Two-step estimation for inhomogeneous spatial point processes

Rasmus Waagepetersen†

Aalborg University, Aalborg, Denmark

Yongtao Guan

Yale University, New Haven, USA

Summary.This paper is concerned with parameter estimation for inhomogeneous spatial point processes with a regression model for the intensity function and tractable second order properties (K-function). Regression parameters are estimated using a Poisson likelihood score estimating function and in a second step minimum contrast estimation is applied for the residual clustering parameters. Asymptotic normality of parameter estimates is established under certain mixing conditions and we exemplify how the results may be applied in ecological studies of rain forests.

1. Introduction

The results in this paper can be applied in many contexts in e.g. biology and spatial epidemiology. The original motivation is, however, studies of biodiversity in tropical rain forest ecology. A question of particular interest is how the very high number of different rain forest tree species continue to coexist, see e.g. Burslem *et al.* (2001) and Hubbell (2001). Conspecific aggregation is hypothesized to promote diversity although the causes of aggregation remain unclear (Seidler and Plotkin, 2006). One explanation is the so called niche assembly hypothesis that different species benefit from different habitats determined e.g. by topography or soil properties. However, the aggregation of conspecific trees may also be due to seed dispersal. Hence it is of interest to quantify residual clustering after adjusting for possible aggregation due to environmental covariates. In recent years huge amounts of data have been collected in tropical rain forest plots in order to investigate the niche assembly and other competing hypotheses (Losos and Leigh, 2004). The data sets consist of measurements of soil properties, digital terrain models, and individual locations of all trees growing in the plots.

In this paper we model the set of tree locations for a particular species as a realization of a spatial point process X on $\mathbb{R}^2$ with intensity function of the form

$$\rho_\beta(u) = \rho(z(u)\beta^\mathsf{T}), u \in \mathbb{R}^2,$$

where ρ is a positive strictly increasing differentiable function, $z(u)$ is the covariate vector associated with the spatial location u, and β is a regression parameter. Evidence of the niche assembly hypothesis may be obtained by assessing the magnitudes of the components of β.

†*Address for correspondence:* Rasmus Waagepetersen, Department of Mathematical Sciences, Aalborg University, Fredrik Bajersvej 7G, DK-9220 Aalborg
E-mail: rw@math.aau.dk

The second-order properties of X is determined by the so-called pair correlation function g (see Section 3). Translation-invariance of g implies second-order intensity reweighted stationarity (Baddeley *et al.*, 2000) in which case the so-called K-function is well-defined and given in terms of an integral involving g (cf. (2) in Section 3). A parametric model g_ψ is often imposed for the pair correlation function and hypotheses regarding clustering may be formulated in terms of ψ. Seidler and Plotkin (2006) e.g. relate estimates of ψ for different species to classes of species determined by their modes of seed dispersal.

Suppose X is observed within a bounded observation window $W \subset \mathbb{R}^2$. If X is a Poisson process, the maximum likelihood estimate of β is obtained by solving $u(\beta) = 0$ where u is the Poisson score estimating function

$$u(\beta) = \sum_{u \in X \cap W} \frac{\rho_\beta^{(1)}(u)}{\rho_\beta(u)} - \int_W \rho_\beta^{(1)}(u)\mathrm{d}u \qquad (1)$$

where here and in the following, $\rho_\beta^{(l)}$ denotes the l'th derivative with respect to β. If X is not Poisson, likelihood-based inference may be carried out using Markov chain Monte Carlo methods (Møller and Waagepetersen, 2003) but especially in case of Cox and cluster processes, this can be computationally very hard. However, also for non-Poisson processes, an estimate $\hat\beta$ may be obtained using the estimating function (1). Consistency of $\hat\beta$ was studied in Schoenberg (2005) while Waagepetersen (2007) obtained asymptotic normality of $\hat\beta$ for a fixed observation window employing infinite divisibility of the inhomogeneous Neyman-Scott processes considered in this paper. Guan and Loh (2007) instead established asymptotic normality using increasing-domain asymptotics for suitably mixing point processes. Regarding ψ, Waagepetersen (2007) suggested a two-step estimation procedure where ψ is estimated using minimum contrast estimation based on the theoretical K-function K_ψ and an estimate of K_ψ depending on β. Waagepetersen (2007), however, did not provide a theoretical study of this approach for estimating ψ.

In some cases, β is the parameter of main interest. Nevertheless, the asymptotic covariance matrix for β still depends on the second-order properties of X. Waagepetersen (2007) suggested a plug-in approach where $g_{\hat\psi}$ is plugged in for the unknown pair correlation function while Guan and Loh (2007) suggested a block bootstrap procedure which avoids the specification of a specific parametric model for the pair correlation.

It is not obvious that the second step of the two-step procedure produces useful estimates of ψ since the estimate of the K-function is not unbiased when $\hat\beta$ is plugged in for the true value of β. Under certain mixing conditions however, we show that the parameter estimate $(\hat\beta, \hat\psi)$ in fact does enjoy the usual desirable properties of consistency and asymptotic normality. Our results thus extend the results for $\hat\beta$ in Waagepetersen (2007) and Guan and Loh (2007) and the results for $\hat\psi$ in Heinrich (1992) and Guan and Sherman (2007) who considered stationary point processes. The asymptotic normality of $\hat\psi$ both gives a theoretical basis for inference regarding ψ and also for the plug-in approach in Waagepetersen (2007).

In section 2 we describe a data example previously considered in Waagepetersen (2007) and Guan and Loh (2007). Some background concerning Cox and cluster point process and their product densities are given and some basic assumptions stated in Section 3. The two-step procedure for parameter estimation and its asymptotic properties are considered in Section 4 and applied to the data example in Section 5. Section 5 also contains a simulation study. A few open problems are discussed in Section 6.
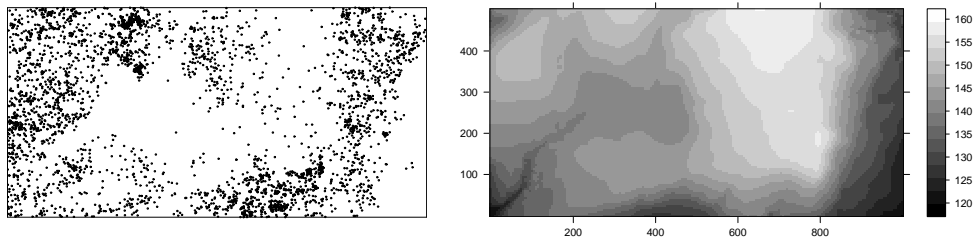
Fig. 1. Locations of *Beilschmiedia pendula Lauraceae* (left) and altitude (right).

2. A data example

The tropical tree data set considered in this section is extracted from a huge data set collected in the 1000 by 500 meter Barro Colorado Island plot, see Condit *et al.* (1996), Condit (1998), and Hubbell and Foster (1983). The left plot in Figure 1 shows all tree positions in 1995 of the species *Beilschmiedia pendula Lauraceae* (3604 trees). The right plot shows altitude on a 5 by 5 meter grid.

Letting $z(u) = (1, z_2(u), z_3(u))$ where z_2 denotes altitude and z_3 slope (the norm of the altitude gradient), Waagepetersen (2007) and Guan and Loh (2007) fitted a log-linear model

$$\rho_\beta(u) = \exp(z(u)\beta^\mathsf{T})$$

and concluded that the intensity of the rain forest trees depends significantly on the slope. Exploratory plots based on the K-function indicated that the assumption of an inhomogeneous Poisson process is not tenable due to clustering which may be due to e.g. unobserved environmental covariates, seed dispersal, and competition due to other species. Waagepetersen (2007) assumed that the data originated from a modified Thomas process (see Section 3) while Guan and Loh (2007) used a block bootstrap to evaluate the variance of $\hat\beta$ without explicit assumptions regarding the nature of the clustering.

We return to this data set in Section 5 but expand the analysis by adding covariates on pH and mineralized nitrogen, phosphorous and potassium contents in the soil and considering inference for the clustering parameters of the Thomas process as well as covariance parameters in a log Gaussian Cox process model (Section 3).

3. Some basic background and assumptions

This section describes the specific examples of point processes considered in this paper and gives some background on product densities for point processes.

3.1. Inhomogeneous Cox point processes

A Cox process X is defined in terms of a random intensity function Λ where given $\Lambda = \lambda$, X is a Poisson process with intensity function λ. For a log Gaussian Cox process (LGCP), $\log \Lambda$ is a Gaussian process. A Neyman-Scott process with Poisson numbers of offspring is a union $\cup_{c \in C} X_c$ where C is a 'mother' Poisson process of intensity $\kappa > 0$. Given C, X_c, $c \in C$, are independent Poisson processes with intensity functions $\alpha k(\cdot - c)$ where $\alpha > 0$ is the expected number of offspring for each mother point and $k(\cdot)$ is a probability density determining the spread of offspring around their mother. A so-called modified Thomas

process is obtained when k is the density of a bivariate normal distribution $N(0, \omega^2 I)$. A Neyman-Scott process can also be viewed as a Cox process with

$$\Lambda(u) = \alpha \sum_{c \in C} k(u - c)$$

and inhomogeneous Neyman-Scott processes are obtained by multiplying $\Lambda(u)$ by e.g. a log-linear term $\exp(z(u)\beta^\mathsf{T})$ (Waagepetersen, 2007).

3.2. *Product densities and basic assumptions*

Let $\rho_{\beta,n}(u_1, \ldots, u_n)$ denote the nth order product density of X. For locations u_i in infinitesimally small regions A_i of volumes $\mathrm{d}A_i$, $i = 1, \ldots, n$, $\rho_{\beta,n}(u_1, \ldots, u_n)\mathrm{d}A_1 \cdots \mathrm{d}A_n$ is the joint probability that X has a point in each A_i. The pair correlation function is

$$g(u, v) = \frac{\rho_{\beta,2}(u, v)}{\rho_\beta(u)\rho_\beta(v)}.$$

Throughout the paper we assume without further notice the following:

B1 bounded covariates

$$\|z(u)\| \leq K_1, u \in \mathbb{R}^2, \quad \text{for some } K_1 < \infty.$$

B2 the product densities $\rho_{\beta,n}$ are of the form

$$\rho_{\beta,n}(u_1, \ldots, u_n) = \rho_n(u_1, \ldots, u_n) \prod_{i=1}^{n} \rho_\beta(u_i)$$

where ρ_n is the nth order product density of a stationary point process on $\mathbb{R}^2$.

B3 ρ_2 and ρ_3 are bounded and there is a K_2 so that for all $u_1, u_2 \in \mathbb{R}^2$, $\int |\rho_3(0, v, v + u_1) - \rho_1(0)\rho_2(0, u_1)|\mathrm{d}v < K_2$ and $\int |\rho_4(0, u_1, v, v + u_2) - \rho_2(0, u_1)\rho_2(0, u_2)|\mathrm{d}v < K_2$.

B4 $W_n = [an, nb] \times [cn, dn]$ where $b - a > 0$ and $d - c > 0$.

The assumption B1 of bounded covariates is not a serious restriction from a practical point of view and implies $k_1 \leq \rho_\beta(u), |\rho_\beta^{(l)}(u)| \leq K_3$, $l = 1, 2, \ldots$ for constants $k_1 > 0$ and $K_3 < \infty$. The assumption B2 e.g holds if X is an independent thinning of a stationary point process with probability of retaining a point at u proportional to $\rho_\beta(u)$. Under B1, both log Gaussian Cox processes and inhomogeneous Neyman-Scott processes fall into this category. Note that stationarity implies $\rho_n(u_1, \ldots, u_n) = \rho_n(0, u_2 - u_1, \ldots, u_n - u_1)$. In the current setting, the pair correlation function g of X coincides with ρ_2 and with a convenient abuse of notation we write $g(u, v) = g(v - u)$ where $g(h) = \rho_2(0, h)$. Moreover, the K-function is given by

$$K(t) = \int_{\|h\| \leq t} g(h)\mathrm{d}h, \quad t \geq 0. \tag{2}$$

If g is isotropic, $g(h)$ can be recovered from $K(t)$ by differentiating: $g(h) = K'(\|h\|)/(2\pi\|h\|)$. For a Thomas process the pair correlation function is

$$g_{(\kappa,\omega)}(h) = 1 + \exp(-\|h\|^2/(4\omega^2))/(4\pi\omega^2\kappa), \quad \kappa, \omega > 0,$$

while it is

$$g_\psi(h) = \exp(c_\psi(h))$$

for a log Gaussian Cox process with covariance function $c_\psi(h)$ for the Gaussian field. One example of covariance function is the exponential

$$c_{(\sigma^2,\phi)}(h) = \sigma^2 \exp(-\|h\|/\phi), \quad \sigma^2, \phi > 0,$$

where σ^2 is the variance and ϕ is the correlation scale parameter.

The assumption B3 of weak dependence holds for many commonly applied point processes including Poisson cluster processes and log Gaussian Cox processes with an absolutely integrable correlation function, see Guan and Sherman (2007). Rectangular observation windows B4 are assumed for ease of exposition and can be relaxed. It is important, though, that for any $h \in \mathbb{R}^2$, $\lim_{n\to\infty} |W_n|/|W_n \cap W_{n,h}| = 1$ where $W_{n,h}$ is W_n translated by h.

4.   The two-step estimation procedure

Suppose X is observed within W_n, $\beta \in \mathbb{R}^p$, and $\psi \in \Psi \subseteq \mathbb{R}^q$. We then first obtain $\hat\beta_n$ by solving $u_{n,1}(\beta) = 0$ where

$$u_{n,1}(\beta) = \sum_{u \in X \cap W_n} \frac{\rho_\beta^{(1)}(u)}{\rho_\beta(u)} - \int_{W_n} \rho_\beta^{(1)}(u)\mathrm{d}u.$$

Second, $\hat\psi_n$ is obtained by minimizing $m_n(\psi) = m_{n,\hat\beta_n}(\psi)$ where

$$m_{n,\beta}(\psi) = \int_{r_l}^r (\hat K_{n,\beta}(t)^c - K_\psi(t)^c)^2 \mathrm{d}t, \tag{3}$$

r_l, r, and c are user-specified constants, and

$$\hat K_{n,\beta}(t) = \sum_{u,v \in X \cap W_n}^{\neq} \frac{1[\|u - v\| \le t]}{\rho_\beta(u)\rho_\beta(v)|W_n \cap W_{n,u-v}|} \tag{4}$$

is an estimate (Baddeley *et al.*, 2000) of the theoretical K-function $K_\psi(t)$. We denote by β^* and ψ^* the true values of β and ψ and note that $\hat K_{n,\beta^*}(t)$ is unbiased for $K_{\psi^*}(t)$. An excellent account of practical aspects of minimum contrast estimation is given in Section 6.1 in Diggle (2003) where $r_l = 0$. However, in Section 4.1 we for technical reasons need $r_l > 0$ when $c < 1$.

4.1.   *Joint asymptotic normality of* $(\hat\beta_n, \hat\psi_n)$

Let

$$u_{n,2}(\beta,\psi) = -|W_n|\frac{\mathrm{d}m_{n,\beta}(\psi)}{\mathrm{d}\psi} = |W_n|2c\int_{r_l}^r (\hat K_{n,\beta}(t)^c - K_\psi(t)^c)K_\psi(t)^{c-1}K_\psi^{(1)}(t)\mathrm{d}t$$

(assuming that K_ψ is differentiable, cf. N2 below). Then the two-step estimating procedure corresponds to solving

$$u_n(\beta,\psi) = (u_{n,1}(\beta), u_{n,2}(\beta,\psi)) = 0.$$

By a Taylor-expansion, $u_{n,2}(\beta^*, \psi^*)$ can be approximated by

$$\tilde{u}_{n,2}(\beta^*, \psi^*) = |W_n|2c^2 \int_{r_l}^r (\hat{K}_{n,\beta^*}(t) - K_{\psi^*}(t)) K_{\psi^*}(t)^{2c-2} K_{\psi^*}^{(1)}(t)\mathrm{d}t$$

and we define

$$\tilde{\Sigma}_n = |W_n|^{-1}\mathbb{V}\mathrm{ar}\big(u_{n,1}(\beta^*), \tilde{u}_{n,2}(\beta^*, \psi^*)\big).$$

From a mathematical point of view, $\tilde{u}_{n,2}(\beta^*, \psi^*)$ is easier to handle than $u_{n,2}(\beta^*, \psi^*)$ since we avoid the exponent c for $\hat{K}_{n,\beta^*}(t)$.

Define further

$$I_{n,11} = \frac{1}{|W_n|} \int_{W_n} \frac{(\rho_{\beta^*}^{(1)}(u))^{\mathsf{T}} \rho_{\beta^*}^{(1)}(u)}{\rho_{\beta^*}(u)}\mathrm{d}u,$$

$$I_{n,12} = -2c^2 \int_{r_l}^r H_{n,\beta^*}(t) K_{\psi^*}^{2c-2}(t) K_{\psi^*}^{(1)}(t)\mathrm{d}t$$

where

$$H_{n,\beta^*}(t) = \mathbb{E}\frac{\mathrm{d}}{\mathrm{d}\beta^{\mathsf{T}}}\hat{K}_{n,\beta}(t)|_{\beta=\beta^*} = -2 \int_{W_n^2} \frac{1[\|u-v\| \le t]}{|W_n \cap W_{n,u-v}|} \frac{\rho_{\beta^*}^{(1)}(u)}{\rho_{\beta^*}(u)} g_{\psi^*}(u-v)\mathrm{d}u\mathrm{d}v$$

and

$$I_{22} = 2c^2 \int_{r_l}^r K_{\psi}(t)^{2c-2} (K_{\psi}^{(1)}(t))^{\mathsf{T}} K_{\psi}^{(1)}(t)\mathrm{d}t.$$

The following result is verified in Appendix A.

THEOREM 1. *Assume*

N1 $r_l > 0$ *if* $c < 1$; *otherwise* $r_l \ge 0$.
N2 ρ_β *and* K_ψ *are twice continuously differentiable as functions of* β *and* ψ.
N3 I_{22} *is positive definite and* $\liminf_{n\to\infty} \min\{\tilde{\lambda}_n, \lambda_{n,11}\} > 0$ *where* $\tilde{\lambda}_n$ *and* $\lambda_{n,11}$ *are the smallest eigenvalues of* $\tilde{\Sigma}_n$ *and* $I_{n,11}$, *respectively.*
N4 $\rho_{4+2\delta}(u_1, \cdots, u_{4+2\delta}) < \infty$ *for some positive integer* δ.
N5 *for some* $a > 8r^2$,

$$\alpha_{a,\infty}(m) = O(m^{-d}) \text{ for some } d > 2(2+\delta)/\delta \tag{5}$$

where, following Politis et al. *(1998), the mixing coefficient* $\alpha_{a_1,a_2}(m)$ *is*

$$\alpha_{a_1,a_2}(m) \equiv sup\{|P(A_1 \cap A_2) - P(A_1)P(A_2)| : A_1 \in \mathcal{F}(E_1),$$
$$A_2 \in \mathcal{F}(E_2), |E_1| \le a_1, |E_2| \le a_2, d(E_1, E_2) \ge m, \ E_1, E_2 \in \mathcal{B}(\mathbb{R}^2)\}$$

and $\mathcal{B}(\mathbb{R}^2)$ *denotes the set of Borel sets in* $\mathbb{R}^2$, $d(E_1, E_2)$ *is the minimal distance between* E_1 *and* E_2, *and* $\mathcal{F}(E_i)$ *is the* σ-*algebra generated by* $X \cap E_i$, $i = 1, 2$.

Then there is a sequence $\{(\hat{\beta}_n, \hat{\psi}_n)\}_{n\ge 1}$ *for which* $u_n(\hat{\beta}_n, \hat{\psi}_n) = 0$ *with a probability tending to one and where*

$$|W_n|^{1/2}[(\hat{\beta}_n, \hat{\psi}_n) - (\beta^*, \psi^*)]I_n\tilde{\Sigma}_n^{-1/2} \overset{d}{\to} N(0, I) \tag{6}$$

with

$$I_n = \begin{bmatrix} I_{n,11} & I_{n,12} \\ 0 & I_{22} \end{bmatrix}. \tag{7}$$

In the following two sections 4.2 and 4.3 we discuss in more detail the practical use of this result and the conditions for it.

4.2. Practical issues

Often $u_n(\beta, \psi) = 0$ has a unique solution which then coincides with $(\hat{\beta}_n, \hat{\psi}_n)$. The practical implication of (6) is that for a given n, $(\hat{\beta}_n, \hat{\psi}_n)$ is approximately normal with mean (β^*, ψ^*) and covariance matrix $(I_n^{\mathsf{T}})^{-1}\tilde{\Sigma}_n I_n^{-1}$ (by N3, I_n^{-1} exists for large enough n). The upper block $\tilde{\Sigma}_{n,11}$ in $\tilde{\Sigma}_n$ is the sum of $I_{n,11}$ and

$$\frac{1}{|W_n|} \int_{W_n^2} (\rho_{\beta^*}^{(1)}(u))^{\mathsf{T}} \rho_{\beta^*}^{(1)}(v)(g_{\psi^*}(u-v)-1)\mathrm{d}u\mathrm{d}v.$$

The more complicated expressions defining $\tilde{\Sigma}_{n,12} = \tilde{\Sigma}_{n,21}^{\mathsf{T}}$ and $\tilde{\Sigma}_{n,22}$ are discussed in Appendix B. Due to the basic assumptions, the entries in $\tilde{\Sigma}_n$ are bounded below and above. We obtain consistent estimates $\hat{I}_n$ and $\hat{\Sigma}_n$ by replacing β^* and ψ^* in I_n and $\tilde{\Sigma}_n$ with $\hat{\beta}_n$ and $\hat{\psi}_n$. The integrals in $\hat{I}_n$ are evaluated using numerical quadrature. For ease of implementation we evaluate the integrals in $\hat{\Sigma}_n$ using Monte Carlo simulations under the fitted model given by $\hat{\beta}_n$ and $\hat{\psi}_n$. Alternatively one might use numerical quadrature. Regarding $I_{n,12}$ the following approximation

$$H_{n,\beta^*}(t) \approx -\frac{2K_{\psi^*}(t)}{|W_n|} \int_{W_n} \frac{\rho_{\beta^*}^{(1)}(u)}{\rho_{\beta^*}(u)}\mathrm{d}u$$

is useful. The matrix $\tilde{\Sigma}_n$ is mainly used for mathematical convenience and in Section 5 we also consider the alternative

$$\Sigma_n = |W_n|^{-1}\mathbb{V}\mathrm{ar}\, u_n(\beta^*, \psi^*).$$

The approximate covariance for $\hat{\psi}_n$ is of the form

$$I_{22}^{-1}\big[(I^{n,12})^{\mathsf{T}}\tilde{\Sigma}_{n,11}I^{n,12} - \tilde{\Sigma}_{n,12}^{\mathsf{T}}I^{n,12} - (I^{n,12})^{\mathsf{T}}\tilde{\Sigma}_{n,12} + \tilde{\Sigma}_{n,22}\big]I_{22}^{-1}$$

where $I^{n,12} = I_{n,11}^{-1}I_{n,12}$. The sum of matrices within the parenthesis corresponds to the variance of $\tilde{u}_{n,2}(\hat{\beta}_n, \psi^*)$ which (at least in our examples) is less than the variance $\tilde{\Sigma}_{n,22}$ of $\tilde{u}_{n,2}(\beta^*, \psi^*)$ due to the effect of of plugging in $\hat{\beta}_n$ rather than β^* in (4). This is related to the observation (Dietrich Stoyan, personal communication) that a more precise estimate of the K-function is obtained in the stationary case when using an estimated intensity rather than the true value of the intensity. As a curious consequence, the variance for $\hat{\psi}_n$ becomes smaller when $\hat{\beta}_n$ is used in the minimum contrast estimation rather than the true value β^*.

4.3. Discussion of conditions for asymptotic normality

The condition N1 is needed for technical reasons so that we can apply Lemma 2 (Appendix D) with $d < 0$ in the proofs of Lemma 4 and Lemma 5 in Appendix D. In practice we approximate the integrals from r_l to r using right endpoint Riemann sums in which case specifying $r_l = 0$ is unproblematic. Condition N2 is satisfied for many examples of Neyman-Scott and log Gaussian Cox processes but exclude the well-known Matérn cluster process. Regarding N3, the matrix I_{22} is positive definite if $K_\psi^{(1)}(t_i)$, $i = 1, \ldots, q$, are linearly independent for distinct $r_l < t_1 < t_2 < \cdots < t_q < r$. It is hard to say something general about the condition on the eigenvalues of $\tilde{\Sigma}_n$ and $I_{n,11}$ since they depend on the behaviour of the covariates on all of $\mathbb{R}^2$. However, the condition seems reasonable since $\tilde{\Sigma}_n$

and $I_{n,11}$ are both covariance matrices ($I_{n,11}$ is the covariance matrix of $u_{n,1}(\beta^*)$ if X is a Poisson process). The condition N4 of bounded product densities is not restrictive.

Condition N5 requires that the dependence between parts of the point process observed in two distinct sets decays at a polynomial rate as a function of the inter-set distance m. In the nonstationary case, it follows from (1') at page 3 in Doukhan (1994) that N5 is satisfied if the process can be regarded as an independent thinning of a stationary process satisfying N5. By the same result, N5 holds for Cox processes if the condition is satisfied for the random intensity function Λ. For many examples of stationary Neyman-Scott processes including the modified Thomas process, N5 can be verified directly, see Appendix E. Regarding stationary LGCPs, simple conditions for mixing of stationary Gaussian fields are provided in Corollary 2 in Doukhan (1994) but are restricted to fields on $\mathbb{Z}^d$, $d \geq 1$. From a practical point of view, however, we can approximate a continuous Gaussian field $(Y_s)_{s \in \mathbb{R}^2}$ arbitrarily well by step functions with step heights Y_s for s on a fine grid $\{\epsilon(i,j) : (i,j) \in \mathbb{Z}^2\}$, $\epsilon > 0$, see also Waagepetersen (2004).

5.   Data examples and simulation studies

Returning to the data example from Section 2, Waagepetersen (2007) considered an inhomogeneous Thomas process with log linear intensity function and covariates altitude and slope. Using $c = 0.25$ and $r = 100$ he obtained minimum contrast estimates $8{\times}10^{-5}$ and 20 for κ and ω. Since then covariates regarding soil properties have been made available and in addition to altitude and slope we now include pH, mineralized nitrogen, phosphorous, and potassium covariates. The solid line in Figure 2 shows the difference between the resulting estimate (4) of the K-function and the K-function for a Poisson process. Even after adjusting for the soil variables considerable extra-Poisson clustering remains.
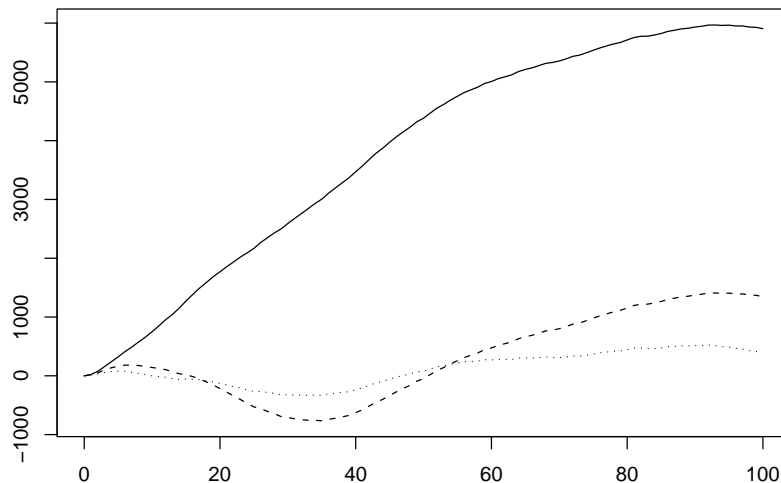

Fig. 2. Solid, dashed, dotted lines: estimate of K-function minus respectively K-function for Poisson process, K-function for fitted inhomogeneous Thomas process, and K-function for fitted LGCP.

Considering an inhomogeneous Thomas process with both topographical and soil covariates, minimum contrast parameter estimates $2.2{\times}10^{-4}$ and 13 are obtained for κ and ω. Note that smaller κ and ω yields less clustering. A comparison with the previous estimates

of κ and ω thus illustrates how the addition of the soil variables has decreased the amount of residual clustering. Using the results from Section 4.1 we can now add confidence intervals to these estimates. Since κ and ω are positive parameters we reparameterize and apply the asymptotic result to $\psi_1 = \log \kappa$ and $\psi_2 = \log \omega$. The approximate confidence intervals $(1.2 \times 10^{-4}; 3.9 \times 10^{-4})$ and $(10; 17)$ for σ^2 and ϕ are obtained by exponentiating approximate 95% confidence intervals based on the asymptotic normality of the estimates of ψ_1 and ψ_2. Following Section 4.2 we also evaluated the asymptotic covariance matrix replacing $\tilde{\Sigma}_n$ with Σ_n and obtained very similar results.

   We also consider a LGCP with an exponential covariance function for the Gaussian field. Proceeding as for the inhomogeneous Thomas process we obtain estimates 1.66 and 21 and confidence intervals $(1.0; 2.7)$ and $(12; 38)$ for σ^2 and ϕ. In this case we obtain somewhat different confidence intervals $(1.2; 2.3)$ and $(13; 33)$ when using Σ_n. Empirical standard errors of simulated estimates of $\log \sigma^2$ and $\log \phi$ obtained from 1000 simulations of the fitted LGCP model were 0.14 and 0.30. This indicates that both the asymptotic standard errors 0.18 and 0.24 obtained with Σ_n and the standard errors 0.24 and 0.30 obtained with $\tilde{\Sigma}_n$ are inaccurate. More simulation results are given in the next Section 5.1.

   The two models are qualitatively quite different with the LGCP providing the best fit to the estimated K-function, see Figure 2. Using the so-called J-summary statistic (Lieshout and Baddeley, 1996) for model assessment indicates that neither of the models are completely satisfactory since the inhomogeneous Thomas process is too tightly clustered while the LGCP has too many isolated points compared with the data.

5.1.   Simulation study

To check how the asymptotic results apply in finite-sample settings we consider simulation studies for an inhomogeneous Thomas process and an LGCP with exponential correlation function. The set-up for the simulation study is similar to the one in Waagepetersen (2007) with observation plot and altitude and slope covariates as in Section 2 and with the altitude and slope parameters β_2^* and β_3^* given by the parameter estimates in Waagepetersen (2007). In the case of the Thomas process we let $\psi = (\log \kappa, \log \omega)$ while $\psi = (\log \sigma^2, \log \phi)$ for the LGCP. In the simulation study we focus on the asymptotic normality of $\hat{\psi}$, the asymptotic standard errors for $\hat{\psi}$, and the coverage properties of approximate confidence intervals based on the asymptotic normality of the parameter estimates.

   In the case of an inhomogeneous Thomas process we vary (κ^*, ω^*) and the expected number of points μ^* to reflect varying degrees of clustering and tree abundance. We let ω^* equal to 10 or 20 while κ^* is 1×10^{-4} or 5×10^{-4} corresponding to expected numbers 50 or 250 of mother points within the plot and recall that larger κ^* and ω^* results in less clustering. The expected number μ^* of simulated points is either 200 or 800 corresponding to "sparse" and "moderately abundant" point patterns. For each combination of κ^*, ω^*, and μ^* we generate 1000 synthetic data sets and obtain simulated parameter estimates by applying our estimation procedure with $r = 100$ and $c = 0.25$. We compute the empirical standard deviation of the simulated parameter estimates and we evaluate for each simulation the asymptotic covariance matrix by plugging in the corresponding simulated parameter estimate. Approximate 95 % confidence intervals are constructed using standard errors extracted from the the estimated asymptotic covariance matrices. We only report results obtained with $\tilde{\Sigma}_n$ since similar results are obtained with Σ_n.

   Except for the 7th row, the simulation results in Table 5.1 shows fine agreement between the empirical standard errors and the median asymptotic standard errors for the simulated

Table 1. Columns 4-6: standard deviation for $\hat{\psi}_1 = \log \hat{\kappa}$ estimated from simulations, median of standard deviations obtained from estimated asymptotic covariance matrices, coverage of nominal 95 % approximate confidence intervals. Columns 7-9: as columns 4-6 but for $\hat{\psi}_2 = \log \hat{\omega}$.

κ^*	ω^*	μ^*	sd_1	$\hat{\text{sd}}_1$	cvrg_1	sd_2	$\hat{\text{sd}}_2$	cvrg_2
1×10^{-4}	10	200	0.28	0.28	0.94	0.15	0.16	0.96
1×10^{-4}	10	800	0.25	0.24	0.94	0.12	0.12	0.96
1×10^{-4}	20	200	0.39	0.38	0.95	0.19	0.2	0.96
1×10^{-4}	20	800	0.30	0.30	0.94	0.12	0.13	0.96
5×10^{-4}	10	200	0.47	0.48	0.98	0.30	0.32	0.95
5×10^{-4}	10	800	0.25	0.24	0.94	0.14	0.14	0.96
5×10^{-4}	20	200	1.72	2.24	1.00	0.72	1.44	1.00
5×10^{-4}	20	800	0.43	0.43	0.95	0.22	0.23	0.96

Table 2. Columns 4-6: standard deviation for $\hat{\psi}_1 = \log \hat{\sigma}^2$ estimated from simulations, median of standard deviations obtained from estimated asymptotic covariance matrices using Σ_n (and $\tilde{\Sigma}_n$ in parentheses), coverage of nominal 95 % approximate confidence intervals. Columns 7-9: as columns 4-6 but for $\hat{\psi}_2 = \log \hat{\phi}$.

$\sigma^{2,*}$	ϕ^*	μ^*	sd_1	$\hat{\text{sd}}_1$	cvrg_1	sd_2	$\hat{\text{sd}}_2$	cvrg_2
0.5	15	800	0.34	0.34 (0.34)	0.96	0.39	0.41 (0.42)	0.97
0.5	30	800	0.24	0.25 (0.26)	0.97	0.35	0.36 (0.37)	0.95
1.5	15	800	0.17	0.20 (0.23)	0.98	0.21	0.23 (0.25)	0.96
1.5	30	800	0.18	0.19 (0.23)	0.95	0.24	0.24 (0.28)	0.92

parameter estimates. The coverages of the confidence intervals are also fairly close (in general within 1%) to the nominal coverages of 95%. The problems in row 7 is probably due to that the parameter values $\kappa^* = 5\times10^{-4}$ and $\omega^* = 20$ corresponds to the least clustered case and with only 200 simulated points on average it may often be hard to distinguish the estimated K-function from that of a Poisson process. This leads to rather extreme values of the parameter estimates and for 3% of the simulated point patterns for row 7, the minimum contrast procedure did in fact not converge.

Quantile plots based on the simulated parameter estimates are shown in Figure 3 and Figure 4. The distributions of the parameter estimates are in general fairly close to normal. Bivariate scatter plots (omitted) of $(\log \hat{\kappa}, \log \hat{\omega})$ indicate that the joint distribution is well approximated by a bivariate normal and that $\log \hat{\kappa}$ is strongly negatively correlated with $\log \hat{\omega}$. However, for reasons discussed in the above paragraph, the case $\kappa^* = 5\times10^{-4}$, $\omega^* = 20$, and $\mu^* = 200$ is an exception where the distributions of both $\log \hat{\kappa}$ and $\log \hat{\omega}$ are very heavy tailed.

For the LGCP we restrict attention to the case $\mu^* = 800$ and values of $\sigma^{2,*} = 0.5, 1.5$ and $\phi^* = 15, 30$. Proceeding as for the Thomas process we obtain Table 5.1. The asymptotic results work rather well for $\sigma^{2,*} = 0.5$. For $\sigma^{2,*} = 1.5$ the asymptotic standard errors tend to overestimate the true standard errors and this is especially so for the asymptotic standard errors obtained with $\tilde{\Sigma}_n$. The quantile plots in Figure 5 show some deviations from normality both when $\sigma^{2,*} = 0.5$ and $\sigma^{2,*} = 1.5$ but the deviations seem rather modest.
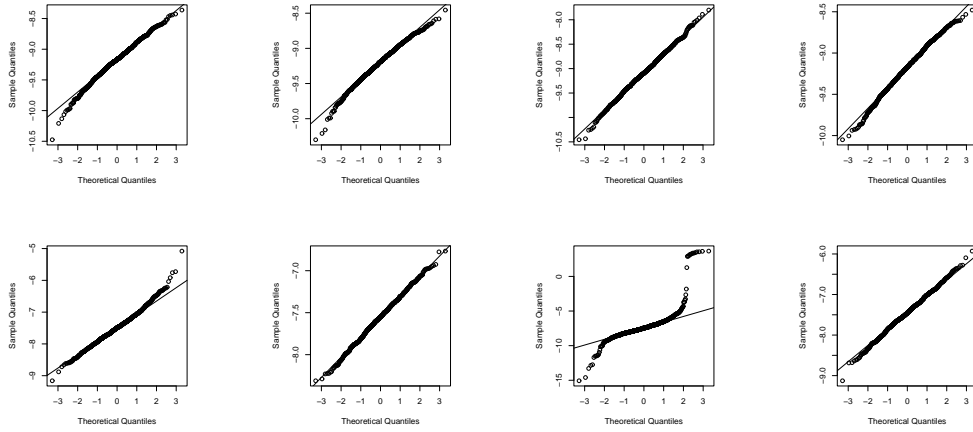
Fig. 3. Empirical quantiles of the simulated parameter estimates for $\log \kappa$ against quantiles for a standard normal distribution. Upper row, $\kappa^* = 1 \times 10^{-4}$, lower row: $\kappa^* = 5 \times 10^{-4}$. First two columns: $\omega^* = 10$ and second two columns: $\omega^* = 20$. First and third column: $\mu^* = 200$, second and fourth column: $\mu^* = 800$.
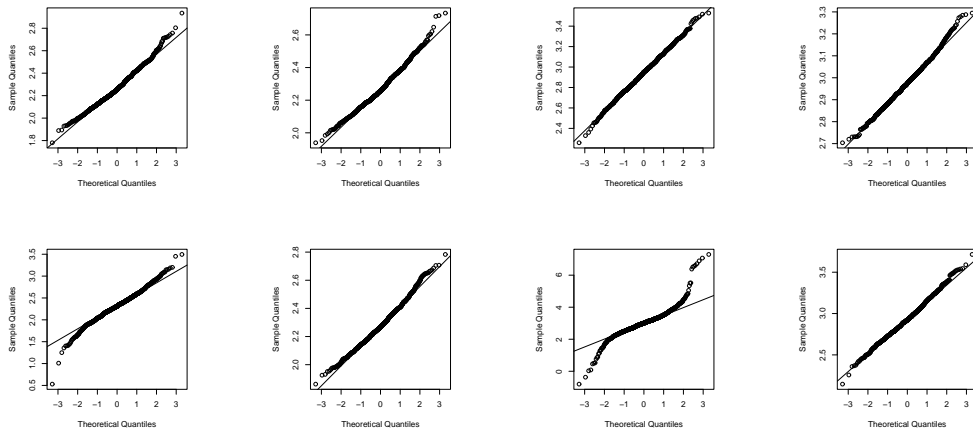

Fig. 4. Empirical quantiles for the simulated parameter estimates for $\log \omega$ against quantiles of a standard normal distribution. Upper row, $\kappa^* = 1 \times 10^{-4}$, lower row: $\kappa^* = 5 \times 10^{-4}$. First two columns: $\omega^* = 10$ and second two columns: $\omega^* = 20$. First and third column: $\mu^* = 200$, second and fourth column: $\mu^* = 800$.
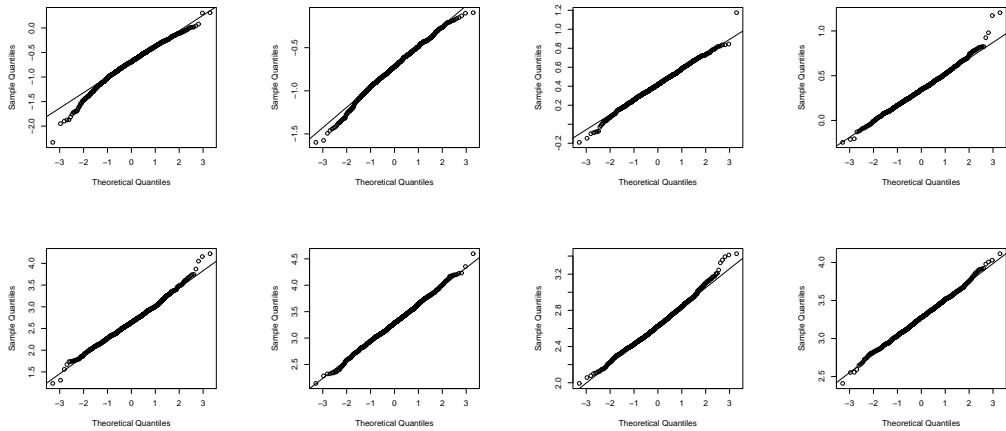
Fig. 5. Empirical quantiles for simulated parameter estimates of $\log \sigma^2$ (upper row) and $\log \phi$ (lower row) against quantiles of a standard normal distribution. From left to right: $(\sigma^{,2,*}, \phi^*) = (0.5, 15), (0.5, 30), (1.5, 15), (1.5, 30)$.

6. Discussion

The asymptotic result in Section 4.1 is applicable to a wide range of mixing inhomogeneous spatial point processes. In Section 5 we considered the spefic examples of an inhomogeneous Thomas process and an LGCP with exponential correlation function. The simulation study suggests that inference based on asymptotic normality is generally quite reliable with a possible exception when σ^2 is large for an LGCP. If one prefers a parametric bootstrap the $O_P(|W_n|^{-1/2})$ consistency implied by (6) is still a useful result.

A drawback of the minimum contrast estimation method is the need to specify r and c. The values $r = 100$ and $c = 0.25$ used in the data example and the simulation studies were taken from Waagepetersen (2007) who chose these values on basis of rules of thumbs from Diggle (2003). For the *Beilschmiedia* data and the Thomas process for example, we obtained estimates and confidence intervals 2.2×10^{-4} (1.2×10^{-4}; 3.9×10^{-4}) and 13 (10;17) for κ and ω. Changing r from 100 to 50 or c from 0.25 to 1 on the other hand yields 3.2×10^{-4} (2.1×10^{-4}; 4.7×10^{-4}) and 9.6 (8.4;11.1) respectively 1.7×10^{-4} (0.8×10^{-4}; 3.9×10^{-4}) and 20.1 (12.9;31.2). Hence quite different parameter estimates and estimates of uncertainty are obtained. The large sensitivity to r and c may partly be caused by the strong negative correlation between $\hat{\kappa}$ and $\hat{\omega}$. The product $\hat{\kappa}\hat{\omega}$ is e.g. not strongly affected by the choice of r and c. Similarly, estimated standard errors for $\hat{\beta}$ do not differ much when plugging in the different estimates of (κ, ω). For the LGCP, the parameter estimates are less sensitive as we obtain estimates (1.66,20.8), (1.82,17,4), and (1.40,24.7) for (σ^2, ϕ) using respectively $(r, c) = (100, 0.25)$, $(r, c) = (50, 0.25)$, and $(r, c) = (100, 1)$.

Simulation studies in Guan (2006) shows in concordance with Diggle (2003) that using $c = 0.25$ for aggregated point patterns generally works well. Regarding the choice of r, plots of the empirical pair correlation function (e.g. (4.21) in Møller and Waagepetersen, 2003) may be helpful. Typically, the pair correlation function converges to one and it is not helpful to use an r so that the pair correlation function becomes very close to one for a

range of values in $[0, r]$. For the *Beilschmiedia* data, inspection of the empirical correlation function suggests a value of r around 70.

Acknowledgment

A. Proof of joint asymptotic normality of $(\hat{\beta}_n, \hat{\psi}_n)$

Below we first establish the existence of a consistent sequence $\{(\hat{\beta}_n, \psi_n)\}_{n \geq 1}$ such that $u_n(\hat{\beta}_n, \hat{\psi}_n) = 0$ with a probability tending to one and $|W_n|^{1/2}\big((\hat{\beta}_n, \hat{\psi}_n) - (\beta^*, \psi^*)\big)$ is bounded in probability (i.e. for each $\epsilon > 0$ there exists a d such that $P(|W_n|^{1/2}\|(\hat{\beta}_n, \hat{\psi}_n) - (\beta^*, \psi^*)\| > d) \leq \epsilon$ for n sufficiently large). Asymptotic normality then follows from Lemma 4 and Lemma 5 in Appendix D, the boundedness of the entries in $\tilde{\Sigma}_n$ (Appendix B), and the Taylor expansion

$$|W_n|^{-1/2}u_n(\hat{\beta}_n, \hat{\psi}_n)\tilde{\Sigma}_n^{-1/2} =$$
$$|W_n|^{-1/2}u_n(\beta^*, \psi^*)\tilde{\Sigma}_n^{-1/2} + |W_n|^{1/2}\big((\hat{\beta}_n, \hat{\psi}_n) - (\beta^*, \psi^*)\big)\frac{J_n(\tilde{\beta}, \tilde{\psi})}{|W_n|}\tilde{\Sigma}_n^{-1/2} \quad (8)$$

where

$$J_n(\beta, \psi) = -\frac{\mathrm{d}}{\mathrm{d}(\beta, \psi)^{\mathsf{T}}}u_n(\beta, \psi) =$$
$$\begin{bmatrix} \frac{\mathrm{d}}{\mathrm{d}\beta^{\mathsf{T}}}u_{n,1}(\beta) & \frac{\mathrm{d}}{\mathrm{d}\beta^{\mathsf{T}}}u_{n,2}(\beta, \psi) \\ 0 & \frac{\mathrm{d}}{\mathrm{d}\psi^{\mathsf{T}}}u_{n,2}(\beta, \psi) \end{bmatrix} = \begin{bmatrix} J_{n,11}(\beta, \psi) & J_{n,12}(\beta, \psi) \\ 0 & J_{n,22}(\beta, \psi) \end{bmatrix} \quad (9)$$

and $(\tilde{\beta}, \tilde{\psi})$ is between $(\hat{\beta}_n, \hat{\psi}_n)$ and (β^*, ψ^*).

Regarding $|W_n|^{1/2}(\hat{\beta}_n - \beta^*)$, we apply Theorem 2 and Remark 1 in Appendix C to $u_{n,1}$ with $V_n = (|W_n|\tilde{\Sigma}_{n,11})^{1/2}$ and $c_n = |W_n|$. The conditions G2-G4 hold by Lemma 3-5 in Appendix D. It thus follows that there exists a sequence $\{\hat{\beta}_n\}_{n \geq 1}$ where $|W_n|^{1/2}\|\hat{\beta}_n - \beta^*\|$ is bounded in probability and $u_{n,1}(\hat{\beta}_n) = 0$ with a probability tending to one.

We proceed in a similar manner for $|W_n|^{1/2}(\hat{\psi}_n - \psi^*)$. Using a Taylor expansion,

$$|W_n|^{-1/2}u_{n,2}(\hat{\beta}_n, \psi^*)\tilde{\Sigma}_{n,22}^{-1/2} =$$
$$|W_n|^{-1/2}u_{n,2}(\beta^*, \psi^*)\tilde{\Sigma}_{n,22}^{-1/2} + |W_n|^{-1/2}(\hat{\beta}_n - \beta^*)J_{n,12}(\tilde{\beta}, \psi^*)\tilde{\Sigma}_{n,22}^{-1/2}$$

where $\|\tilde{\beta} - \beta^*\| \leq \|\hat{\beta}_n - \beta^*\|$. Letting $V_n = (|W_n|\tilde{\Sigma}_{n,22})^{1/2}$ it follows that $u_{n,2}(\hat{\beta}_n, \psi^*)V_n^{-1}$ is bounded in probability. Applying Theorem 2 in Appendix C to $u_{n,2}(\hat{\beta}_n, \psi)$ it follows as for $u_{n,1}$ that there exists a sequence $\{\hat{\psi}_n\}_{n \geq 1}$ where $|W_n|^{1/2}\|\hat{\psi}_n - \psi^*\|$ is bounded in probability and $u_{n,2}(\hat{\beta}_n, \hat{\psi}_n) = 0$ with a probability tending to one.

B.   The matrices $\tilde{\Sigma}_{n,12}$ and $\tilde{\Sigma}_{n,22}$

Note that we may rewrite $\tilde{u}_{n,2}(\beta^*, \psi^*)$ as

$$|W_n| \sum_{u,v \in X \cap W_n}^{\neq} \frac{f(u,v)}{|W_n \cap W_{n,u-v}|} - |W_n| 2c^2 \int_{r_l}^{r} K_{\psi^*}(t)^{2c-1} K_{\psi^*}^{(1)}(t) \mathrm{d}t$$

where

$$f(u,v) = \frac{2c^2 \int_{\max\{r_l, \|u-v\|\}}^{r} K_{\psi^*}(t)^{2c-2} K_{\psi^*}^{(1)}(t) \mathrm{d}t}{\rho_{\beta^*}(u) \rho_{\beta^*}(v)}.$$

Hence we can compute $\tilde{\Sigma}_{n,22} = |W_n|^{-1} \mathbb{V}\mathrm{ar}\tilde{u}_{n,2}(\beta^*, \psi^*)$ using the expansion (12) in Appendix D. Similarly, letting $h(u) = \rho_{\beta^*}^{(1)}(u)/\rho_{\beta^*}(u)$,

$$\begin{aligned}
\tilde{\Sigma}_{n,12} =& |W_n^{-1}| \mathbb{E}u_{n,1}(\beta^*)^{\mathsf{T}} \tilde{u}_{n,2}(\beta^*, \psi^*) \\
=& \int_{W_n^3} h(w) \frac{f(u,v)}{|W_n \cap W_{n,u-v}|} [\rho_{\beta^*,3}(w,u,v) - \rho_{\beta^*}(w)\rho_{\beta^*,2}(u,v)] \mathrm{d}w \mathrm{d}u \mathrm{d}v \\
& + 2 \int_{W_n^2} h(u) \frac{f(u,v)}{|W_n \cap W_{n,u-v}|} \rho_{\beta^*,2}(u,v) \mathrm{d}u \mathrm{d}v.
\end{aligned}$$

The boundedness of the entries in $\tilde{\Sigma}_{n,12}$ and $\tilde{\Sigma}_{n,22}$ follows from $f(u,v) = 0$ if $\|u - v\| > r$ and the basic assumptions B1-B4.

C.   A general asymptotic result

The following result is inspired by unpublished lecture notes by Professor Jens L. Jensen, University of Aarhus. Consider a parameterized family of probability measures P_θ, $\theta \in \mathbb{R}^p$, and a sequence of estimating functions $u_n : \mathbb{R}^p \to \mathbb{R}^p$, $n \geq 1$. The distribution of $\{u_n(\theta)\}_{n \geq 1}$ is governed by $P = P_{\theta^*}$ where θ^* denotes the 'true' parameter value. For a matrix $A = [a_{ij}]$, $\|A\|_M = \max_{ij} |a_{ij}|$ and we let $J_n(\theta) = -\frac{\mathrm{d}}{\mathrm{d}\theta^{\mathsf{T}}} u_n(\theta)$.

THEOREM 2.   *Assume that there exists a sequence of invertible symmetric matrices V_n such that*

G1 $\|V_n^{-1}\| \to 0.$
G2   *There exists a $l > 0$ so that $P(l_n < l)$ tends to zero where*

$$l_n = \inf_{\|\phi\|=1} \phi V_n^{-1} J_n(\theta^*) V_n^{-1} \phi^{\mathsf{T}}.$$

G3   *For any $d > 0$,*

$$\sup_{\|(\theta - \theta^*)V_n\| \leq d} \|V_n^{-1}[J_n(\theta) - J_n(\theta^*)]V_n^{-1}\|_M = \gamma_{nd} \to 0$$

in probability under P.
G4   *The sequence $u_n(\theta^*)V_n^{-1}$ is bounded in probability (i.e. for each $\epsilon > 0$ there exists a d so that $P(\|u_n(\theta^*)V_n^{-1}\| > d) \leq \epsilon$ for n sufficiently large).*

Then for each $\epsilon > 0$, there exists a $d > 0$ such that

$$P(\exists \tilde{\theta}_n : u_n(\tilde{\theta}_n) = 0 \ and \ \|(\tilde{\theta}_n - \theta^*)V_n\| < d) > 1 - \epsilon \tag{10}$$

whenever n is sufficiently large.

REMARK 1. *Suppose that there is a sequence $\{c_n\}_{n \geq 1}$ and matrices I_n so that $J_n(\theta^*)/c_n^2 - I_n$ tends to zero in probability. In condition G2 we can then replace $V_n^{-1}J_n(\theta^*)V_n^{-1}$ by $(V_n/c_n)^{-1}I_n(V_n/c_n)^{-1}$. Let $\hat{\theta}_n = 0$ if $u_n(\theta)$ has no solution and otherwise the root closest to θ^*. Then by (10), with a probability tending to one $\hat{\theta}_n$ is a root and $(\hat{\theta}_n - \theta^*)V_n$ is bounded in probability.*

PROOF. The event

$$\{\exists \hat{\theta}_n : u_n(\hat{\theta}_n) = 0 \text{ and } \|(\hat{\theta}_n - \theta^*)V_n\| < d\}$$

occurs if $u_n(\theta^* + \phi V_n^{-1})V_n^{-1}\phi^{\mathsf{T}} < 0$ for all ϕ with $\|\phi\| = d$ since this implies $u_n(\theta^* + \phi V_n^{-1}) = 0$ for some $\|\phi\| < d$ (Lemma 2 in Aitchison and Silvey, 1958). Hence we need to show that there is a d such that

$$P(\sup_{\|\phi\|=d} u_n(\theta^* + \phi V_n^{-1})V_n^{-1}\phi^{\mathsf{T}} \geq 0) \leq \epsilon$$

for sufficiently large n. To this end we write

$$u_n(\theta^* + \phi V_n^{-1})V_n^{-1}\phi^{\mathsf{T}} = u_n(\theta^*)V_n^{-1}\phi^{\mathsf{T}} - \phi \int_0^1 V_n^{-1}J_n(\theta(t))V_n^{-1}\mathrm{d}t\phi^{\mathsf{T}}$$

where $\theta(t) = \theta^* + t\phi V_n^{-1}$. Then

$$P(\sup_{\|\phi\|=d} u_n(\theta^* + \phi V_n^{-1})V_n^{-1}\phi^{\mathsf{T}} \geq 0) \leq$$

$$P(\sup_{\|\phi\|=d} u_n(\theta^*)V_n^{-1}\phi^{\mathsf{T}} \geq \inf_{\|\phi\|=d} \phi \int_0^1 V_n^{-1}J_n(\theta(t))V_n^{-1}\mathrm{d}t\phi^{\mathsf{T}}) \leq$$

$$P(\|u_n(\theta^*)V_n^{-1}\| \geq d \inf_{\|\phi\|=1}[\phi V_n^{-1}J_n(\theta^*)V_n^{-1}\phi^{\mathsf{T}}/2] - p\gamma_{nd}]) \leq$$

$$P(\|u_n(\theta^*)V_n^{-1}\| \geq dl_n/2) + P(p\gamma_{nd} > l_n/2)$$

The first term can be made arbitrarily small by picking a sufficiently large d and letting n tend to infinity. The second term converges to zero as n tends to infinity.

D. Auxiliary results

In this appendix we collect a number of lemmas used in the previous appendices. Recall that we always assume that the basic assumptions B1-B4 hold.

LEMMA 1. *The variance*

$$\mathrm{Var} \sum_{u,v \in X \cap W_n}^{\neq} \frac{1[\|u - v\| \leq t]f(u,v)}{|W_n|\rho_{\beta^*}(u)\rho_{\beta^*}(v)} \tag{11}$$

is $O(|W_n|^{-1})$ for any bounded function $f(u,v)$.

PROOF. Let $\phi(u,v) = \frac{1[\|u-v\|\le b_n t] f(u,v)}{|W_n|\rho_{\beta^*}(u)\rho_{\beta^*}(v)}$. Then by the Campbell formulae, (11) is equal to

$$2\int_{W_n^2} \phi(u,v)^2 \rho_{\beta^*}^{(2)}(u,v)\mathrm{d}u\mathrm{d}v + 4\int_{W_n^3} \phi(u,v)\phi(v,w)\rho_{\beta^*}^{(3)}(u,v,w)\mathrm{d}u\mathrm{d}v\mathrm{d}w+$$

$$\int_{W_n^4} \phi(u,v)\phi(w,z)(\rho_{\beta^*}^{(4)}(u,v,w,z) - \rho_{\beta^*}^{(2)}(u,v)\rho_{\beta^*}^{(2)}(w,z))\mathrm{d}u\mathrm{d}v\mathrm{d}w\mathrm{d}z. \quad (12)$$

It then follows from straightforward calculations that each of the three terms is $O(|W_n|^{-1})$.

LEMMA 2. *For any $d \in \mathbb{R}$,*

$$\sup_{r_l \le t \le r_u} |\hat{K}_{n,\beta^*}^d - K_{\psi^*}(t)^d|$$

is $o_P(1)$ for any $0 < r_l < r_u < \infty$. If $d \ge 0$ we may take $r_l = 0$.

PROOF. By Lemma 1, $\hat{K}_{n,\beta^*}(t)$ tends to $K_{\psi^*}(t)$ in probability for each $t \ge 0$. Using the monotonicity of $\hat{K}_{n,\beta^*}(t)^d$ and $K_{\psi^*}(t)^d$, the result follows by arguments as in the proof of the Glivenko-Cantelli theorem (e.g. page 266 in Van der Vaart, 1998).

LEMMA 3. *Assume N3. Then $\liminf_{n\to\infty} \min\{l_{n,11}, l_{n,22}\} > 0$ where*

$$l_{n,11} = \inf_{\|\phi_1\|=1} \phi_1 \tilde{\Sigma}_{n,11}^{-1/2} I_{n,11} \tilde{\Sigma}_{n,11}^{-1/2} \phi_1^{\mathsf{T}} \text{ and } l_{n,22} = \inf_{\|\phi_2\|=1} \phi_2 \tilde{\Sigma}_{n,22}^{-1/2} I_{22} \tilde{\Sigma}_{n,22}^{-1/2} \phi_2^{\mathsf{T}}.$$

PROOF. For a symmetric matrix A with eigen values λ_i and a vector ϕ it follows from the spectral decomposition that there exists another vector $\tilde{\phi}$ where $\|\phi\| = \|\tilde{\phi}\|$ and $\phi A\phi^T = \sum_i \tilde{\phi}_i^2 \lambda_i$. This implies that that the eigen values of $\tilde{\Sigma}_{n,11}$ and $\tilde{\Sigma}_{n,22}$ are bounded by the maximal eigenvalue $\tilde{\lambda}_{n,\max}$ of $\tilde{\Sigma}_n$ and that $\tilde{\lambda}_{n,\max} < \tilde{\lambda}_{\max}$ for some $\tilde{\lambda}_{\max} < \infty$ (since the entries in $\tilde{\Sigma}_n$ are bounded). Hence, the eigenvalues of $\tilde{\Sigma}_{n,11}^{-1}$ are greater than $1/\tilde{\lambda}_{\max}$, $\|\phi_1 \tilde{\Sigma}_{n,11}^{-1/2}\|^2 \ge 1/\tilde{\lambda}_{\max}$, and $l_{n,11} \ge \lambda_{n,11}/\tilde{\lambda}_{\max}$. Similarly, $l_{n,22} \ge \lambda_{22}/\tilde{\lambda}_{\max}$ where λ_{22} is the smallest eigen value of I_{22}.

LEMMA 4. *Assume N1-N2 and define $J_n(\beta, \psi)$ as in (9) in Appendix A.*

(a) *For any $d > 0$,*

$$\sup_{(\beta,\psi):\|((\beta,\psi)-(\beta^*,\psi^*))|W_n|^{1/2}\|\le d} \|J_n(\beta,\psi)/|W_n| - J_n(\beta^*,\psi^*)/|W_n|\|$$

tends to zero in probability.
(b) *$|W_n|^{-1}J_n(\beta^*,\psi^*) - I_n$ converges to zero in probability where I_n is given in (7).*

PROOF. The result (a) follows easily by arguments involving continuity of ρ_β and K_ψ and their derivatives. Regarding (b), we consider the blocks in J_n and I_n one at a time.

$$|W_n|^{-1}J_{n,11}(\beta^*,\psi^*) - I_{n,11} =$$

$$\frac{1}{|W_n|} \sum_{u\in X\cap W_n} \frac{(\rho_{\beta^*}^{(1)}(u))^{\mathsf{T}}\rho_{\beta^*}^{(1)}(u)}{\rho_{\beta^*}(u)^2} - I_{n,11} - \frac{1}{|W_n|} \sum_{u\in X\cap W_n} \frac{\rho_{\beta^*}^{(2)}(u)}{\rho_{\beta}^*(u)} + \frac{1}{|W_n|} \int_{W_n} \rho_{\beta^*}^{(2)}(u)\mathrm{d}u.$$

By the Campbell formulae $|W_n|^{-1}J_{n,11}(\beta^*,\psi^*)-I_{n,11}$ has mean zero and variance $O(|W_n|^{-1})$. Hence $|W_n|^{-1}J_{n,11}(\beta^*,\psi^*)-I_{n,11}$ tends to zero in probability. The matrix $|W_n|^{-1}J_{n,12}(\beta^*,\psi^*)$ is

$$-2c^2\int_{r_l}^r(\hat{K}_{n,\beta^*}(t)^{c-1}-K_{\psi^*}(t)^{c-1})K_{\psi^*}(t)^{c-1}K_{\psi^*}^{(1)}(t)\frac{\mathrm{d}}{\mathrm{d}\beta^{\mathsf{T}}}\hat{K}_{n,\beta}(t)|_{\beta=\beta^*}\mathrm{d}t+$$

$$-2c^2\int_{r_l}^r K_{\psi^*}(t)^{2c-2}K_{\psi^*}^{(1)}(t)\frac{\mathrm{d}}{\mathrm{d}\beta^{\mathsf{T}}}\hat{K}_{n,\beta}(t)|_{\beta=\beta^*}\mathrm{d}t.$$

where the first term tends to zero in probability by Lemma 2 and Lemma 1. The last term minus $I_{n,12}$ is

$$-2c^2\int_{r_l}^r K_{\psi^*}(t)^{2c-2}K_{\psi^*}^{(1)}(t)[\frac{\mathrm{d}}{\mathrm{d}\beta^{\mathsf{T}}}\hat{K}_{n,\beta}(t)|_{\beta=\beta^*}-H_{n,\beta^*(t)}]\mathrm{d}t$$

which tends to zero by Lemma 1. Regarding $J_{n,22}(\beta^*,\psi^*)$,

$$|W_n|^{-1}J_{n,22}(\beta^*,\psi^*)=I_{n,22}+$$
$$2c\int_{r_l}^r(\hat{K}_{n,\beta^*}(t)^c-K_{\psi^*}(t)^c)[(c-1)K_{\psi^*}(t)^{c-2}(K_{\psi^*}^{(1)}(t))^{\mathsf{T}}K_{\psi^*}^{(1)}(t)+K_{\psi^*}(t)^{c-1}K_{\psi^*}^{(2)}(t)]\mathrm{d}t$$

where the last term converges to zero in probability by Lemma 2.

LEMMA 5. *Assume N1-N5. Then* $|W_n|^{-1/2}\big(u_{n,1}(\beta^*),u_{n,2}(\beta^*,\psi^*)\big)\tilde{\Sigma}_n^{-1/2}$ *is asymptotically standard normal.*

PROOF. Note
$$u_{n,2}(\beta^*,\psi^*)=\tilde{u}_{n,2}(\beta^*,\psi^*)+V_{n,2}(\beta^*,\psi^*)$$
where

$$V_{n,2}(\beta^*,\psi^*)=$$
$$2c^2|W_n|\int_{r_l}^r(\hat{K}_{n,\beta^*}(t)-K_{\psi^*}(t))(\tilde{K}_n(t)^{c-1}-K_{\psi^*}(t)^{c-1})K_{\psi^*}(t)^{c-1}K_{\psi^*}^{(1)}(t)\mathrm{d}t$$

and $|\tilde{K}_n(t)-K_{\psi^*}(t)|\le|\hat{K}_{n,\beta^*}(t)-K_{\psi^*}(t)|$. The term $|W_n|^{-1/2}V_{n,2}(\beta^*,\psi^*)$ tends to zero in probability since $(\tilde{K}_n(t)^{c-1}-K_{\psi^*}(t)^{c-1})$ tends to zero uniformly in t by Lemma 2 and since $\mathbb{V}\mathrm{ar}|W_n|^{1/2}r\hat{K}_{n,\beta^*}(r)$ is $O(1)$. Hence $|W_n|^{-1/2}u_n(\beta^*,\psi^*)\tilde{\Sigma}_n^{-1/2}$ has the same weak limit as $|W_n|^{-1/2}\big(u_{n,1}(\beta^*),\tilde{u}_{n,2}(\beta^*,\psi^*)\big)\tilde{\Sigma}_n^{-1/2}$.

For the asymptotic normality of $|W_n|^{-1/2}\big(u_{n,1}(\beta^*,\psi^*),\tilde{u}_{n,2}(\beta^*,\psi^*)\big)$, let $s=\sqrt{4r^2+\epsilon/2}-2r$ where $\epsilon=a-8r^2>0$, cf. N5. For $(i,j)\in\mathbb{Z}^2$, let $A_{ij}=[is,(i+1)s)\times[js,(j+1)s)$ be the $s\times s$ box with lower right corner at (is,js) and define

$$X_{ij}=\sum_{u\in X\cap A_{ij}}\frac{\rho_\beta^{(1)}(u)}{\rho_\beta(u)}-\int_{A_{ij}}\rho_\beta^{(1)}(u)$$

whereby
$$|W_n|^{-1/2}u_{n,1}(\beta)=|W_n|^{-1/2}\sum_{(i,j)\in\mathbb{Z}^2:A_{ij}\subseteq W_n}X_{ij}+o_P(1).$$

Regarding $\tilde{u}_{n,2}$ we replace $\hat{K}_{n,\beta^*}(t)$ by

$$\frac{1}{|W_n|} \sum_{u \in X \cap W_n} \sum_{v \in X} \frac{1[0 < \|u-v\| \leq t]}{\rho_{\beta^*}(u)\rho_{\beta^*}(v)}$$

and define

$$
\begin{aligned}
Y_{ij} &= 2c^2 \sum_{u \in X \cap A_{ij}} \int_{r_l}^{r} \sum_{v \in X} \frac{1[0 < \|u-v\| \leq t]}{\rho_{\beta^*}(u)\rho_{\beta^*}(v)} K_{\psi^*}(t)^{2c-2} K_{\psi^*}^{(1)}(t) \mathrm{d}t \\
&\quad - 2c^2 s^2 \int_{r_l}^{r} K_{\psi^*}(t)^{2c-1} K_{\psi^*}^{(1)}(t) \mathrm{d}t
\end{aligned}
$$

whereby

$$|W_n|^{-1/2} \tilde{u}_{n,2}(\beta^*, \psi^*) = |W_n|^{-1/2} \sum_{(i,j) \in \mathbb{Z}^2 : A_{ij} \subseteq W_n} Y_{ij} + o_P(1).$$

Let $x = (x_1, \cdots, x_p)$ and $y = (y_1, \cdots, y_q)$ be two arbitrary non-zero vectors and define

$$Z_{ij} = X_{ij} x^\mathsf{T} + Y_{ij} y^\mathsf{T}.$$

Asymptotic normality of $|W_n|^{-1/2} \sum_{(i,j) \in \mathbb{Z}^2} Z_{ij}$ now implies the asymptotic normality of $|W_n|^{-1/2} \big(u_{n,1}(\beta^*), \tilde{u}_{n,2}(\beta^*, \psi^*)\big)$. Let $|\Lambda|$ denote cardinality of a subset $\Lambda \subseteq \mathbb{Z}^2$ and $\mathcal{F}(Z, \Lambda)$ the σ-algebra generated by $\{Z_{ij} : (i,j) \in \Lambda\}$. Define the mixing coefficient

$$
\begin{aligned}
\alpha_{p_1,p_2}(m; Z) = \sup\{&|P(A_1 \cap A_2) - P(A_1)P(A_2)| : A_i \in \mathcal{F}(Z, \Lambda_i), |\Lambda_i| \leq p_i, \\
&\Lambda_i \subseteq \mathbb{Z}_2, i = 1, 2, d(\Lambda_1, \Lambda_2) \geq m\}.
\end{aligned}
$$

Since the random field $Z = \{Z_{ij} : (i,j) \in \mathbb{Z}^2\}$ inherits the mixing properties of X we can now invoke the central limit Theorem 3.3.1 in Guyon (1991) which is an extension to the nonstationary case of Bolthausen (1982)'s central limit theorem. Specifically, we need for some $\delta > 0$,

(a) $\liminf_{n \to \infty} |W_n|^{-1}(x,y) \mathbb{V}\mathrm{ar}\big((u_{n,1}(\beta^*), \tilde{u}_{n,2}(\beta^*, \psi^*))\big)(x,y)^\mathsf{T} > 0$,
(b) $\sup_{ij} \mathbb{E}(|Z_{ij}|^{2+\delta}) < \infty$
(c) $\sum_{m \geq 1} m \alpha_{2,\infty}(m; Z)^{\delta/(2+\delta)} < \infty$,

These conditions hold due to N3, N4, and N5, respectively. Note in particular regarding the last condition that Y_{ij} and hence Z_{ij} only depends on X through $X \cap A_{ij} \oplus r$ where $A_{ij} \oplus r = [is - r, i(s+1) + r) \times [js - r, j(s+1) + r)$ whose area equals $a/2$.

E.   A sufficient condition for mixing for Neyman-Scott processes

Recall the definition in Section 3 of a Neyman-Scott process $X = \cup_{c \in C} X_c$ where the X_c are independent offspring Poisson processes with intensity functions $\alpha k(\cdot - c)$ and k is the dispersal density for the offspring. Below we verify that a sufficient condition for mixing is that

$$\int_{\mathbb{R}^2 \setminus [-m,m]^2} k(v - w) \mathrm{d}v \text{ is } O(m^{-d-2}) \tag{13}$$

whenever the distance from w to $\mathbb{R}^2 \setminus [-m,m]^2$ is bigger than $m/2$.

Consider regions $E_1 = [-h, h]^2$, and $E_2 = \mathbb{R}^2 \setminus [-m, m]^2$ where $m = 2n > h$. Let $X_1 = \cup_{c \in C \cap [-n,n]^2} X_c$ and $X_2 = X \setminus X_1$. Then X_1 and X_2 are independent cluster processes. Let $A_i = \{X \cap E_i \in G_i\}$, $i = 1, 2$, where G_1 and G_2 are sets of point configurations. Further let $B_1 = \{X_1 \cap E_2 = \emptyset\}$, $B_2 = \{X_2 \cap E_1 = \emptyset\}$ and $B = B_1 \cap B_2$. Then

$$P(A_1 \cap A_2) = P(A_1 \cap A_2 \cap B) + P(A_1 \cap A_2 \cap B^c)$$

where

$$P(A_1 \cap A_2 \cap B) = P(X_1 \cap E_1 \in G_1, X_1 \cap E_2 = \emptyset)P(X_2 \cap E_2 \in G_2, X_2 \cap E_1 = \emptyset).$$

Similarly,

$$\begin{aligned}
&P(A_1)P(A_2) \\
&= P(X_1 \cap E_1 \in G_1, X_1 \cap E_2 = \emptyset)P(X_2 \cap E_2 \in G_2, X_2 \cap E_1 = \emptyset)P(B) \\
&\quad + P(A_1 \cap B)P(A_2 \cap B^c) + P(A_1 \cap B^c)P(A_2 \cap B) + P(A_1 \cap B^c)P(A_2 \cap B^c).
\end{aligned}$$

Thus,

$$|P(A_1 \cap A_2) - P(A_1)P(A_2)| \leq 5P(B^c) \leq 5P(B_1^c) + 5P(B_2^c)$$

Let $n(X_1 \cap E_2)$ denote the cardinality of $X_1 \cap E_2$. Then

$$P(B_1^c) \leq \mathbb{E}n(X_1 \cap E_2) = \alpha\kappa \int_{[-n,n]^2} \int_{\mathbb{R}^2 \setminus [-m,m]^2} k(u - c)\mathrm{d}u\mathrm{d}c$$

and

$$P(B_2^c) \leq \mathbb{E}n(X_2 \cap E_1) = \alpha\kappa \int_{[-h,h]^2} \int_{\mathbb{R}^2 \setminus [-n,n]^2} k(u - c)\mathrm{d}c\mathrm{d}u.$$

Both of these are $O(m^{-d})$ if (13) holds.

References

Aitchison, J. & Silvey, S. D. (1958). Maximum-likelihood estimation of parameters subject to restraints. *The Annals of Mathematical Statistics* **29**, 813–825.

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

Bolthausen, E. (1982). On the central limit theorem for stationary mixing random fields. *The Annals of Probability* **10**, 1047–1050.

Burslem, D. F. R. P., Garwood, N. C. & Thomas, S. C. (2001). Tropical forest diversity: the plot thickens. *Science* **291**, 606–607.

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

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

Diggle, P. J. (2003). *Statistical Analysis of Spatial Point Patterns*. Oxford University Press, 2nd edition.

Doukhan, P. (1994). *Mixing: properties and examples*. Number 85 in Lecture Notes in Statistics, Springer, New York.

Guan, Y. (2006). On model fitting procedures for inhomogeneous Neyman-Scott processes. Preprint.

Guan, Y. & Loh, J. M. (2007). A thinned block bootstrap procedure for modeling inhomogeneous spatial point patterns. *Journal of the American Statistical Association*, to appear.

Guan, Y. & Sherman, M. (2007). On least squares fitting for stationary spatial point processes. *Journal of the Royal Statistical Society Series B* **69**, 31–49.

Guyon, X. (1991). *Random fields on a network*. Springer-Verlag, New York.

Heinrich, L. (1992). Minimum contrast estimates for parameters of spatial ergodic point processes. In: *Transactions of the 11th Prague Conference on Random Processes, Information Theory and Statistical Decision Functions*, Academic Publishing House, Prague, 479–492.

Hubbell, S. P. (2001). *The unified neutral theory of biodiversity and biogeography*. Number 32 in Monographs in Population Biology, Princeton University Press.

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

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

Losos, E. C. & Leigh, Jr., E. G., eds. (2004). *Tropical Forest Diversity and Dynanism. Findings from a Large-Scale Plot Network*. The University of Chicago Press, Chicago.

Møller, J. & Waagepetersen, R. P. (2003). *Statistical inference and simulation for spatial point processes*. Chapman and Hall/CRC, Boca Raton.

Politis, D. M., Paparoditis, E. & Romano, J. P. (1998). Large sample inference for irregularly spaced dependent observations based on subsampling. *Sankhya* **60**, 274–292.

Schoenberg, F. P. (2005). Consistent parametric estimation of the intensity of a spatial-temporal point process. *Journal of Statistical Planning and Inference* **128**, 79–93.

Seidler, T. G. & Plotkin, J. B. (2006). Seed dispersal and spatial pattern in tropical trees. *PLoS Biology* **4**, 2132–2137.

Van der Vaart, A. W. (1998). *Asymptotic statistics*. Cambridge University Press, Cambridge.

Waagepetersen, R. (2004). Convergence of posteriors for discretized log Gaussian Cox processes. *Statistics and Probability Letters* **66**, 229–235.

Waagepetersen, R. (2007). An estimating function approach to inference for inhomogeneous Neyman-Scott processes. *Biometrics* **63**, 252–258.