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

In this paper we study theoretical properties of an estimation procedure for inhomogeneous spatial point processes and show how the results may be used in studies of 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). Aggregation of trees of the same species 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 may also be due to seed dispersal around parent trees. Incisive studies of the different diversity hypotheses require statistical methods which allow to disentangle the various sources of aggregation.

In recent years huge amounts of data have been collected in tropical rain forest plots. The data sets consist of measurements of soil properties, digital terrain models, and individual locations of all trees growing in the plots. 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.

The second-order properties of X are determined by the so-called pair correlation function g (see Section 2). Translation-invariance of g implies second-order intensity reweighted

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

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. (1) in Section 2). 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. study the relation between estimates of ψ and the modes of seed dispersal for different species. They, however, neglect possible aggregation due to covariates. For a Poisson process, g is identically equal to one due to the independence properties of Poisson processes.

If X is not a Poisson process, evaluation of the likelihood function in general requires application of Markov chain Monte Carlo (Møller and Waagepetersen, 2003) which especially in case of Cox and cluster processes can be computationally very difficult. However, regardless whether X is a Poisson process or not, a computationally easy approach to estimate β is to use the Poisson likelihood score function as an estimating function (Schoenberg, 2005), see Section 3. Consistency of the resulting estimate $\hat{\beta}$ is studied in Schoenberg (2005) while Waagepetersen (2007) obtains asymptotic normality of $\hat{\beta}$ for a fixed observation window employing infinite divisibility of the inhomogeneous Neyman-Scott processes considered in this paper. Regarding ψ, Waagepetersen (2007) suggests 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, does not provide a theoretical study of this approach for estimating ψ.

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 biased 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 extend the results for $\hat{\psi}$ in Heinrich (1992) and Guan and Sherman (2007) to the important case of inhomogeneous point processes which are indispensable in modern point process statistics. We moreover provide less restrictive conditions for the asymptotic normality of $\hat{\beta}$ than those in Waagepetersen (2007) and Guan and Loh (2007).

Our asymptotic results for $\hat{\psi}$ are important for two reasons. First, the asymptotic normality enables inference regarding ψ when hypotheses regarding the clustering of X are considered. Second, the consistency of $\hat{\psi}$ gives a theoretical basis for the plug-in approach in Waagepetersen (2007) who suggested to plug in $g_{\hat{\psi}}$ for the unknown g when evaluating the asymptotic covariance matrix for $\hat{\beta}$.

Some background concerning Cox and cluster point processes and product densities is given and some basic assumptions stated in Section 2. The two-step procedure for parameter estimation and its asymptotic properties are considered in Section 3. In Section 4 we consider a data example where we relate clustering to different modes of seed dispersal as in Seidler and Plotkin (2006). We, however, refine the analysis by taking into account inhomogeneity due to environmental covariates and uncertainty of clustering parameter estimates. A few open problems are discussed in Section 5.

2. 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.

2.1. 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 Thomas process with e.g. a large κ and a small ω is composed of many spatially tight clusters while a small κ and a large ω produces few and widely dispersed clusters.

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). Shot-noise Cox processes provide a further extension where the fixed mean cluster size α is replaced by a random variable, see Møller (2003) and Waagepetersen and Schweder (2006).

2.2. Product densities and basic assumptions

Let $\rho_{\beta,k}(u_1, \ldots, u_k)$ denote the kth order product density of X. For locations u_i in infinitesimally small regions A_i of volumes $\mathrm{d}A_i$, $i = 1, \ldots, k$, $\rho_{\beta,k}(u_1, \ldots, u_k)\mathrm{d}A_1 \cdots \mathrm{d}A_k$ 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 the following:

B1 bounded covariates

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

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

$$\rho_{\beta,k}(u_1, \ldots, u_k) = \rho_k(u_1, \ldots, u_k) \prod_{i=1}^{k} \rho_\beta(u_i)$$

where ρ_k is the kth 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 $W_n = [an, bn] \times [cn, dn]$ 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. Letting $\rho_\beta^{(l)}$ denote the lth order derivative of ρ_β, B1 implies $k_1 \leq \rho_\beta(u), |\rho_\beta^{(l)}(u)| \leq$

K_3, $l = 1, 2$, 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 $\rho_k(u_1, \dots, u_k) = \rho_k(0, u_2 - u_1, \dots, u_k - u_1)$. 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

$$K(t) = \int_{\|h\| \le t} g(h) \mathrm{d}h, \quad t \ge 0. \tag{1}$$

If g is isotropic, $g(h)$ can be recovered from $K(t)$ by differentiating: $g(h) = K'(\|h\|)/(2\pi\|h\|)$.

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 this 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.

We call assumptions B1-B4 basic assumptions since they allow us to verify (see Appendix D) the basic property of consistency for the semi-parametric K-function estimate (3) in Section 3. Hence these assumptions are not specific for our result on asymptotic normality in Section 3.1.

2.3. Specific examples of pair correlation functions

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 a 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.

Note that although the parameters of a Thomas process or a log Gaussian Cox process have distinct interpretations in terms of the cluster generating mechanism or the Gaussian field this is not entirely so in relation to the pair correlation function. For a given h and the Thomas process, for example, both increasing values of κ and decreasing values of ω lead to decreasing pair correlation $g_{(\kappa,\omega)}(h)$. Similarly, for a log Gaussian Cox process with exponential covariance function, both decreasing σ^2 and decreasing ϕ lead to decreasing $g_\psi(h)$.

The K-function for a Thomas process is

$$K_{(\kappa,\omega)}(t) = \pi t^2 + [1 - \exp(-t^2/(4\omega^2))]/\kappa$$

where πt^2 is the K-function for a Poisson process. Hence, the difference between the Thomas process and the Poisson process K-functions is less than $1/\kappa$ and increases monotonely to $1/\kappa$ as t increases. The increase to $1/\kappa$ is quicker when ω is small. In other words, for fixed t and κ, $K_{(\kappa,\omega)}(t)$ tends to the Poisson K-function at t when ω increases.

3. The two-step estimation procedure

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

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

If X is Poisson, $u_{n,1}$ is simply the score function given by the first derivative of the log likelihood function. Second, $\hat{\psi}_n$ is obtained by minimizing $m_n(\psi) = m_{n,\hat{\beta}_n}(\psi)$ where

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

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

$$\hat{K}_{n,\beta}(t) = \sum_{u,v \in X \cap W_n}^{\neq} \frac{1[\|u-v\| \leq t]}{\rho_\beta(u)\rho_\beta(v)|W_n \cap W_{n,u-v}|} \tag{3}$$

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 3.1 we for technical reasons need $r_l > 0$ when $c < 1$. As in Guan and Sherman (2007) it is possible to introduce a weight function in (2) but we skip this to keep the notation simpler.

3.1. *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

$$I_{n,11} = \frac{1}{|W_n|} \int_{W_n} \frac{(\rho_{\beta^*}^{(1)}(u))^\mathsf{T} \rho_{\beta^*}^{(1)}(u)}{\rho_{\beta^*}(u)} \mathrm{d}u, \tag{4}$$

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

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

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

We finally need the mixing coefficient (Politis *et al.*, 1998)

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

where $\mathcal{B}(\mathbb{R}^2)$ 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 $X \cap E_i$, $i = 1, 2$.

The following result is verified in Appendix A.

THEOREM 1. *Consider a point process X with intensity function ρ_{β^*} and K-function K_{ψ^*}. In addition to B1-B4 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,* $\tilde{\Sigma}_n$ *converges to a positive definite matrix* $\tilde{\Sigma}$, *and* $\liminf_{n\to\infty} \lambda_{n,11} > 0$ *where* $\lambda_{n,11}$ *is the smallest eigenvalue of* $I_{n,11}$.
N4 $\rho_{4+2\delta}(u_1, \cdots, u_{4+2\delta}) < \infty$ *for some positive integer* δ.
N5 for some $a > 8r^2$,

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

Then there is a sequence $\{(\hat{\beta}_n, \hat{\psi}_n)\}_{n\ge1}$ *for which* $u_n(\hat{\beta}_n, \hat{\psi}_n) = 0$ *with a probability tending to one and where*

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

with

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

In the following two sections 3.2 and 3.3 we discuss in more detail the practical use of this result and the conditions for it. A simulation study is summarized in Section 3.4.

3.2. *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

$$\frac{1}{|W_n|} \int_{W_n^2} (\rho_{\beta^*}^{(1)}(u))^\mathsf{T} \rho_{\beta^*}^{(1)}(v) (g_{\psi^*}(u-v) - 1) \mathrm{d}u\mathrm{d}v. \tag{8}$$

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

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 (3). 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. The intuition is that the estimated $\hat{\beta}$ adjust for the variation of the number of points used in the estimate of K. A curious consequence is that, in terms of the asymptotic variance for $\hat{\psi}_n$, we are better off using $\hat{\beta}_n$ in the minimum contrast estimation even if the true β^* was known.

3.3. *Discussion of conditions for asymptotic normality*

The condition N1 is a nuisance but 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 integral (2) using an equispaced right endpoint Riemann sum. For the numerical value of this approximation the specific choice of r_l is not crucial as long as we just use a very small value of r_l. Hence, in practice we just use $r_l = 0$. 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 there exist distinct $r_l < t_1 < t_2 < \cdots < t_q < r$ so that the matrix with rows $K_{\psi}^{(1)}(t_i)$ has full rank. For a Thomas process with $\psi = (\log\kappa, \log\omega)$, for example, this is easy to verify with $t_1 = 2\omega\sqrt{-\log(0.5)}$ and $t_2 = \sqrt{2}t_1$. It is hard to say something general about the conditions on $\tilde{\Sigma}_n$ and the smallest eigenvalue of $I_{n,11}$ since they depend on the behaviour of the covariates on all of $\mathbb{R}^2$. The convergence of $\tilde{\Sigma}_n$ may be viewed as convergence of a spatial average over W_n cf. e.g. (4) and (8). Note moreover that $\tilde{\Sigma}_n$ 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). The conditions in Kolmogorov and Rozanov (1960) for strong mixing of Gaussian processes on $\mathbb{R}$ e.g. hold in the case of an exponential covariance function but Kolmogorov and Rozanov (1960) do not consider spatial processes.

3.4. Simulation study

To check how the asymptotic results apply in finite-sample settings we considered a simulation study for an inhomogeneous Thomas process and an LGCP with exponential covariance function. Both point processes were simulated with varying parameter settings within a 1000 by 500 meter region and we focused on normality of parameter estimates obtained from the simulations and coverage properties of approximate 95 % confidence intervals. The simulation study in general confirmed that valid inferences can be based on the asymptotic results even for moderately sized point patterns (the expected numbers of simulated points were either 200 or 800). One exception was the case of an inhomogeneous Thomas process with parameter values $\kappa^* = 5{\times}10^{-4}$, $\omega^* = 20$, and on average 200 simulated points. This is a case of weak clustering with on average less than one offspring per mother. With the small numbers of simulated points it was often hard to distinguish the K-function estimated from the simulated data from the K-function of a Poisson process. In such cases rather extreme clustering parameter estimates were obtained since the Poisson process is not nested within the inhomogeneous Thomas process. From a practical point of view one would in such cases typically use a Poisson process model anyway. For LGCPs, larger variances for the Gaussian field seemed to imply slower convergence to normality. Details of the entire simulation study are given in Waagepetersen and Guan (2008).

4. A data example

The tropical tree data sets considered in this section are 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 plots in Figure 1 show positions of alive trees in 1995 of the species *Acalypha diversifolia* (528 trees), *Lonchocarpus heptaphyllus* (836 trees), *Capparis frondosa* (3299 trees), and altitude on a 5 by 5 meter grid. The three species have distinct modes of seed dispersal (Wright *et al.*, 2007). The seeds are dispersed by exploding capsules for *Acalypha*, by the wind for *Lonchocarpus*, and by birds and mammals for *Capparis*. Seidler and Plotkin (2006) hypothesize that the modes of seed dispersal is reflected in the spatial patterns of tree locations with tight clusters for exploding capsules, loose clusters for bird and mammal dispersal, and tightness of clustering somewhere in between for species with wind dispersal. Seidler and Plotkin (2006) fit homogeneous Thomas processes to a large number of species and quantify tightness of clustering using the parameter ω. We restrict attention to only three species but take into account inhomogeneity due to the environment and uncertainty of clustering parameter estimates.

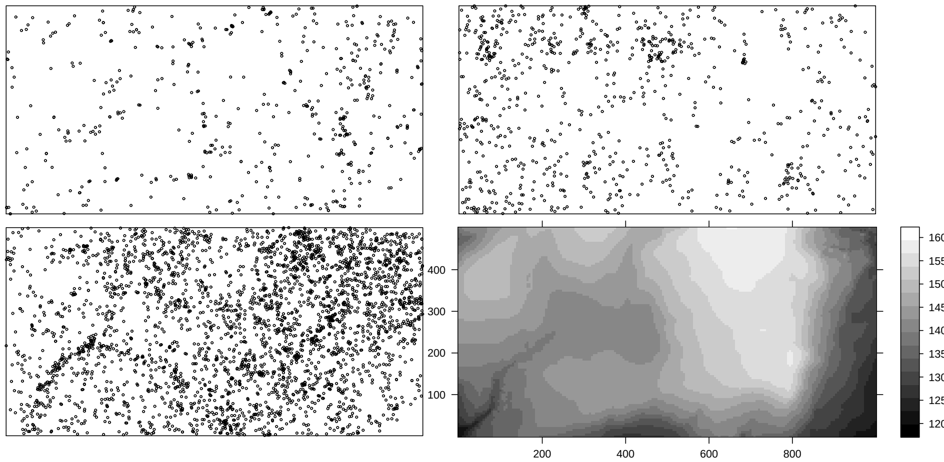

Fig. 1. Locations of *Acalypha diversifolia* (top left), *Lonchocarpus heptaphyllus* (top right), *Capparis frondosa* (bottom left), and altitude (bottom right).

Note that the parameter ω is just a quantitative summary of clustering and it can not be directly related to e.g. a mean range of seed dispersal around trees of a given species. This is because a Thomas process only involves one generation of offspring while the patterns of rain forest trees are produced by multiple generations of offspring.

For each species we fitted an inhomogeneous Thomas process with

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

where the covariate vector $z(u)$ had entries altitude, slope, pH, and soil contents of mineralized nitrogen, phosphorous, and potassium. A stepwise model reduction was subsequently performed using Wald-tests based on approximate normality of the regression parameter estimates (cf. (6)). Covariates significant at the 5 % level were elevation and potassium for *Acalypha* and *Capparis* and nitrogen and phosphorous for *Lonchocarpus*. For *Acalypha*, the estimates of the regression parameters for elevation and potassium were 0.02 (0.004,0.04) and 0.005 (0.002;0.007), respectively (approximate 95 % confidence intervals in parantheses). For *Capparis* we obtained 0.03 (0.01;0.05) and 0.004 (0.002;0.006). Hence *Acalypha* and *Capparis* appear to be similar in terms of niche assembly characteristics with preference for high elevation and high potassium content. *Lonchocarpus* is quite different and with nitrogen and phosphorous parameter estimates -0.03 (-0.04;-0.02) and -0.15 (-0.27;-0.03) it seems to be a frugal species adapted to soils with low nutrition contents.

The left plot in Figure 2 shows the estimates of ω obtained using only the significant covariates and with values $r = 100$ and $c = 0.25$ chosen on basis of rules of thumb in Diggle (2003). The lines in the plot show approximate 95 % confidence intervals based on approximate normality of $\log \hat{\omega}$. According to exploratory analyses using cross K-functions, the three species may be considered uncorrelated. Based on approximately normal test-statistics given by differences of the log parameter estimates and assuming independence of the parameter estimates for different species we can clearly reject that *Acalypha* and *Lonchocarpus* share the same ω while this is not the case when considering *Lonchocarpus* and *Capparis*. For comparison we also show the estimates obtained for homogeneous Thomas processes, i.e. without adjusting for the covariate effects. It appears that the estimates of

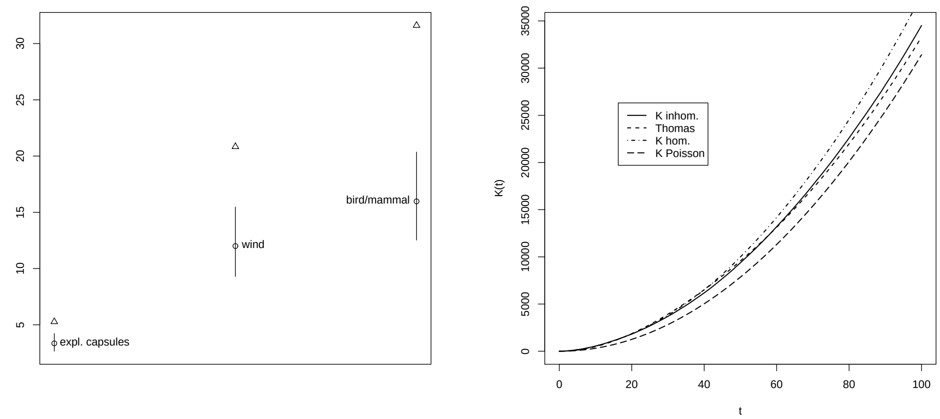

Fig. 2. Left: dots show estimates of ω for *Acalypha* (expl. capsules), *Lonchocarpus* (wind), and *Capparis* (bird/mammal) (left to right). The segments indicate approximate 95 % confidence intervals and triangles show estimates of ω obtained without adjusting for aggregation due to covariates. Right: curves show for the *Capparis* data K-function estimated adjusting for inhomogeneity, K-function for fitted Thomas process, K-function estimated assuming homogeneity, and K-function for Poisson process.

ω are very sensitive to whether covariate effects are accounted for or not. The right plot of Figure 2 shows estimates of the K-function for the *Capparis* data assuming respectively inhomogeneity due to covariates and homogeneity. A parametric bootstrap based on simulation from the fitted inhomogeneous Thomas process shows that the difference between the K-function estimates can not be explained by sampling variation.

We carried out model assessment using the J-function (Lieshout and Baddeley, 1996) and visual comparison of the point patterns and simulations from the fitted models. There is scope for improving the modelling of the *Acalypha* and *Lonchocarpus* point patterns since, compared with these data, the fitted inhomogeneous Thomas processes produce point patterns with too distinct clusters and too much empty space between the clusters. We also fitted LGCPs with exponential covariance functions and in terms of the J-function and visually, these models provide better fits for the *Acalypha* and *Lonchocarpus* data. We leave it as a topic of further research to explore in more detail LGCPs, Neyman-Scott processes with other types of offspring densities, or perhaps more flexible classes of cluster processes like generalized shot-noise Cox processes (Møller and Torrisi, 2005).

5. Discussion

The two-step estimation procedure only requires specification of the intensity function and the pair correlation function. In this paper we considered parametric pair correlation functions generated by specific point process models. An interesting open problem is, given a proposed model for the pair correlation, how to determine whether there indeed exist a point process with this pair correlation. The evaluation of the asymptotic covariance matrix for the pair correlation parameters moreover requires a consistent specification of the third and fourth order product densities.

A drawback of the minimum contrast estimation method is the need to specify r and c. Experiments for a Thomas process show that the estimates of κ and ω are quite sensitive to the choice of r and c. This sensitivity may partly be linked to 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 estimates of (κ, ω) obtained with different r and c. For an LGCP, the parameter estimates appear to be less sensitive to r and c.

Simulation studies in Guan (2006) show 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 beyond the point where the pair correlation function has essentially converged to one.

Acknowledgment

Rasmus Waagepetersen's and Yongtao Guan's research was supported respectively by grant 272-06-0442 'Point process modelling and statistical inference' from the Danish Natural Science Research Council and NSF grant DMS-0706806. The BCI forest dynamics research project was made possible by National Science Foundation grants to Stephen P. Hubbell: DEB-0640386, DEB-0425651, DEB-0346488, DEB-0129874, DEB-00753102, DEB-9909347, DEB-9615226, DEB-9615226, DEB-9405933, DEB-9221033, DEB-9100058, DEB-8906869, DEB-8605042, DEB-8206992, DEB-7922197, support from the Center for Tropical Forest Science, the Smithsonian Tropical Research Institute, the John D. and Catherine T. MacArthur Foundation, the Mellon Foundation, the Celera Foundation, and numerous private individuals, and through the hard work of over 100 people from 10 countries over the past two decades. The plot project is part of the Center for Tropical Forest Science, a global network of large-scale demographic tree plots. The BCI soils data set were collected and analyzed by J. Dalling, R. John and K. Harms with support from NSF DEB021104, 021115, 0212284, 0212818 and OISE 0314581 and OISE 031458, STRI and CTFS. We thank Joe Wright for sharing data on dispersal modes for the BCI trees and two anonymous referees and the associate editor for helpful comments on previous versions of the paper.

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

Below we first establish the existence of a consistent sequence $\{(\hat{\beta}_n, \hat{\psi}_n)\}_{n\geq 1}$ 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

$$|W_n|^{-1/2} u_n(\beta^*, \psi^*)\tilde{\Sigma}_n^{-1/2} = |W_n|^{-1/2} u_n(\hat{\beta}_n, \hat{\psi}_n)\tilde{\Sigma}_n^{-1/2} + |W_n|^{1/2}\big((\hat{\beta}_n, \hat{\psi}_n) - (\beta^*, \psi^*)\big)\frac{J_n(\tilde{\beta}, \tilde{\psi})}{|W_n|}\tilde{\Sigma}_n^{-1/2} \quad (9)$$

where

$$J_n(\beta,\psi) = -\frac{\mathrm{d}}{\mathrm{d}(\beta,\psi)^{\mathsf{T}}} u_n(\beta,\psi) = \begin{bmatrix} \frac{\mathrm{d}}{\mathrm{d}\beta^{\mathsf{T}}} u_{n,1}(\beta) & \frac{\mathrm{d}}{\mathrm{d}\beta^{\mathsf{T}}} u_{n,2}(\beta,\psi) \\ 0 & \frac{\mathrm{d}}{\mathrm{d}\psi^{\mathsf{T}}} u_{n,2}(\beta,\psi) \end{bmatrix} = \begin{bmatrix} J_{n,11}(\beta,\psi) & J_{n,12}(\beta,\psi) \\ 0 & J_{n,22}(\beta,\psi) \end{bmatrix} \tag{10}$$

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 (13) in Appendix D. Similarly, letting $h(u) = \rho_{\beta^*}^{(1)}(u)/\rho_{\beta^*}(u)$,

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

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

$$P(\exists \tilde{\theta}_n : u_n(\tilde{\theta}_n) = 0 \text{ and } \|(\tilde{\theta}_n - \theta^*)V_n\| < d) > 1 - \epsilon \tag{11}$$

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 (11), 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

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

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

$$
\begin{aligned}
&P(\sup_{\|\phi\|=d} u_n(\theta^* + \phi V_n^{-1})V_n^{-1}\phi^\mathsf{T} \geq 0) \leq \\
&P(\sup_{\|\phi\|=d} u_n(\theta^*)V_n^{-1}\phi^\mathsf{T} \geq \inf_{\|\phi\|=d} \phi \int_0^1 V_n^{-1} J_n(\theta(t)) V_n^{-1} \mathrm{d}t \phi^\mathsf{T}) \leq \\
&P(\|u_n(\theta^*)V_n^{-1}\| \geq d \inf_{\|\phi\|=1} [\phi V_n^{-1} J_n(\theta^*) V_n^{-1} \phi^\mathsf{T}] - dp\gamma_{nd}]) \leq \\
&P(\|u_n(\theta^*)V_n^{-1}\| \geq dl_n/2) + P(p\gamma_{nd} > l_n/2).
\end{aligned}
$$

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 B1-B4.

LEMMA 1. *The variance*

$$
\mathbb{V}\mathrm{ar} \sum_{u,v \in X \cap W_n}^{\neq} \frac{1[\|u-v\| \leq t] f(u,v)}{|W_n| \rho_{\beta^*}(u) \rho_{\beta^*}(v)} \tag{12}
$$

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

PROOF. Let $\phi(u,v) = \frac{1[\|u-v\| \leq t] f(u,v)}{|W_n| \rho_{\beta^*}(u) \rho_{\beta^*}(v)}$. Then by the Campbell formulae, (12) is equal to

$$
\begin{aligned}
2\int_{W_n^2} \phi(u,v)^2 \rho^{(2)}_{\beta^*}(u,v) \mathrm{d}u\mathrm{d}v + 4\int_{W_n^3} \phi(u,v)\phi(v,w) \rho^{(3)}_{\beta^*}(u,v,w) \mathrm{d}u\mathrm{d}v\mathrm{d}w + \\
\int_{W_n^4} \phi(u,v)\phi(w,z)(\rho^{(4)}_{\beta^*}(u,v,w,z) - \rho^{(2)}_{\beta^*}(u,v)\rho^{(2)}_{\beta^*}(w,z)) \mathrm{d}u\mathrm{d}v\mathrm{d}w\mathrm{d}z.
\end{aligned} \tag{13}
$$

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

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

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

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

PROOF. By Lemma 1, $\hat{K}_{n,\beta^*}(t)$ tends to $K_{\psi^*}(t)$ in probability for each $t \geq 0$. 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*

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

PROOF. For a symmetric matrix A with eigenvalues λ_i and a vector ϕ it follows from the spectral decomposition that there exists another vector $\tilde{\phi}$ where $\|\phi\| = \|\tilde{\phi}\|$ and $\phi A \phi^\mathsf{T} = \sum_i \tilde{\phi}_i^2 \lambda_i$. This implies that that the eigenvalues 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 $1/\tilde{\lambda}_{\max}$, $\|\phi_1 \tilde{\Sigma}_{n,11}^{-1/2}\|^2 \geq 1/\tilde{\lambda}_{\max}$, and $l_{n,11} \geq \lambda_{n,11}/\tilde{\lambda}_{\max}$. Similarly, $l_{n,22} \geq \lambda_{22}/\tilde{\lambda}_{\max}$ where λ_{22} is the smallest eigenvalue of I_{22}.

LEMMA 4. *Assume N1-N2 and define $J_n(\beta,\psi)$ as in (10) in Appendix A.*

(a) For any $d > 0$,

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

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.

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

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

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

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

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

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

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

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

Regarding $|W_n|^{-1/2}\big(u_{n,1}(\beta^*,\psi^*), \tilde{u}_{n,2}(\beta^*,\psi^*)\big)$, 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

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

and define

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

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

$$Z_{ij} = X_{ij}x^\mathsf{T} + Y_{ij}y^\mathsf{T}, \quad \sigma_n^2 = |W_n|^{-1}\mathbb{V}\mathrm{ar}\sum_{(i,j)\in\mathbb{Z}^2} Z_{ij} = (x,y)\tilde{\Sigma}_n(x,y)^\mathsf{T} + o(1).$$

We show below that $(\sigma_n|W_n|)^{-1/2}\sum_{(i,j)\in\mathbb{Z}^2} Z_{ij}$ is asymptotically standard normal. This and N3 implies that $|W_n|^{-1/2}\sum_{(i,j)\in\mathbb{Z}^2} Z_{ij}$ is asymptotically $N(0,(x,y)\tilde{\Sigma}(x,y)^{\mathsf{T}})$. The asymptotic normality of $|W_n|^{-1/2}\big(u_{n,1}(\beta^*),\tilde{u}_{n,2}(\beta^*,\psi^*)\big)\tilde{\Sigma}_n^{-1/2}$ then follows by the Cramér-Wold device and N3.

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

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

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) $\liminf_{n\to\infty}\sigma_n^2>0$,
(b) $\sup_{ij}\mathbb{E}(|Z_{ij}|^{2+\delta})<\infty$,
(c) $\sum_{m\ge1} m\alpha_{2,\infty}(m;Z)^{\delta/(2+\delta)}<\infty$.

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 2 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

$$\sup_{\omega\in[-m/2,m/2]^2}\int_{\mathbb{R}^2\setminus[-m,m]^2} k(v-w)\mathrm{d}v \text{ is } O(m^{-d-2}). \tag{14}$$

Consider for a given $h>0$ 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,

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

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 (14) 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* **102**, 1377–1386.

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.

Kolmogorov, A. N. & Rozanov, Y. A. (1960). On strong mixing for stationary Gaussian processes. *Theory of Probability and its Applications* **5**, 294–208.

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

Møller, J. (2003). Shot noise Cox processes. *Advances in Applied Probability* **35**, 614–640.

Møller, J. & Torrisi, G. L. (2005). Generalized shot noise Cox processes. *Advances in Applied Probability* **37**, 48–74.

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.

Waagepetersen, R. & Guan, Y. (2008). Two-step estimation for inhomogeneous spatial point processes - a simulation study. Available at `www.math.aau.dk/~rw/publications.html`.

Waagepetersen, R. & Schweder, T. (2006). Likelihood-based inference for clustered line transect data. *Journal of Agricultural, Biological, and Environmental Statistics* **11**, 264–279.

Wright, S. J., Hernandéz, A. & Condit, R. (2007). The bushmeat harvest alters seedling banks by favoring lianas, large seeds, and seeds dispersed by bats, birds, and wind. *Biotropica* **39**, 363–371.