let  $\sigma^2(t)$  denote  $\sigma^2(\mathbf{x};t)$ . A "method-of-moment" estimator of  $\sigma^2(t)$  is given as follows:

$$\hat{\sigma}^2(t) = \frac{1}{N(D) - 1} \sum_{\mathbf{x} \in N \cap D} [M(\mathbf{x}; t) - \bar{M}(t)]^2,$$

where  $\overline{M}(t) = \sum_{\mathbf{x} \in N \cap D} M(\mathbf{x}; t) / N(D)$ . In other words,  $\hat{\sigma}^2(t)$  is simply the sample variance of the observed marks. Section 3 establishes the strong consistency of  $\hat{\sigma}^2(t)$  for  $\sigma^2(t)$  under an increasing domain setting. Specifically it will be shown that  $\hat{\sigma}^2(t)$  converges to  $\sigma^2(t)$  uniformly almost surely for  $t \in (0, r)$ .

In terms of each observed mark process  $M(\mathbf{x}; t)$ , a discrepancy measure that is analogous to (2) can be defined as:

$$U(\theta; \mathbf{x}) = \int_0^r \frac{1}{\hat{\sigma}^2(t)} [M(\mathbf{x}; t) - \lambda K(t; \theta)]^2 dt.$$

This further leads to the following overall discrepancy measure:

$$U(\theta) = \frac{1}{|D|} \sum_{\mathbf{x} \in N \cap D} \int_0^r \frac{1}{\hat{\sigma}^2(t)} [M(\mathbf{x}; t) - \lambda K(t; \theta)]^2 dt.$$
(4)

Typically the estimation of  $\lambda$  is done outside of the minimization of (4). For example, a strongly consistent estimator for  $\lambda$  is given by  $\hat{\lambda} = N(D)/|D|$ . Replace  $\lambda$  in (4) by  $\hat{\lambda}$  and define  $\hat{K}(t) = \hat{J}(t)/\hat{\lambda}^2$ . Let  $U^{(i)}(\theta)$  and  $f^{(i)}(t;\theta)$  denote the *i*th gradient vectors of  $U(\theta)$  and  $f(t;\theta)$  with respect to  $\theta$ , respectively, where  $f(t;\theta)$  is an arbitrary function depending on both t and  $\theta$ . Then

$$\begin{split} U^{(1)}(\theta) &= -\frac{2\hat{\lambda}}{|D|} \sum_{\mathbf{x}\in N\cap D} \int_0^r \frac{K^{(1)}(t;\theta)}{\hat{\sigma}^2(t)} [M(\mathbf{x};t) - \hat{\lambda}K(t;\theta)] dt \\ &= -\frac{2\hat{\lambda}}{|D|} \int_0^r \frac{K^{(1)}(t;\theta)}{\hat{\sigma}^2(t)} \left[ \sum_{\mathbf{x}\in N\cap D} M(\mathbf{x};t) - \hat{\lambda}N(D)K(t;\theta) \right] dt \\ &= -2\hat{\lambda}^3 \int_0^r \frac{K^{(1)}(t;\theta)}{\hat{\sigma}^2(t)} \left[ \frac{\hat{J}(t)}{\hat{\lambda}^2} - K(t;\theta) \right] dt \\ &= -2\hat{\lambda}^3 \int_0^r \frac{K^{(1)}(t;\theta)}{\hat{\sigma}^2(t)} \left[ \hat{K}(t) - K(t;\theta) \right] dt. \end{split}$$

Ignoring the term  $\hat{\lambda}^3$ , note that  $U^{(1)}(\theta)$  is also the derivative of the following discrepancy measure with respect to  $\theta$ :

$$U(\theta) = \int_0^r \frac{1}{\hat{\sigma}^2(t)} \left[ \hat{K}(t) - K(t;\theta) \right]^2 dt.$$
(5)

Thus (5) is an equivalent but computationally simpler alternative to (4). Note that (5) (and thus (4)) is a special case of (2) with  $w(t) = 1/\hat{\sigma}^2(t)$ .

A potential problem with using (4) and (5), and (2) in general, is the weight function being used is often unbounded when  $t \to 0$ , which may lead to numerical instability for the resulting estimation procedure (e.g., Diggle 2003). A solution for this is to define a lower limit  $r_0$  and let  $w(t) = w(r_0)$  for all  $t < r_0$ . In terms of (4) and (5), this implies to set  $\sigma^2(t)$  and  $\hat{\sigma}^2(t)$  equal to  $\sigma^2(r_0)$  and  $\hat{\sigma}^2(r_0)$  for  $t < r_0$ , respectively. Guan and Sherman (2007) found that the choice of  $r_0$  affected the estimation minimally in their setting where w(t) was obtained by subsampling. The same holds true for (4) and (5). Thus the use of a lower limit  $r_0$  to bound w(t) presents little practical constraint.

We study in the next section the asymptotic properties of the estimator of  $\theta$  that is

obtained by minimizing (4). With a slight abuse of notation, let  $\sigma^2(t)$  and  $\hat{\sigma}^2(t)$  denote the truncated versions of the true and estimated variances of  $M(\mathbf{x}; t)$ , i.e.  $\sigma^2(t) = \sigma^2(r_0)$  and  $\hat{\sigma}^2(t) = \hat{\sigma}^2(r_0)$  for all  $t < r_0$ . Assume that  $\sigma^2(t)$  is bounded below from zero for  $t \ge r_0$ . The resulting weight function w(t) will thus be bounded above from infinity.

## 3. ASYMPTOTIC PROPERTIES OF $\hat{\theta}$ FOR STATIONARY POINT PROCESSES

#### 3.1 Conditions

Let N be a stationary and ergodic spatial point process whose distribution is determined by a  $p \times 1$  parameter vector,  $\theta$ . In what follows, let  $\theta_0$  denote the true parameter vector, and  $\Theta$  denote the parameter space of  $\theta$ , where  $\Theta$  is compact and  $\theta_0$  is an interior point of  $\Theta$ . Define the kth-order intensity function of N:

$$\lambda_k(\mathbf{x}_1,\cdots,\mathbf{x}_k) \equiv \lim_{|d\mathbf{x}_1|,\cdots,|d\mathbf{x}_k|\to 0} \left\{ \frac{\mathrm{E}[N(d\mathbf{x}_1)\cdots N(d\mathbf{x}_k)]}{|d\mathbf{x}_1|\cdots |d\mathbf{x}_k|} \right\}$$

Since N is stationary,  $\lambda_k(\mathbf{x}_1, \dots, \mathbf{x}_k) = \lambda_k(\mathbf{x}_2 - \mathbf{x}_1, \dots, \mathbf{x}_k - \mathbf{x}_1)$ , where  $\lambda_k(\cdot)$  now is function defined on  $\mathbb{R}^{k-1}$ . Let C denote a constant that may assume different values. Assume

$$\lambda_k(\cdot), k = 1, \cdots, 4$$
, are bounded, and  $\int |\lambda_4^*(\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_2 + \mathbf{u}_3)| d\mathbf{u}_2 < C$ , (6)

where  $\lambda_4^*(\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_2 + \mathbf{u}_3) = \lambda_4(\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_2 + \mathbf{u}_3) - \lambda_2(\mathbf{u}_1)\lambda_2(\mathbf{u}_3)$ . Condition (6) ensures the existence of the (standardized) limiting variance of  $G_n(r; \theta)$  to be defined in (9). Heinrich (1988) showed that (6) holds if N is Brillinger-mixing, examples of which include Poisson processes, Poisson cluster processes, a certain class of stationary renewal processes and others. For a stationary log Gaussian Cox process (Møller *et al.*, 1998), let  $R(\mathbf{u})$  denote the correlation function of the underlying Gaussian random field. Then (6) holds if  $\int |R(\mathbf{u})| d\mathbf{u} < \infty$  (Guan and Sherman, 2007).

Consider a sequence of domains of interest  $D_n$ . Let  $U_n(\theta)$  be  $U(\theta)$  in (4) obtained on  $D_n$  and  $\hat{\theta}_n$  be the estimator of  $\theta$  by minimizing  $U_n(\theta)$ . The large sample properties of  $\hat{\theta}_n$  will be studied under an increasing domain setting. Specifically let  $\partial D_n$  denote the boundary of  $D_n$  and  $|\partial D_n|$  denote the length of  $\partial D_n$ . Assume

$$|D_n| = \mathbf{O}(n^2), \ |\partial D_n| = \mathbf{O}(n).$$
(7)

Condition (7) ensures that  $D_n$  becomes increasingly large in all directions as n increases. Many commonly encountered domain shapes satisfy this assumption, e.g., circles with radius of order O(n), rectangles with the lengths of all sides of order O(n), etc.

The dependence in N is quantified by a strong mixing coefficient (Rosenblatt 1956). For two arbitrary points in  $\mathbf{x}, \mathbf{y} \in \mathbb{R}^2$ , let  $d_0(\mathbf{x}, \mathbf{y}) \equiv \max_{1,2}\{|x_1 - y_1|, |x_2 - y_2|\}$ , where  $\mathbf{x} = (x_1, x_2)$  and  $\mathbf{y} = (y_1, y_2)$ . Then for any subsets A, B of  $\mathbb{R}^2$  define the distance between them as  $d(A, B) = \inf\{d_0(\mathbf{x}, \mathbf{y}) : \mathbf{x} \in A, \mathbf{y} \in B\}$ . Let  $E_1$  be a compact and convex subset of  $\mathbb{R}^2$  and  $\mathcal{F}(E_1)$  be the  $\sigma$ -algebra generated by the random points falling in  $N \cap E_1$ . Define the following strong mixing coefficient:

$$\alpha(p;k) \equiv \sup\{|P(A_1 \cap A_2) - P(A_1)P(A_2)| : A_1 \in \mathcal{F}(E_1), A_2 \in \mathcal{F}(E_2), \\ E_2 = E_1 + \mathbf{x}, |E_1| = |E_2| \le p, d(E_1, E_2) \ge k\},$$

where the supremum is taken over all  $E_1 \subset \mathbb{R}^2$ , and over all  $\mathbf{x} \in \mathbb{R}^2$  such that  $d(E_1, E_2) \ge k$ . Assume the following mixing condition:

$$\sup_{p} \frac{\alpha(p;k)}{p} = \mathbf{O}(k^{-\epsilon}) \text{ for some } \epsilon > 2.$$
(8)

Condition (8) ensures that the dependence in N decreases as the distance k increases. Specifically the dependence decreases at a polynomial rate in k. As the area increases, (8) allows the dependence to increase at a rate controlled by the area p. Any point process with finite dependence range, for example, the Matérn cluster process (e.g., Stoyan and Stoyan 1994), satisfies this condition. It is also satisfied by the Strauss point process due to Jensen (1993a, b) and by the log Gaussian Cox process (e.g., Møller and Waagepetersen 2003) if the correlation of the underlying Gaussian random field decays at a polynomial rate faster than  $2 + \epsilon$  and has a spectral density bounded below due to Corollary 2 of Doukhan (1994, p.59).

## 3.2 Results

The following theorem establishes the consistency of  $\hat{\sigma}_n^2(t)$  and  $\hat{\theta}_n$ .

Theorem 1. Assume that condition (7) holds, N is ergodic,  $K(t;\theta)$  is bounded and continuous with respect to both t and  $\theta$ , and  $K(t;\theta_1) \neq K(t;\theta_2)$  on a set of positive Lebesgue measure if  $\theta_1 \neq \theta_2$ . Then

- 1.  $\sup_{t \in (0,r)} |\hat{\sigma}_n^2(t) \sigma^2(t)| \to 0$  in probability,
- 2.  $\hat{\theta}_n$  exists and is consistent for  $\theta$ .

Proof. See Appendix A.

To establish the asymptotic normality, let  $F(t; \theta) \equiv \lambda K(t; \theta)$ . Define

$$G_n(r;\theta) = \frac{1}{|D_n|} \sum_{\mathbf{x},\mathbf{y}\in N\cap D_n, \mathbf{y}\neq\mathbf{x}} \int_{||\mathbf{y}-\mathbf{x}||}^r \frac{F^{(1)}(t;\theta)}{\sigma^2(t)} dt.$$
(9)

Assume the following mild moment condition on  $G_n(r; \theta_0)$ :

$$\sup_{n} E\left[\left|\sqrt{|D_{n}|} \times \left\{G_{n}(r;\theta_{0}) - E[G_{n}(r;\theta_{0})]\right\}\right|^{2+\delta}\right] \le C \text{ for some } \delta > 0.$$
(10)

Condition (10) is only slightly stronger than the existence of the (standardized) asymptotic variances of  $G_n(r; \theta_0)$ . If N is Brillinger-mixing and  $F^{(1)}(t; \theta_0)$  is bounded, then Theorem 3 of Jolivet (1978) yields that (10) holds for all  $\delta > 0$ .

Theorem 2. Assume all conditions in Theorem 1, conditions (6), (8) and (10) hold,  $F^{(i)}(t;\theta_0), i = 1, 2$ , are bounded. Then  $\sqrt{|D_n|}(\hat{\theta}_n - \theta_0) \rightarrow N\{0, B(\theta_0)^{-1}\Sigma(\theta_0) [B(\theta_0)^{-1}]'\}$ , where the notation ' signifies the matrix transpose operation,

$$B(\theta_0) = \lambda \int_0^r \frac{F^{(1)}(t;\theta_0) \left[F^{(1)}(t;\theta_0)\right]'}{\sigma^2(t)} dt,$$

$$\Sigma(\theta_0) = \lim_{n \to \infty} |D_n| Var[G_n(r;\theta_0)].$$
(11)

Proof. See Appendix B.

Let  $\sigma^2(t_1, t_2) \equiv \lim_n |D_n| Cov[\hat{J}_n(t_1), \hat{J}_n(t_2)]$ . An alternative expression of  $\Sigma(\theta_0)$  which follows from some elementary derivations is given as:

$$\int_{0}^{r} \int_{0}^{r} \frac{\sigma^{2}(t_{1}, t_{2})}{\sigma^{2}(t_{1})\sigma^{2}(t_{2})} F^{(1)}(t_{1}; \theta_{0}) \left[ F^{(1)}(t_{2}; \theta_{0}) \right]' dt_{1} dt_{2}.$$
(12)

Let  $\hat{\theta}_n^c$  be the estimate of  $\theta$  from (1) that is obtained on  $D_n$ . Guan and Sherman (2007) showed that  $\sqrt{|D_n|}(\hat{\theta}_n^c - \theta_0) \to N\{0, B(\theta_0; c)^{-1}\Sigma(\theta_0; c) [B(\theta_0; c)^{-1}]'\}$  under suitable conditions, where

$$B(\theta_0; c) = \int_0^r \frac{J^{(1)}(t; \theta_0) \left[ J^{(1)}(t; \theta_0) \right]'}{\left[ J(t; \theta_0) \right]^{2-2c}} dt,$$
(11')

$$\Sigma(\theta_0) = \int_0^r \int_0^r \frac{\sigma^2(t_1, t_2)}{\left[J(t_1; \theta_0)\right]^{2-2c} \left[J(t_2; \theta_0)\right]^{2-2c}} J^{(1)}(t_1; \theta_0) \left[J^{(1)}(t_2; \theta_0)\right]' dt_1 dt_2.$$
(12')

Thus the proposed approach is closely related to (1) in the sense that  $\sigma^2(t)$  in (11) and (12) essentially plays the same role as  $[J(t; \theta_0)]^{2-2c}$  in (11') and (12'). Unlike (1), the proposed approach depends only on r but not on any other tuning parameter such as c in (1) and thus is more objective.

#### 4. SIMULATION

This section evaluates the finite-sample performance of the proposed methods in two important classes of spatial point process models. The first class is a Poisson cluster process whereas the second is a simple inhibition process of the first type given by Matérn (1960, Chapter 3). One thousand realizations from each type of process were simulated on a unit square with  $\lambda = 100, 400$ . For the Poisson cluster process, the expected number of parents  $\rho = 25, 100$ , and the position of each offspring relative to their parents followed a radially symmetric Gaussian random variable (see, e.g., Diggle 2003). The spread parameter  $\sigma$  was set at .01, .02 and .04 and the expected number of offspring per parent  $\mu = 4$ . For the simple inhibition process, the inhibition parameter  $\delta$  was .02 and .03 for  $\lambda = 100$  and .01

and .015 for  $\lambda = 400$ . All pairs with an inter-point distance less than  $\delta$  were removed from a simulated homogeneous Poisson process.

For each realization, the proposed method was used to estimate the pertaining parameters. The smallest lag used in all estimations was .001, whereas the largest lag was  $3\sigma$ ,  $4\sigma$  and  $5\sigma$  in the Poisson cluster case and  $2\delta$ ,  $3\delta$  and  $4\delta$  in the simple inhibition case. The MCE procedure given in (1) was also applied with c = .25, 1 in the Poisson cluster case and c = .25, .5 in the simple inhibition case. Note that c = .25 and c = .5 are the recommended values of c, whereas c = 1 and c = .25 are some "inappropriate" alternatives for cluster and inhibition processes, respectively.

Table 1 lists the mean squared errors (MSEs) in the Poisson cluster case. Note that the parameter  $\mu$  is a nuisance parameter and thus only the results on  $\hat{\rho}$  and  $\hat{\sigma}$  are reported. An interesting finding is that when (1) was used c = 1 not necessarily led to inferior results to c = .25. In fact when  $r = 3\sigma$ , c = 1 almost consistently yielded better results than c = .25. When  $r = 5\sigma$ , however, c = .25 often was better than c = 1. This observation agrees with the finding of Guan and Sherman (2007) that the choice of c is highly dependent on r. Given the fact that r is mostly chosen arbitrarily, the recommendation of using c = .25 (or smaller) for this type of process appears to be questionable. On the other hand, the MSEs from (5) are consistently smaller than the larger MSEs between c = .25 and c = 1 and are comparable to (and in fact often smaller than) the smaller MSEs between c = .25 and c = .25 and c = 1. This is true regardless the value of r. In particular, in the case of  $\lambda = 400$ , the MSEs from (5) for both  $\hat{\rho}$  and  $\hat{\sigma}$  are the smallest in in almost all cases.

Table 2 lists the MSEs in the simple inhibition case. Note that  $\delta$  can be estimated by the nearest neighbor distance between two events but not by any of the discussed MCE procedures. Thus only results on the intensity parameter of the original Poisson process are reported. Unlike the Poisson cluster case, it can be seen that the recommended value c = .5 always yielded better results than c = .25. However, this does not necessarily mean that this will hold true for all other inhibition type of processes or for other values of the upper limit r. Nevertheless, the MSEs from (5) are all very close to and often virtually the same as those from c = .5. Combining the findings from Table 1 and Table 2, it can be concluded that the proposed model fitting procedure using (5) works well for both the cluster and inhibition process models being considered. Note that the strength of the procedure is that it has eliminated the tuning parameter c used in (1) and thus is more objective than the commonly used procedure based on (1).

## 5. AN APPLICATION

The data set being studied in this section was collected by Kari Leinonen and Markku Nygren at the University of Helsinki, Finland. It consists of the locations of 114 birch trees in a  $50m \times 50m$  region. Figure 1 plots these locations which were transformed onto a unit square. Møller *et al.* (1998) fit a log Gaussian Cox process model to the data. Specifically, they used an exponential covariance function  $R(t) = \sigma^2 \exp(-\beta t)$  for the underlying Gaussian process generating the intensity of the point process and obtained the fit ( $\hat{\beta} = 16, \hat{\sigma}^2 = 1.65$ ). However, the fit to data for the estimated model was very unsatisfactory due to the fact that the point-to-nearest-event empirical density function (EDF)  $\hat{F}(t)$  (not shown) is way above the upper simulation envelope obtained from the fitted model.

The same class of log Gaussian Cox process models was fit to the data using (5) with r = .2, .3, .4. The estimates are  $(\hat{\beta} = 8.1413, \hat{\sigma}^2 = 1.7973)$  for  $r = .2, (\hat{\beta} = 8.4799, \hat{\sigma}^2 = 1.8124)$  for r = .3 and  $(\hat{\beta} = 8.7423, \hat{\sigma}^2 = 1.8215)$  for r = .4. Thus all new estimates suggest a much stronger dependence in the underlying Gaussian random field  $Z(\mathbf{x})$  than Møller *et al.*'s estimates. Figure 2 plots the nearest neighbor EDF  $\hat{G}(t)$ , the point-to-nearest-event EDF  $\hat{F}(t)$  and the simulated envelopes from the fitted model with  $(\hat{\beta} = 1.8215)$  for r = .4.

8.7423,  $\hat{\sigma}^2 = 1.8215$ ). The fit appears to be quite good since both  $\hat{G}(t)$  and  $\hat{F}(t)$  are completely within the corresponding simulation envelopes. Thus the log Gaussian Cox process model appears to be appropriate for the data. This result agrees well with those of Guan (2006) and Guan (2007) but is in contrast to Møller *et al.*'s result.

### APPENDIX A: PROOF OF THEOREM 1

*Proof.* Let  $\hat{\lambda}_n \equiv \frac{N(D_n)}{|D_n|}$ .  $\hat{\lambda}_n \to \lambda$  almost surely due to ergodicity and condition (7). Note that  $\hat{J}_n(t)$  is non-decreasing in t. Following from the proof of the Glivenko-Cantelli theorem (e.g., Durrett 1996, 59-60),  $\sup_{t \in (0,r)} |\hat{J}_n(t) - J(t;\theta_0)| \to 0$  almost surely due to ergodicity and (7).

Proof of consistency of  $\hat{\sigma}_n^2(t)$ : Note that  $\bar{M}_n(t) = \hat{J}_n(t)/\hat{\lambda}_n$ . Thus if  $t \ge r_0$ , then

$$\hat{\sigma}_n^2(t) = \frac{|D_n|}{N(D_n) - 1} \left\{ \frac{1}{|D_n|} \sum_{\mathbf{x} \in N \cap D_n} [M(\mathbf{x}; t)]^2 \right\} - \frac{N(D_n)}{N(D_n) - 1} \left[ \hat{J}_n(t) / \hat{\lambda}_n \right]^2$$

The second term on the right side of the equation converges uniformly almost surely to  $[J(t;\theta_0)/\lambda]^2$ . For the first term, note that  $M(\mathbf{x};t)$  is a non-decreasing function of t. Thus

$$\sup_{t \in (0,r)} \left| \frac{1}{|D_n|} \sum_{\mathbf{x} \in N \cap D_n} [M(\mathbf{x};t)]^2 - E\left\{ [M(\mathbf{x};t)]^2 | \mathbf{x} \in N \right\} \right| \to 0$$

almost surely due to ergodicity and (7). Thus  $\sup_{t \in (0,r)} |\hat{\sigma}_n^2(t) - \sigma^2(t)| \to 0$  almost surely.

Proof of existence of  $\hat{\theta}_n$ : Consider any sequence  $\theta_m \to \theta$ . Then  $|U_n(\theta_m) - U_n(\theta)| \to 0$ due to continuity of  $K(t; \theta)$ . Thus  $U_n(\theta)$  is continuous. The existence of  $\hat{\theta}_n$  then follows due to the fact that  $\Theta$  is compact.

*Proof of consistency of*  $\hat{\theta}_n$ : First define

$$U_n^*(\theta) = 2 \int_0^r \frac{1}{\hat{\sigma}_n^2(t)} \left[ \hat{J}_n(t) - \frac{\hat{\lambda}_n}{\lambda} J(t;\theta_0) \right] \left[ J(t;\theta_0) - J(t;\theta) \right] dt + \frac{\hat{\lambda}_n}{\lambda} \int_0^r \frac{1}{\hat{\sigma}_n^2(t)} \left[ J(t;\theta_0) - J(t;\theta) \right]^2 dt.$$

Note that  $\hat{\theta}_n = \operatorname{argmin}_{\theta} U_n^*(\theta)$  and  $U_n^*(\hat{\theta}_n) \leq U_n^*(\theta_0)$ . Note also  $U_n^*(\theta_0) = 0$ . Thus

$$\int_{0}^{r} \frac{1}{\hat{\sigma}_{n}^{2}(t)} \left[ J(t;\theta_{0}) - J(t;\hat{\theta}_{n}) \right]^{2} dt$$

$$\leq -\frac{2\lambda}{\hat{\lambda}_{n}} \int_{0}^{r} \frac{1}{\hat{\sigma}_{n}^{2}(t)} \left[ \hat{J}_{n}(t) - \frac{\hat{\lambda}_{n}}{\lambda} J(t;\theta_{0}) \right] \left[ J(t;\theta_{0}) - J(t;\hat{\theta}_{n}) \right] dt.$$
(13)

Since  $\hat{\sigma}_n^2(t)$  converge to  $\sigma^2(t)$  uniformly almost surely and  $\sigma^2(t)$  is bounded below from zero due to the truncation, we have

$$\sup_{t \in (0,r)} \left| \frac{1}{\hat{\sigma}_n^2(t)} - \frac{1}{\sigma^2(t)} \right| \to 0$$
(14)

almost surely. Note that (14) implies

$$\sup_{t \in (0,r)} \left| \frac{1}{\hat{\sigma}_n^2(t)} \right| < C \tag{15}$$

almost surely. The right side of the inequality given in (13) thus goes to zero due to (15), the assumption that  $K(t;\theta)$  is bounded,  $\hat{\lambda}_n \to \lambda$  almost surely and  $\hat{J}_n(t) \to J(t;\theta_0)$ uniformly almost surely for  $t \in (0,r)$ . Further since  $K(t;\theta_1) \neq K(t;\theta_2)$  on a set of positive Lebesgue measure,  $\int_0^r [J(t;\theta_0) - J(t;\theta)]^2 dt$  is continuous with respect to  $\theta$  and is zero only when  $\theta = \theta_0$ . Thus  $\hat{\theta}_n \to \theta$  in probability.

#### APPENDIX B: PROOF OF THEOREM 2

*Proof.* To prove Theorem 2, first note that  $U_n^{(1)}(\theta_0) + U_n^{(2)}(\theta_n^*)(\hat{\theta}_n - \theta_0) = 0$  for some  $\theta_n^* = \theta_0 + \Lambda(\hat{\theta}_n - \theta_0)$ , where  $\Lambda$  is a  $p \times p$  matrix. Further

$$U_n^{(1)}(\theta_0) = -\frac{2}{|D_n|} \sum_{\mathbf{x} \in N \cap D_n} \int_0^r \frac{1}{\hat{\sigma}_n^2(t)} \left[ \sum_{\mathbf{y} \in N \cap D_n, \mathbf{y} \neq \mathbf{x}} I(||\mathbf{y} - \mathbf{x}|| \le t) - F(t;\theta_0) \right] F^{(1)}(t;\theta_0) dt.$$

$$U_{n}^{(2)}(\theta_{n}^{*}) = \frac{2}{|D_{n}|} \sum_{\mathbf{x} \in N \cap D_{n}} \int_{0}^{r} \frac{1}{\hat{\sigma}_{n}^{2}(t)} F^{(1)}(t;\theta_{n}^{*}) \left[F^{(1)}(t;\theta_{n}^{*})\right]' dt$$
$$- \frac{2}{|D_{n}|} \sum_{\mathbf{x} \in N \cap D_{n}} \int_{0}^{r} \frac{1}{\hat{\sigma}_{n}^{2}(t)} \left[\sum_{\mathbf{y} \in N \cap D_{n}, \mathbf{y} \neq \mathbf{x}} I(||\mathbf{y} - \mathbf{x}|| \leq t) - F(t;\theta_{n}^{*})\right] F^{(2)}(t;\theta_{n}^{*}) dt$$

Let  $X_n \sim Y_n$  denote that  $X_n$  and  $Y_n$  have the same limiting distribution. Define

$$A_{n}(\theta_{0}) = \frac{1}{|D_{n}|} \sum_{\mathbf{x} \in N \cap D_{n}} \int_{0}^{r} \frac{1}{\sigma^{2}(t)} \left[ \sum_{\mathbf{y} \in N \cap D_{n}, \mathbf{y} \neq \mathbf{x}} I(||\mathbf{y} - \mathbf{x}|| \le t) - F(t;\theta_{0}) \right] F^{(1)}(t;\theta_{0}) dt.$$

The proof of asymptotic normality consists of the following four steps:

Step 1: 
$$U_n^{(2)}(\theta_n^*) \to 2B(\theta_0)$$
 in probability,  
Step 2:  $\sqrt{|D_n|}(\hat{\theta}_n - \theta_0) \sim -B(\theta_0)^{-1}\sqrt{|D_n|}U_n^{(1)}(\theta_0)/2$ ,  
Step 3:  $\sqrt{|D_n|}U_n^{(1)}(\theta_0)/2 \sim \sqrt{|D_n|}A_n(\theta_0)$ ,  
Step 4:  $\sqrt{|D_n|}A_n(\theta_0) \to N\left\{0, B(\theta_0)^{-1}\Sigma(\theta_0) \left[B(\theta_0)^{-1}\right]'\right\}$  in distribution.

*Proof of Step 1:* The first term of  $U_n^{(2)}(\theta_n^*)$  converges to  $2B(\theta_0)$  due to ergodicity,  $\theta_n^* \to \theta_0, F^{(1)}(t;\theta)$  is bounded and continuous, and (14). Note the second term of  $U_n^{(2)}(\theta_n^*)$  can be rewritten as

$$\int_0^r \frac{1}{\hat{\sigma}_n^2(t)} \left[ \hat{J}_n(t) - \hat{\lambda}_n F(t; \theta_n^*) \right] F^{(2)}(t; \theta_n^*) dt,$$

which converges to zero due to the facts that  $\theta_n^* \to \theta_0$ ,  $\hat{\lambda}_n \to \lambda$  almost surely, and  $\hat{J}_n(t) \to J(t;\theta_0)$  uniformly almost surely, where  $J(t;\theta_0) = \lambda F(t;\theta_0)$ , and (14). Thus Step 1 is proved.

Proof of Step 2: This follows directly from Step 1 and Slutsky's theorem.

Proof of Step 3: This is due to (14).

Proof of Step 4: Note that

$$G_n(r;\theta_0) = \frac{1}{|D_n|} \sum_{\mathbf{x} \in N \cap D_n} \int_0^r \frac{1}{\sigma^2(t)} \left[ \sum_{\mathbf{y} \in N \cap D_n, \mathbf{y} \neq \mathbf{x}} I(||\mathbf{y} - \mathbf{x}|| \le t) \right] F^{(1)}(t;\theta_0) dt,$$

where  $G_n(r; \theta_0)$  is defined as in (9). Since  $A_n(\theta_0) = G_n(r; \theta_0) - E[G_n(r; \theta_0)]$ , it is only needed to show that

$$\sqrt{|D_n|} \left\{ G_n(r;\theta_0) - E\left[G_n(r;\theta_0)\right] \right\} \to N\left\{ 0, B(\theta_0)^{-1} \Sigma(\theta_0) \left[ B(\theta_0)^{-1} \right]' \right\}$$

in distribution. The existence of  $\Sigma(\theta_0)$  is guaranteed by condition (6). The rest follows similarly as in the proof of Guan *et al.* (2004) due to conditions (7), (8) and (10). Thus Theorem 2 is proved.

#### REFERENCES

- Cressie, N. A. C. (1993), Statistics for Spatial Data, New York: Wiley.
- Diggle, P. J. (2003), *Statistical Analysis of Spatial Point Patterns*, New York: Oxford University Press Inc.
- Doukhan, P. (1994), Mixing: Properties and Examples, New York: Springer-Verlag.
- Durrett, R. (1996), Probability: theory and examples, New York: Duxbury Press.
- Guan, Y. (2006), "A Composite Likelihood Approach in Fitting Spatial Point Process Models", *Journal of the American Statistical Association*, 101, 1502–1512.
- Guan, Y. (2007), "A Composite Likelihood Cross-Validation Approach in Selecting Bandwidth for the Estimation of the Pair Correlation Function", *Scandinavian Journal of Statistics*, 34, 336–346.
- Guan, Y. and Sherman, M. (2007), "On Least Squares Fitting for Stationary Spatial Point Processes", *Journal of the Royal Statistical Society, Series B*, 69, 31–49.
- Guan, Y., Sherman, M. and Calvin, J. A. (2004), "A Nonparametric Test for Spatial Isotropy Using Subsampling", *Journal of the American Statistical Assocation, Theory* and Methods, 99, 810–821.
- Heinrich, L. (1988), "Asymptotic Gaussianity of Some Estimators for Reduced Factorial Moment Measures and Product Densities of Stationary Poisson Cluster Processes", *Statistics*, 19, 87–106.

- Jensen, J. L. (1993a), "A Note on Asymptotic Expansions for Sums over a Weakly Dependent Random Field with Application to the Poisson and Strauss Processes", *Annals of the Institute of Statistical Mathematics*, 45, 353-360.
- Jensen, J. L. (1993b), "Asymptotic Normality of Estimates in Spatial Point Processes", *Scandinavian Journal of Statistics*, 20, 97-109.
- Jolivet, E. (1978), "Central Limit Theorem and Convergence of Empirical Processes for Stationary Point Processes", Colloquia Mathematica Socitatis Janos Bolyai 24. Point Processes and Queuing Problems, Keszthely, Hungary, 117–161.
- Loh, J. M. and Stein, M. L. (2004), "Bootstrapping a Spatial Point Process", *Statistica Sinica*, 14, 69–101.
- Matérn (1960), Spatial Variation, Stockholm: Statens Skogsforsningsinstitut.
- Møller, J., Syversveen, A. R. and Waagepetersen, R. P. (1998), "Log Gaussian Cox Processes", *Scandinavian Journal of Statistics*, 25, 451–482.
- Møller, J. and Waagepetersen, R. P. (2003), *Statistical Inference and Simulation for Spatial Point Processes*, New York: Chapman & Hall.
- Rosenblatt, M. (1956), "A Central Limit Theorem and a Strong Mixing Condition", *Proceedings of the National Academy of Sciences*, 42, 43–47.
- Stoyan, D. and Stoyan, H. (1994), Fractals, Random Shapes and Point Fields, New York: Wiley.



Figure 1. Locations of birches.



Figure 2. Goodness-of-fit plots using the nearest neighbor (left) and the pointto-nearest-event (right) distribution functions. The solid curves are the empirical distribution functions, whereas the dotted curves are the envelopes from 19 simulations of the fitted model.

Table 1. Estimated mean squared errors (MSEs) of the different estimators from simulations in the stationary Poisson cluster case. Each MSE value is divided by the squared value of the target parameter that it corresponds to.

				$r = 3\sigma$			$r = 4\sigma$			$r = 5\sigma$			
	$\lambda$	$\sigma$	c = .25	c = 1	(5)	c = .25	c = 1	(5)	c = .25	c = 1	(5)		
$\hat{ ho}$	100	.01	.0796	.0576	.0585	.0549	.0526	.0526	.0548	.0550	.0549		
	100	.02	.1084	.0856	.0876	.0837	.0837	.0836	.0889	.0928	.0919		
	100	.04	.2828	.2764	.2725	.3084	.3457	.3339	.3783	.4444	.4314		
	400	.01	.0235	.0215	.0211	.0216	.0216	.0213	.0234	.0240	.0234		
	400	.02	.0501	.0489	.0455	.0524	.0553	.0526	.0603	.0656	.0618		
	400	.04	.2665	.2728	.1851	.3166	.3499	.2917	.3730	.4356	.3734		
$\hat{\sigma}$	100	.01	.2066	.0203	.0248	.0197	.0105	.0113	.0133	.0097	.0100		
	100	.02	.4054	.0328	.0659	.0541	.0173	.0197	.0247	.0190	.0192		
	100	.04	.7788	.0967	.1221	.0686	.0424	.0448	.0554	.0534	.0501		
	400	.01	.0085	.0054	.0057	.0052	.0043	.0044	.0054	.0053	.0050		
	400	.02	.0169	.0102	.0103	.0112	.0101	.0096	.0125	.0139	.0121		
	400	.04	.0681	.0383	.0283	.0381	.0376	.0315	.0439	.0538	.0430		

Table 2. Estimated mean squared errors (MSEs) of the different estimators from simulations in the simple inhibition case. Each MSE value is divided by the squared value of the target parameter that it corresponds to.

			$r = 3\delta$				$r = 4\delta$				
$\lambda$	$\delta$	c = .25	c = .5	(5)	c = .25	c = .5	(5)		c = .25	c = .5	(5)
100	.02	.0468	.0345	.0354	.0225	.0190	.0195		.0170	.0153	.0155
	.03	.0512	.0463	.0417	.0324	.0306	.0306		.0291	.0285	.0284
400	.01	.0095	.0082	.0084	.0055	.0050	.0051		.0044	.0041	.0041
	.015	.0110	.0100	.0100	.0080	.0077	.0077		.0073	.0071	.0071