Statistics and Its Interface Volume 2 (2009) 91–99

A minimum contrast estimation procedure for estimating the second-order parameters of inhomogeneous spatial point processes

Yongtao Guan*

In this paper we propose a new model fitting procedure to estimate the second-order parameters for a class of inhomogeneous spatial point processes called second-order intensity reweighted stationary processes. The proposed approach is essentially a 'minimum contrast estimation' procedure but is based on the pair correlation function instead of the commonly used K-function of the process. We show through simulations that the new procedure gives more stable estimates than the approach based on the K-function. We apply the proposed method to a tropical forest data example to illustrate its practical use.

Keywords and phrases: Minimum contrast estimation, pair correlation function, second-order intensity reweighted stationary process.

1. INTRODUCTION

Since the early 1980, the *Center for Tropical Forest Science* has established 20 long-term, large-scale (50 ha) forest plots worldwide. Currently over 3 millions trees representing 8,200 species are being monitored. In all plots, all woody tree stems > 1 cm dbh have been mapped to the nearest 0.1 m and identified to species. In addition, abiotic covariates such as topographical and soil nutrient content have been collected. Given this wealth of new data, ecologists would like to gain insight on the biological processes that shape species spatial patterning and to ultimately understand how tropical forests maintain their extremely high biodiversity (Condit et al. 2000). To that end, a multitude of theories have been proposed, the majority of which can be broken down into two categories: niche assembly and dispersal assembly. Niche-assembly theories post that environmental heterogeneity and biological interactions are responsible for species coexistence and community structure (Hubbell and Foster 1986). In contrast, dispersal-assembly theories hypothesize that chance, history and dispersal explain species coexistence and community structure (Hubbell 2001). An ongoing debate exists in the ecological community concerning the relative importance of niche- versus dispersal-assembly theories of diversity (Bell et al. 2005). However, it is likely that in reality neither theory alone can fully explain the maintenance of high tropical tree species diversity (Gravel et al. 2006).

Recent advances in spatial point pattern analysis have provided promising new tools to more effectively separate the niche- and dispersal-assembly effects, so that their respective contributions in shaping the spatial patterning of trees can be assessed (Møller and Waagepetersen 2007). To apply these tools, we view the spatial locations of each tree species as a realization from a stochastic process called spatial point process. The intensity function of the process, which can be roughly understood as the mean or the first-order structure of the process, is typically assumed to be a parametric function of the abiotic covariates so as to incoporate potential niche-assembly effects (Waagepetersen 2007). In addition, a parametric model independent of these covariates is then used for the second-order structure (i.e., the dependence structure) of the process in order to gain insight on potential dispersal-assembly effects (Waagepetersen and Guan 2008). This kind of model fitting scheme is particularly useful for the so-called second-order intensity reweighted stationary processes (see Section 2 for its definition) and we will thus restrict our attention to such processes in what follows.

The main purpose of this paper is to develop a new model fitting procedure to estimate the second-order parameters in a second-order intensity reweighted stationary process, a subject that has not been well studied in literature. The estimation of such parameters is important for at least two reasons: 1) they carry extremely useful information regarding the degree of dispersal-assembly effects as discussed above, and 2) they are critical for inference on the intensity function of the process, i.e., inference on niche-assembly effects, because the variance of the estimated regression parameters in an intensity model typically depends on these second-order parameters. Our proposed method is based on the simple yet extremely useful idea of 'minimum contrast estimation'. However, unlike most existing related procedures, it is based on the pair correlation function but not the more commonly used K-function (Diggle 2003; Møller and Waagepetersen 2004). Alternatively, maximum likelihood estimation is possible for inhibitive types of spatial point patterns but is often too computationally intensive for it to be feasible for

*This research was supported by National Science Foundation grants DMS-0706806 and DMS-0845368.

clustered spatial point patterns (Møller and Waagepetersen 2004). For many tropical forestry data, the latter are much more common than the former.

Most minimum contrast estimation procedures involve some unknown tuning parameters to be determined and the value of these parameters may greatly affect the accuracy of the resulting estimators. This is no exception for our proposed method. We will develop data-driven procedures to select these parameters. In light of the simulation results in Section 4, we will also provide guidelines on selecting the tuning parameters when the estimation is based on the K-function. However, the choice in this case is somewhat more arbitrary than the proposed procedure. Moreover, our simulation results indicate that the proposed procedure is often more stable across the different choices of the tuning parameter involved.

The remainder of this article is organized as follows. We give some background on second-order intensity reweighted stationary processes in Section 2 and develop the proposed estimation procedure in Section 3. We then assess its numerical performance through simulations in Section 4 and apply it to a tropical forestry data example in Section 5. Some technical details are given in the Appendix.

2. BACKGROUND

2.1 Second-order intensity reweighted stationary processes

Consider a spatial point process N that is observed on a spatial domain of interest D. Let $d\mathbf{s}$ be a small region containing $\mathbf{s}$ and let $|d\mathbf{s}|$ denote the area of $d\mathbf{s}$. Write $N(d\mathbf{s})$ for the number of events of N falling in $d\mathbf{s}$. Following Diggle (2003), we define the first- and second-order intensity functions of the process

$$\lambda(\mathbf{s}) = \lim_{|d\mathbf{s}|\to 0} \frac{E[N(d\mathbf{s})]}{|d\mathbf{s}|},$$

$$\lambda_2(\mathbf{s}_1, \mathbf{s}_2) = \lim_{|d\mathbf{s}_1|,|d\mathbf{s}_2|\to 0} \frac{E[N(d\mathbf{s}_1)N(d\mathbf{s}_2)]}{|d\mathbf{s}_1||d\mathbf{s}_2|}.$$

Intuitively, $\lambda(\mathbf{s})|d\mathbf{s}|$ and $\lambda_2(\mathbf{s}_1,\mathbf{s}_2)|d\mathbf{s}_1||d\mathbf{s}_2|$ are the approximate probabilities for $d\mathbf{s}$ and for $d\mathbf{s}_1$ and $d\mathbf{s}_2$ to each contain an event of N. By convention, we refer to the first-order intensity function as the intensity function. We say that a spatial point process is second-order intensity reweighted stationary (SOIRS; Baddeley et al. 2000) if $\lambda(\mathbf{s}_1,\mathbf{s}_2) = \lambda(\mathbf{s}_1)\lambda(\mathbf{s}_2)g(\mathbf{s}_1-\mathbf{s}_2)$ for some function $g(\cdot)$, where $g(\cdot)$ is called the pair correlation function (PCF; Møller and Waagepetersen 2004). If the PCF is isotropic, then the reduced second moment measure, or the K-function can be expressed as $K(t) = 2\pi\int_0^t ug(u)du$.

The class of SOIRS process models contains many commonly used spatial point process models as special examples (Møller and Waagepetersen 2007). Among these, a very popular choice is the inhomogeneous Neyman-Scott process model (Waagepetersen 2007). To simulate realizations from such a process, a spatial Poisson process with some constant intensity ρ needs be first generated. Each event of the process is called a parent which in turn will generate a Poisson number of offspring with an expected value μ. Conditional on the location of each parent, the offspring are dispersed independently following some common probability density function. For any location $\mathbf{s}$, let $\mathbf{X}(\mathbf{s})$ be a $p\times 1$ vector of covariates recorded at this location. An offspring at $\mathbf{s}$ is retained with a probability $f[\mathbf{X}(\mathbf{s})^T\beta]/M$ for some function $f(\cdot)$, where $M = \max\{f[\mathbf{X}(\mathbf{s})^T\beta]\}$ and β is a $p\times 1$ vector of unknown parameters. The resulting offspring process then forms an inhomogeneous Neyman-Scott process.

A nice property about the inhomogeneous Neyman-Scott process is that its summary functions are often available in closed forms. Assume that the first elements of $\mathbf{X}(\mathbf{s})$ and β are equal to one and $\rho\mu/M$, respectively. Then the first- and second-order intensity functions of the process are

$$(1)\qquad \lambda(\mathbf{s};\beta) = f[\mathbf{X}(\mathbf{s})^T\beta],$$
$$\lambda(\mathbf{s}_1,\mathbf{s}_2;\beta,\theta) = \lambda(\mathbf{s}_1;\beta)\lambda(\mathbf{s}_2;\beta)g(\mathbf{s}_1-\mathbf{s}_2;\theta),$$

respectively, where $g(\mathbf{s}_1-\mathbf{s}_2;\theta)$ depends on some unknown parameter vector θ. For example, if the probability density function used to generate the offspring locations is a bivariate radially symmetric normal distribution (Diggle 2003), then

$$(2)\quad g(\mathbf{s}_1-\mathbf{s}_2;\theta) = 1+\exp[-||\mathbf{s}_1-\mathbf{s}_2||^2/(4\sigma^2)]/(4\pi\rho\sigma^2),$$

where σ^2 is the variance of the normal variables and $||\cdot||$ is the Euclidean norm. Note that here $\theta = (\rho,\sigma^2)$. In terms of the tropical forestry data examples discussed in Section 1, potential niche-assembly effects can be easily incorporated into the intensity function model. In particular, note from the definition of the model that the covariates control the survival rate of offspring (i.e., trees in this setting) at a given location. On the other hand, the dispersal parameter θ contains important information about the dispersal pattern of the species. For instance, Seidler and Plotkin (2006) illustrated that the parameter σ^2 in (2) can be used to distinguish among species with different modes of seed dispersal.

2.2 Estimation of β

The first-order parameter β is often estimated by an estimating equation approach based on the Poisson maximum likelihood (Schoenberg 2005). Assume that the intensity function admits the general form (1) and that $f(\cdot) = \exp(\cdot)$. Waagepetersen (2007) proposed to estimate β by solving

$$(3)\quad \mathbf{u}(\beta) = \sum_{\mathbf{s}\in N\cap D}\mathbf{X}(\mathbf{s})^T - \int_D \mathbf{X}(\mathbf{s})^T\exp[\mathbf{X}(\mathbf{s})^T\beta]d\mathbf{s} = \mathbf{0}.$$

Write $\hat{\beta}$ and β_0 for the resulting estimator obtained by solving (3) and the target parameter, respectively. Statistical properties of $\hat{\beta}$ can be found in Schoenberg (2005), Guan and Loh (2007) and Waagepetersen (2007). In particular, Guan and Loh (2007) showed that the variance of $\hat{\beta}$ can be written as

$$\Sigma = \text{cov}(\hat{\beta}) \approx \mathbf{A}^{-1}(\mathbf{A}+\mathbf{B})\mathbf{A}^{-1}, \tag{4}$$

where

$$\begin{aligned}\mathbf{A} &= \int_D \mathbf{X}(\mathbf{s})\mathbf{X}(\mathbf{s})^T\lambda(\mathbf{s};\beta_0)d\mathbf{s},\\ \mathbf{B} &= \int_D\int_D \mathbf{X}(\mathbf{u})\mathbf{X}(\mathbf{v})^T\lambda(\mathbf{u};\beta_0)\lambda(\mathbf{v};\beta_0)\\ &\qquad\qquad \times[g(\mathbf{u}-\mathbf{v};\theta)-1]d\mathbf{u}d\mathbf{v}.\end{aligned}$$

Note that although the calculation of $\hat{\beta}$ does not depend on θ, the variance of $\hat{\beta}$ involves θ.

2.3 Estimation of θ

Assume that the PCF is isotropic from now on. Based on an estimated intensity function, we can obtain the empirical K-function (Møller and Waagepetersen 2004)

$$\hat{K}(t) = \sum\sum_{\mathbf{s}_1,\mathbf{s}_2\in N\cap D}^{\neq} e(\mathbf{s}_1,\mathbf{s}_2)\frac{I(||\mathbf{s}_1-\mathbf{s}_2||\le t)}{\lambda(\mathbf{s}_1;\hat{\beta})\lambda(\mathbf{s}_2;\hat{\beta})},$$

where $\sum\sum^{\neq}$ signifies summation over distinct pairs, $I(\cdot)$ is an indicator function and $e(\mathbf{s}_1,\mathbf{s}_2)$ is an edge correction term. For examples of $e(\mathbf{s}_1,\mathbf{s}_2)$, see Diggle (2003) and Stoyan and Stoyan (1994). In the simulation, we set $e(\mathbf{s}_1,\mathbf{s}_2) = 1/|D\cap(D-\mathbf{s}_1+\mathbf{s}_2)|$, where $|D\cap(D-\mathbf{s}_1+\mathbf{s}_2)|$ is a common area between D and its copy shifted by $\mathbf{s}_1-\mathbf{s}_2$.

To estimate θ, the empirical K-function is often contrasted against the theoretical K-function through the discrepancy measure

$$U_K(\theta) = \int_0^r \{[\hat{K}(t)]^c - [K(t;\theta)]^c\}^2 dt. \tag{5}$$

An estimate for θ is then defined as the minimizer of $U_K(\theta)$. This is an example of the 'minimum contrast estimation' (MCE) (Møller and Waagepetersen 2004). The parameters c and r in (5) are two tuning parameters that need be preselected. Diggle (2003) suggested using $r \le .25$ for data on a unit square and using $c = .25$ or smaller for clustered point patterns, both in the homogeneous case. Note that the use of c is mainly to make the variance of $[\hat{K}(t)]^c$ more stable than that of $\hat{K}(t)$. Our experience is that different choices of r and c can lead to very different estimates, which often makes it difficult to interpret the obtained results. We will give some specific recommendations on the choice of these parameters based on our simulation results in Section 4.

3. THE PROPOSED PROCEDURE

3.1 The empirical PCF

Let $k(\cdot)$ denote a kernel function defined over $\mathbb{R}$ and let h be a bandwidth. The PCF can then be estimated by

$$\hat{g}(t;h) = \frac{1}{2\pi h}\sum\sum_{\mathbf{s}_1,\mathbf{s}_2\in N\cap D}^{\neq} e(\mathbf{s}_1,\mathbf{s}_2)\frac{k[(t-||\mathbf{s}_1-\mathbf{s}_2||)/h]}{\lambda(\mathbf{s}_1;\hat{\beta})\lambda(\mathbf{s}_2;\hat{\beta})||\mathbf{s}_1-\mathbf{s}_2||}. \tag{6}$$

In our simulation in Section 4, we set $k(\cdot)$ to be the Epanečnikov kernel, i.e., $k(x) = .75(1-x^2)$ when $-1\le x\le 1$ and 0 otherwise, and $e(\mathbf{s}_1,\mathbf{s}_2) = 1/|D\cap(D-\mathbf{s}_1+\mathbf{s}_2)|$ as in the K-function case. An analogous estimate for the PCF based on the conditional intensity can also be developed in the spatial-temporal point process setting, following the results in Adelfio and Schoenberg (2009). Here we focus on only the spatial point process case.

Write $\{D_n\}$ and $\{h_n\}$ for a sequence of domains and bandwidths satisfying condition (14) in the Appendix. Let $\hat{\beta}_n$ be $\hat{\beta}$ defined on D_n. Under some suitable regularity conditions, $\hat{\beta}_n$ is consistent for β and $\hat{\beta}-\beta_0$ is typically of order $|D_n|^{-1/2}$. We will therefore conveniently ignore the effect of $\hat{\beta}_n$ on $\hat{g}(t;h_n)$ in the subsequent development. If we further ignore the edge effect, then

$$E[\hat{g}(t;h_n)] \approx \int_{\mathbb{R}} k(u)g(t-h_nu)du.$$

Thus, $\hat{g}(t;h_n)$ is asymptotically unbiased for $g(t)$ given that $g(t)$ is continuous and $h_n \to 0$. Furthermore, we derive in the Appendix that

$$\begin{aligned}&Var[\hat{g}(t;h_n)]\\ &\quad\approx \frac{c_n g(t;\theta)}{|D_n|t}\int_{D_n}\int_0^{2\pi}\frac{1}{\lambda(\mathbf{s};\beta_0)\lambda[\mathbf{s}+\mathbf{u}(t,\psi);\beta_0]}d\psi d\mathbf{s},\end{aligned} \tag{7}$$

where $\{c_n\}$ is a sequence of real values independent of t and $\mathbf{u}(t,\psi) = [t\cos(\psi), t\sin(\psi)]$. Assume that t is relatively small compared to the size of D_n and that $\lambda(\mathbf{s};\beta_0)$ is sufficiently smooth such that $\lambda(\mathbf{s};\beta_0)\approx\lambda[\mathbf{s}+\mathbf{u}(t,\psi);\beta_0]$. Then the variance of $\hat{g}(t;h_n)$ may be simplified as

$$Var[\hat{g}(t;h_n)] \propto g(t;\theta)/t. \tag{8}$$

The above result will play a key role in developing our model fitting procedure below.

3.2 Bandwidth selection

To calculate (6), it is also important to select the bandwidth h, preferably by some data-driven methods. In the homogeneous case, Guan (2007) introduced a cross-validation approach based on the so-called composite likelihood. We will show that a similar approach can be developed in the inhomogeneous case. To begin, we first note that at $\theta = \theta_0$,

$$f(\mathbf{s}_1, \mathbf{s}_2; \theta) = \frac{\lambda(\mathbf{s}_1, \mathbf{s}_2; \beta_0, \theta)}{\int_D \int_D \lambda_2(\mathbf{u}, \mathbf{v}; \beta_0, \theta) d\mathbf{u} d\mathbf{v}}$$

is the probability density function for two arbitrary events in $D \cap N$ to be at $\mathbf{s}_1$ and $\mathbf{s}_2$, where $\lambda_2(\mathbf{s}_1, \mathbf{s}_2; \beta_0, \theta) = \lambda(\mathbf{s}_1; \beta_0)\lambda(\mathbf{s}_2; \beta_0) g(\mathbf{s}_1 - \mathbf{s}_2; \theta)$. Now sum up all the resulting pairwise log-likelihoods. We then obtain the following composite likelihood (Lindsay 1988) criterion

$$\begin{aligned} L(\theta) = \sum_{\mathbf{s}_1, \mathbf{s}_2 \in D \cap N}^{\neq}\sum & \Big\{ \log[\lambda_2(\mathbf{s}_1, \mathbf{s}_2; \beta_0, \theta)] \\ & - \log\left[\int_D \int_D \lambda_2(\mathbf{u}, \mathbf{v}; \beta_0, \theta) d\mathbf{u} d\mathbf{v}\right] \Big\}. \end{aligned}$$

In practice, β_0 is unknown but can be replaced by its consistent estimator $\hat{\beta}$. Estimation of θ can then be obtained by maximizing $L(\theta)$. For purpose of bandwidth selection, we can treat the bandwidth h as the unknown parameter that needs to be estimated. To be specific, we may choose h as the maximizer of

$$(9) \quad \begin{aligned} \sum_{\mathbf{s}_1, \mathbf{s}_2 \in D \cap N}^{\neq}\sum & \Big\{ \log[\tilde{g}(||\mathbf{s}_1 - \mathbf{s}_2||; h)] \\ & - \log\left[\int_D \int_D \lambda(\mathbf{u}; \hat{\beta})\lambda(\mathbf{v}; \hat{\beta})\hat{g}(||\mathbf{u} - \mathbf{v}||; h) d\mathbf{u} d\mathbf{v}\right] \Big\}. \end{aligned}$$

In the above, $\tilde{g}(||\mathbf{s}_1 - \mathbf{s}_2||; h)$ is the cross-validated version of $\hat{g}(||\mathbf{s}_1 - \mathbf{s}_2||; h)$ obtained by deleting the pair $(\mathbf{s}_1, \mathbf{s}_2)$. The removal of the observed pairs is important because otherwise (9) will always take its maximum at $h = 0$. This type of likelihood based cross-validation idea is not new and has been extensively used in density estimations (Silverman 1998). The main difference is that in the latter setting, the maximum likelihood is often used instead of the composite likelihood being used here. Furthermore, note that we delete a pair of events but not a single observation as in density estimations.

The calculation of the double integral in (9) can be quite computationally intensive. For a computationally faster alternative, we first note that a weighted version of $L(\theta)$ can be obtained as

$$\begin{aligned} \sum_{\mathbf{s}_1, \mathbf{s}_2 \in D \cap N}^{\neq}\sum W(\mathbf{s}_1, \mathbf{s}_2) & \Big\{ \log[\lambda_2(\mathbf{s}_1, \mathbf{s}_2; \beta_0, \theta)] \\ & - \log\left[\int_D \int_D W(\mathbf{u}, \mathbf{v})\lambda_2(\mathbf{u}, \mathbf{v}; \beta_0, \theta) d\mathbf{u} d\mathbf{v}\right] \Big\}, \end{aligned}$$

where $W(\mathbf{s}_1, \mathbf{s}_2)$ are some preselected weights. Inspired by this fact, a weighted version of (9) can be obtained as

$$(10) \quad \begin{aligned} \sum_{\mathbf{s}_1, \mathbf{s}_2 \in D \cap N}^{\neq}\sum W(\mathbf{s}_1, \mathbf{s}_2) & \Big\{ \log[\tilde{g}(||\mathbf{s}_1 - \mathbf{s}_2||; h)] \\ & - \log\left[\int_D \int_D W(\mathbf{u}, \mathbf{v})\lambda(\mathbf{u}; \hat{\beta})\lambda(\mathbf{v}; \hat{\beta})\hat{g}(||\mathbf{u} - \mathbf{v}||; h) d\mathbf{u} d\mathbf{v}\right] \Big\}. \end{aligned}$$

Let r_h be some preselected constant. Define

$$(11) \qquad W(\mathbf{s}_1, \mathbf{s}_2) = \frac{I(||\mathbf{s}_1 - \mathbf{s}_2|| \leq r_h)}{\lambda(\mathbf{s}_1; \hat{\beta})\lambda(\mathbf{s}_2; \hat{\beta})|D \cap D - \mathbf{s}_1 + \mathbf{s}_2|}.$$

The double integral in (10) then reduces to $2\pi \int_0^{r_h} t\hat{g}(t; h) dt$, which is much easier to calculate. We then obtain the composite likelihood cross-validation criterion

$$\begin{aligned} C(h) = \sum_{\mathbf{s}_1, \mathbf{s}_2 \in D \cap N}\sum W(\mathbf{s}_1, \mathbf{s}_2) & \Big\{ \log[\tilde{g}(||\mathbf{s}_1 - \mathbf{s}_2||; h)] \\ & - \log\left[\int_0^{r_h} t\hat{g}(t; h) dt\right] \Big\}, \end{aligned}$$

where $W(\mathbf{s}_1, \mathbf{s}_2)$ is as defined in (11). For the tuning parameter r_h, we generally set r_h to be around the dependence range. A rough estimate for the dependence range can be easily obtained by examining an empirical PCF plot based on a pilot bandwidth. Our experience is that the choice of r_h only has very limited effect on the choice of the bandwidth.

3.3 An MCE procedure for θ

The MCE procedure discussed in Section 2.3 can be extended to the case of using the PCF. Specifically, we may define the discrepancy measure

$$(12) \qquad U_g(\theta) = \int_0^r \left\{[\hat{g}(t; h)]^c - [g(t; \theta)]^c\right\}^2 dt,$$

where $\hat{g}(t; h)$ is the kernel estimator defined in (6). However, such a procedure is often difficult to apply because of the need to select the tuning parameters r and c. In general, the tuning parameter c is used to control the sampling fluctuation in the empirical summary function (i.e., $\hat{g}(t; h)$ in this case) involved in the MCE procedure. In light of (8), the same goal can be achieved by introducing a weight function $w(t) = t/\hat{g}(t; h)$ in the MCE discrepancy measure as follows:

$$(13) \qquad U(\theta) = \int_0^r w(t) \left[\hat{g}(t; h) - g(t; \theta)\right]^2 dt.$$

An advantage of using (13) is that now there is only one tuning parameter r to be selected. We will provide guidelines for the selection of r based on our simulation results in the next section.

4. A SIMULATION STUDY

To assess the performance of the proposed method, we apply it to realizations from inhomogeneous Neyman-Scott

Table 1. Mean squared errors of $\hat{\rho}$ and $\hat{\sigma}$ for MCE procedures based on (5) (denoted by MCEK) and (12) (denoted by MCEP), and the proposed procedure (MCEW). Each mean squared error is divided by the squared target parameter. $*$ denotes extremely unstable estimates

			$\rho = 25$						$\rho = 12.5$					
			MCEK			MCEP		MCEW	MCEK			MCEP		MCEW
	σ	r	.125	.25	.5	.25	.5		.125	.25	.5	.25	.5	
$\hat{\rho}$	.01	3σ	$*$	.087	.060	.057	.058	.054	.194	.108	.092	.090	.091	.088
		4σ	.096	.054	.050	.055	.055	.053	.107	.088	.085	.087	.087	.086
		5σ	.060	.051	.051	.057	.055	.055	.091	.086	.085	.088	.087	.086
		6σ	.055	.052	.052	.058	.056	.055	.089	.087	.086	.088	.087	.085
	.02	3σ	$*$	.097	.077	.077	.075	.074	.205	.114	.101	.096	.096	.095
		4σ	.113	.076	.078	.082	.078	.081	.118	.102	.099	.096	.094	.095
		5σ	.086	.083	.089	.085	.080	.084	.107	.104	.104	.098	.095	.094
		6σ	.084	.090	.097	.086	.081	.085	.109	.111	.113	.099	.095	.094
	.04	3σ	$*$	.250	.264	.187	.183	.201	.241	.204	.197	.148	.147	.166
		4σ	.254	.284	.334	.229	.223	.255	.207	.225	.235	.170	.165	.192
		5σ	.267	.371	.455	.246	.238	.271	.246	.278	.302	.178	.173	.201
		6σ	.331	.468	.612	.251	.243	.276	.296	.339	.390	.180	.174	.202
$\hat{\sigma}$	.01	3σ	$*$	.047	.018	.022	.025	.017	.327	.027	.013	.015	.017	.012
		4σ	.105	.017	.010	.015	.017	.013	.038	.012	.008	.011	.012	.010
		5σ	.046	.012	.009	.015	.016	.014	.021	.009	.008	.012	.011	.012
		6σ	.034	.011	.009	.016	.016	.015	.017	.009	.008	.013	.011	.012
	.02	3σ	$*$	.047	.020	.038	.040	.026	.392	.027	.015	.022	.024	.016
		4σ	.131	.020	.015	.031	.032	.025	.045	.014	.012	.018	.018	.016
		5σ	.059	.018	.017	.031	.031	.026	.026	.014	.014	.019	.018	.019
		6σ	.046	.020	.021	.031	.031	.027	.023	.016	.018	.019	.018	.019
	.04	3σ	$*$	.099	.049	.144	.145	.062	.388	.028	.021	.045	.047	.024
		4σ	.653	.048	.045	.128	.129	.054	.047	.024	.026	.037	.039	.026
		5σ	.125	.047	.058	.115	.117	.058	.032	.029	.037	.036	.037	.028
		6σ	.087	.053	.073	.113	.114	.064	.032	.037	.050	.036	.037	.029

processes simulated on a unit square. The offspring dispersion of the processes follows a bivariate radially symmetric normal distribution. We introduce inhomogeneity to the model by assigning $\lambda(\mathbf{s}) = \exp(\beta_0 + \beta_1 s_x)$, where $\mathbf{s} = (s_x, s_y)$ and $\beta_1 = 1$. Thus, the likelihood for an offspring to survive increases as s_x increases. The expected number of events per simulation is 100, and the expected number of parents $\rho = 12.5, 25$. The dispersal parameter $\sigma = .01, .02, .04$, representing relatively tight, moderate and loose clusters, respectively. For the tuning parameters, we set $c = .125, .25, .5, 1$ and $r = 3\sigma, 4\sigma, 5\sigma, 6\sigma$ for the MCE procedures based on (5) and (12). We set $r_h = 5\sigma$ for our proposed method and use the same r values as in the other two procedures. Note that 4σ is often regarded as the dependence range for this type of process (e.g., Diggle 2003). So we simply try to link the tuning parameter r to the dependence range of the process. A guideline on the choice of r in terms of the dependence range of the process is practically appealing since a rough estimate of the dependence range can often be obtained with relative ease, e.g., by examining an empirical PCF plot. For ease of presentation, we will use MCEK, MCEP, MCEW to denote the three estimation procedures, respectively, where MCEK and MCEP are the two MCE procedures based on (5) and (12), and MCEW is the proposed procedure.

Table 1 lists the mean squared errors for the three methods. For MCEK, the results for $c = 1$ are either similar to or slightly worse than those for $c = .5$. For MCEP, the results for $c = .125$ are similar to those for $c = .25$. We thus omit these results. From the results in Table 1 we see that the choice of tuning parameters greatly affects the performance of MCEK. Note that the combination of $r = 3\sigma$ and $c = .125$ often leads to unstable estimates. In contrast, MCEP and MCEW are relatively insensitive to the choice of the tuning parameter r. Moreover, MCEW appears to be generally better than MCEP in terms of estimating σ, which will be our main interest in the next section. For MCEW, $r = 3\sigma, 4\sigma$ yield overall the best results. From a practical point of view, this means that we should choose r to be around or slightly smaller than the dependence range of the process. Compared to the best results from MCEK, the results from MCEP are still quite competitive across all levels of r, especially when estimating ρ. It's worth noting that for MCEK, a small mean squared error for $\hat{\rho}$ is often accompanied by a large mean squared error for $\hat{\sigma}$ and vice versa. This is especially true when $\sigma = .04$.

We also comment on how to select r and c for MCEK given that this is a very popular approach in practice. As noted earlier, a combination of a small r value relative to the dependence range and a small c value should be avoided.

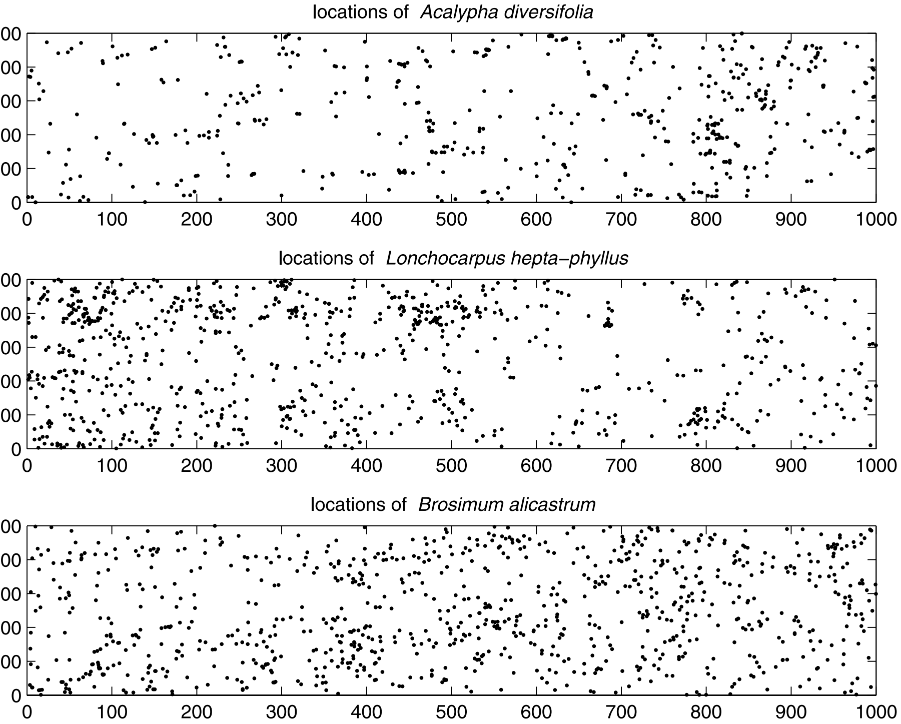

Figure 1. Locations of Acalypha diversifolia (526 trees), Lonchocarpus hepta-phyllus (837 trees) and Brosimum alicastrum (909 trees) in the BCI plot.

For a process with a short dependence range (i.e., with tight clusters as is the case for $\sigma = .01$ in the simulation), $c = .5$ consistently yields the best results across all r values. Note that this contradicts the general perception that $c = .25$ or less gives better results for clustered point patterns. To select r in this case, we recommend using $r = 4\sigma, 5\sigma$, i.e., around or slightly larger than the dependence range. For a process with a medium dependence range that is comparable to the case of $\sigma = .02$, $c = .25$ and $.5$ both work well. Although $c = .125$ sometime yields better results when estimating ρ, this is often achieved at the price of a highly variable estimate for σ. To select r in this case, we recommend using $r = 4\sigma$, i.e., around the dependence range of the process. For a process with a relatively long dependence range (e.g., $\sigma = .04$ in our simulation), $c = .125$ generally performs better than $c = .25$ and $.5$ when estimating ρ except for $r = 3\sigma$. However, it often yields an unstable estimate for σ. We thus do not recommend using $c = .125$ in this case. To select r, we recommend using $r = 4\sigma$ for $c = .25$ or $r = 3\sigma$ for $c = .5$. For a more general process, this means that r should be around or slightly smaller than the dependence range for $c = .25$ and $.5$, respectively.

The performance of all procedures is affected by the model parameters ρ, μ and σ. In particular, σ can be estimated more accurately when ρ is smaller. Note that a smaller ρ means a larger cluster size so that more information in each cluster can be used to estimate σ. For ρ, more accurate estimates can be obtained if the cluster is tighter (i.e., if σ is smaller). This is because tighter clusters can be more easily distinguished from each other. We have also conducted simulations for a bigger sample size with $\lambda = 200$ and our main conclusions regarding the performance of these procedures and our general recommendations still remain valid. Due to space constraint, we thus omit the detailed presentation of these results.

5. AN APPLICATION

We apply the proposed method to analyze spatial distribution of three tree species in a 1000 meters by 500 meters plot in the Barro Colorado Island (BCI). The BCI plot is one of the 20 permanent plots established by the *Center for Tropical Forest Science*, see Condit et al. (1996), Condit (1998), and Hubbell & Foster (1983) for more information. The three tree species being considered are *Acalypha diversifolia* (526 trees), *Lonchocarpus hepta-phyllus* (837 trees) and *Brosimum alicastrum* (909 trees). Figure 1 plots the spatial locations of these trees.

The seed dispersal modes for the three species vary greatly, with exploding capsules for *Acalypha*, wind for *Lonchocarpus*, and birds and mammals for *Brosimum.* It is hypothesized that the different seed dispersal modes will lead to different spatial patterns of tree locations, with tight clusters for exploding capsules, loose clusters for bird and mammal dispersal, and somewhere in between for wind dispersal

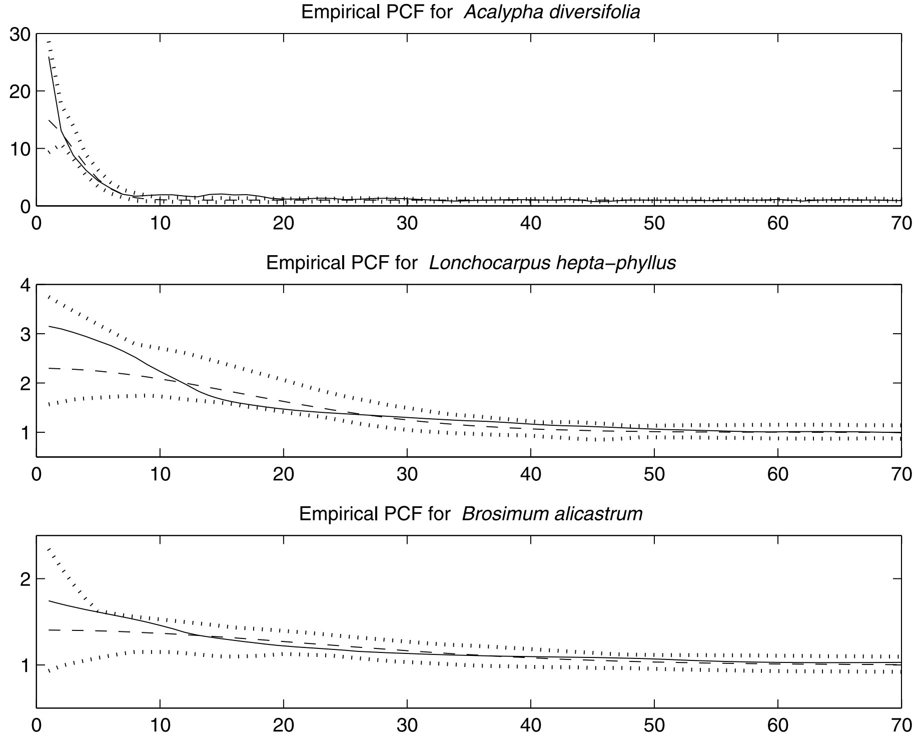

Figure 2. Plots of the parametric (dashed lines) and nonparametric estimates (solid lines) of the pair correlation functions and the associated confidence envelopes (dotted lines) obtained by simulations for Acalypha diversifolia, Lonchocarpus hepta-phyllus and Brosimum alicastrum in the BCI plot.

(Seidler and Plotkin 2006). We will assess the dispersal patterns of these three tree species by fitting inhomogeneous Neyman-Scott process models to these data. To be specific, we use a bivariate radially symmetric normal distribution for the offspring dispersal and define the intensity function model as

$$\lambda(\mathbf{s}) = \exp\left[\mathbf{X}(\mathbf{s})^T\beta\right],$$

where the covariates $\mathbf{X}(\mathbf{s})$ include elevation, slope, soil contents of potassium, phosphorous, mineralized nitrogen and and soil pH level. The same set of covariates were used in Waagepetersen and Guan (2008). Note that the fitted intensity function models will allow a direct assessment of potential niche-assembly effects in terms of these covariates.

Figure 2 shows the empirical PCF plots for the three species. Clearly *Acalypha* is much more clustered than the other two species and also has a much shorter dependence range. It also appears that *Lonchocarpus* is slightly more clustered and has a slightly longer dependence range than *Brosimum*. Table 2 shows the estimates and their associated standard errors and 95% confidence intervals for the dispersal parameter σ from both MCEK and MCEW. The standard errors are obtained from 100 Monte Carlo simulations based on the estimated regression parameters $\hat{\beta}$ and the estimated second-order parameters from each procedure. Following Waagepetersen and Guan (2008), we assume that the distribution of $\hat{\sigma}$ is asymptotically normal, based on which we construct the confidence intervals. We do not consider MCEP here due to its poor performance in estimating σ, as can be seen from the simulation results in Section 4. For MCEK, we use two different choices for the tuning parameters for each species. The first is an arbitrary choice with $c = .25$ and $r = 100$ meters for all species as in Waagepetersen and Guan (2008), whereas the second is based on our general recommendation given in Section 4 with $r = 12, 40, 55$ meters in conjunction with $c = .5$ bing used for *Acalypha*, *Lonchocarpus* and *Brosimum*, respectively. Note that the selected r values are around the dependence range for *Acalypha* but slightly smaller than the dependence range for both *Lonchocarpus* and *Brosimum* (see Figure 2). For ease of presentation, we refer to the resulting estimates as MCEK$_1$ and MCEK$_2$, respectively. Note that all three procedures give similar estimated values for σ for *Lonchocarpus* and *Brosimum*. However, MCEK$_1$ yields a much larger estimate than MCEK$_2$ and MCEW for *Acalypha*. In all cases, MCEK$_1$ also has much larger standard deviations than MCEK$_2$ and MCEW. This provides further supports for the need to carefully select these tuning parameters. As discussed in the previous paragraph, we expect the smallest σ for *Acalypha*, the largest for *Brosimum*, and somewhere in between for *Lonchocarpus*. The estimated σ values are indeed in this order. Based on the obtained 95%

Table 2. Estimation results for the dispersal parameter σ for the tropical forestry data. MCEK$_1$ is the MCE approach using the K-function with $c = .25$ and $r = 100$ meters; MCEK$_2$ is the MCE approach using the K-function with $c = .5$ and $r = 12, 40, 55$ meters for Acalypha, Lonchocarpus and Brosimum, respectively; MCEW is the proposed estimation procedure with the same r values as MCEK$_2$. EST, STD and CI stand for estimate, standard deviation and confidence interval, respectively.

	MCEK$_1$			MCEK$_2$			MCEW		
Species	EST	STD	95% CI	EST	STD	95% CI	EST	STD	95% CI
Acalypha	3.35	.18	(2.99,3.70)	1.93	.10	(1.73,2.13)	2.11	.11	(1.89,2.34)
Lonchocarpus	10.30	1.52	(7.33,13.28)	9.46	1.10	(7.30,11.62)	11.70	1.12	(9.49,13.90)
Brosimum	13.72	3.04	(7.67,19.67)	13.76	2.44	(8.99,18.54)	15.82	2.44	(11.02,20.61)

Table 3. Estimation results for the first-order parameters for the tropical forestry data. EST, STD and CI stand for estimate, standard deviation and confidence interval, respectively.

	Acalypha			*Lonchocarpus*			*Brosimum*		
Covariates	EST	STD	95% CI	EST	STD	95% CI	EST	STD	95% CI
Elevation	.207	.084	**(.042,.371)**	.193	.100	(-.003,.388)	.015	.076	(-.135,.165)
Slope	.129	.070	(-.008,.266)	.072	.087	(-.099,.243)	-.022	.065	(-.149,.106)
Potassium	.215	.109	**(.001,.430)**	.113	.124	(-.130,.356)	-.017	.098	(-.208,.175)
Phosphorous	.002	.078	(-.150,.154)	-.272	.098	**(-.464,-.080)**	.068	.069	(-.067,.202)
Nitrogen	.054	.091	(-.125,.233)	-.315	.105	**(-.520,-.110)**	.062	.081	(-.097,.221)
pH	.041	.089	(-.133,.216)	-.236	.104	**(-.440,-.033)**	.035	.079	(-.121,.190)

confidence intervals, we can further conclude that *Acalypha* has a much smaller σ value than the other two species. However, we do not detect a significant difference between the σ values for *Lonchocarpus* and *Brosimum*.

Figure 2 also plots the resulting parametric estimates of the PCF based on the estimated parameters from MCEW and the associated simulation envelopes from 99 simulations. The fits appear to be reasonable for all three species, except that the empirical PCF for *Acalypha* slightly exceeds the upper simulation envelope from the fitted model. Table 3 shows the estimates and their associated standard errors and 95% confidence intervals for the regression parameters β. The standard errors are estimated by a plug-in estimator for (4) based on the estimated regression parameters $\hat{\beta}$ and the estimated second-order parameters from MCEW. Following Guan and Loh (2007), we assume that the distribution of $\hat{\beta}$ is asymptotically normal, based on which we construct the confidence intervals. For *Acalypha*, both elevation and potassium are significant. The positive signs suggest that this particular species prefers both higher elevation and higher potassium. For *Lonchocarpus*, phosphorous, nitrogen and soil pH level are all significant. The negative signs for phosphorous and nitrogen suggest that this is a 'a frugal species adapted to soils with low nutrition contents' (Waagepetersen and Guan 2008). The negative sign for pH indicates that *Lonchocarpus* prefers more acidic soil conditions. For *Brosimum*, it is quite interesting to see that neither coefficient is significant. Thus, it appears that dispersal effects alone determine the spatial distribution of *Brosimum*. This is not the case for either *Acalypha* nor *Lonchocarpus*.

APPENDIX: DERIVATION OF (6)

Assume that $\lambda(\mathbf{s})$ is bounded below from zero. Consider a sequence of regions D_n and bandwidths h_n. Let ∂D_n denote the boundary of D_n and $|\partial D_n|$ denote the length of ∂D_n. We assume the following condition on D_n and h_n:

$$(14)\qquad |D_n| = \mathrm{O}(n^2),\ |\partial D_n| = \mathrm{O}(n),\ \text{and}\ h_n = \mathrm{O}(n^{-\beta})\ \text{for some}\ \beta \in (0,2).$$

For ease of presentation, let $e(\mathbf{s}_1, \mathbf{s}_2) = 1/|D_n|$, that is, we do not consider any edge correction term. Assume also that $\hat{\beta}_n$ can be replaced safely with β_0 without altering the asymptotic results. This is generally true given that $\hat{\beta}_n$ is consistent for β_0.

Let $\lambda_k(\mathbf{s}_1, \dots, \mathbf{s}_k)$ denote the kth-order intensity function of the process, defined analogously as the first- and second-order intensity functions. Assume that $\lambda_k(\mathbf{s}_1, \dots, \mathbf{s}_k)$ has the general form of $\lambda(\mathbf{s}_1) \dots \lambda(\mathbf{s}_k) g_k(\mathbf{s}_2 - \mathbf{s}_1, \dots, \mathbf{s}_k - \mathbf{s}_1)$ for $k = 2, 3, 4$. This condition is not restrictive and holds for commonly used spatial point process models such as the inhomogeneous Neyman-Scott process and the log-Gaussian Cox process (Møller et al. 1998) models. Given these conditions, then

$$\begin{aligned}
&4\pi^2(|D_n|)^2\,Var[\hat{g}_n(t;h)]\\
&= 2\iint \frac{\{k[(t - ||\mathbf{s}_1 - \mathbf{s}_2||)/h_n]\}^2 g(||\mathbf{s}_1 - \mathbf{s}_2||)}{\lambda(\mathbf{s}_1)\lambda(\mathbf{s}_2)||\mathbf{s}_1 - \mathbf{s}_2||^2(h_n)^2}\, d\mathbf{s}_1 d\mathbf{s}_2\\
&+ 4\iiint \frac{k[(t - ||\mathbf{s}_1 - \mathbf{s}_2||)/h_n]k[(t - ||\mathbf{s}_1 - \mathbf{s}_3||)/h_n]g_3(\mathbf{s}_2 - \mathbf{s}_1, \mathbf{s}_3 - \mathbf{s}_1)}{\lambda(\mathbf{s}_1)||\mathbf{s}_1 - \mathbf{s}_2||||\mathbf{s}_1 - \mathbf{s}_3||(h_n)^2}\\
&\qquad \times d\mathbf{s}_1 d\mathbf{s}_2 d\mathbf{s}_3
\end{aligned}$$

$$+ \iiiint \frac{k[(t-||\mathbf{s}_1-\mathbf{s}_2||)/h_n]k[(t-||\mathbf{s}_3-\mathbf{s}_4||)/h_n]}{||\mathbf{s}_1-\mathbf{s}_2||||\mathbf{s}_3-\mathbf{s}_4||(h_n)^2}$$

$$\times [g_4(\mathbf{s}_2-\mathbf{s}_1,\mathbf{s}_3-\mathbf{s}_1,\mathbf{s}_4-\mathbf{s}_1) - g(||\mathbf{s}_1-\mathbf{s}_2||)g(||\mathbf{s}_3-\mathbf{s}_4||)]d\mathbf{s}_1 d\mathbf{s}_2 d\mathbf{s}_3 d\mathbf{s}_4,$$

where all the integrations are over D_n. Given condition (14) and the so-called Brillinger mixing (Heinrich 1988), it follows from lengthy yet elementary algebra that the first term on the right-hand side of the equality dominates over the other two terms. Moreover, the first term is approximately equal to

$$\frac{g(t;\theta)}{h_n t}\int_{D_n}\int_0^{2\pi}\frac{1}{\lambda(\mathbf{s})\lambda[\mathbf{s}+\mathbf{u}(t,\psi)]}d\psi d\mathbf{s}\int [k(u)]^2 du.$$

This thus completes the proof.

Received 6 December 2008

REFERENCES

Adelfio, G. and Schoenberg, F. P. (2009), "Point Process Diagnostics Based on Weighted Second-Order Statistics and Their Asymptotic Properties", *Annals of the Institute of Statistical Mathematics*, in press.

Baddeley, A. J., Møller, J. and Waagepetersen, R. (2000), "Non- and Semi-Parametric Estimation of Interaction in Inhomogeneous Point Patterns", *Statistica Neerlandica*, 54, 329–350.

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

Condit, R., Ashton, P. S., Baker, P., Bunyavejchewin, S., Gunatilleke, S., Gunatilleke, N., Hubbell, S. P., Foster, R. B., Itoh, A., Lafrankie, J. V., Lee, H. S., Losos, E., Manokaran, N., Sukumar, R. and Yamakura, T. (2000), "Spatial Patterns in the Distribution of Tropical Tree Species", *Science*, 288, 1414–1418.

Condit, R., Hubbell, S. P. and 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*, New York: Oxford University Press Inc.

Gravel, D., Canham, C. D., Beaudet, M. and Messier, C. (2006), "Reconciling Niche and Neutrality: the Continuum Hypothesis", *Ecology Letters*, 9, 399–409.

Guan, Y. (2007), "A Composite Likelihood Cross-Validation Approach in Selecting Bandwidth for the Estimation of the Pair Correlation Function", *Scandinavian Journal of Statistics*, 34, 336–346.

Guan, Y. and Loh, J. M. (2007), "A Thinned Block Bootstrap Variance Estimation Procedure for Inhomogeneous Spatial Point Patterns", *Journal of the American Statistical Association*, 102, 1377–1386.

Heinrich, L. (1988), "Asymptotic Gaussianity of Some Estimators for Reduced Factorial Moment Measures and Product Densities of Stationary Poisson Cluster Processes", *Statistics*, 19, 87–106.

Hubbell, S. P. (2001). *The unified neutral theory of biodiversity and biogeography*, Princeton, NJ: Princeton University Press.

Hubbell, S. P. and Foster, R. B. (1983), "Diversity of Canopy Trees in a Neotropical Forest and Implications for Conservation", *Tropical Rain Forest: Ecology and Management*, Sutton, S. L., Whitmore, T. C. and Chadwick, A. C. (eds.), Oxford: Blackwell Scientific Publications, 25–41.

Hubbell, S. P. and Foster, R. B. (1986), "Biology, Chance and History and the Structure of Tropical Rain Forest Tree Communities". In: J. Diamond and T. J. Case (eds.), *Community Ecology*, 314–329, New York: Harper and Row.

Lindsay, B. G. (1988), "Composite Likelihood Methods", *Contemporary Mathematics*, 80, 221–239.

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

Møller, J. and Waagepetersen, R. P. (2004), *Statistical Inference and Simulation for Spatial Point Processes*, New York: Chapman & Hall.

Møller, J. and Waagepetersen, R. P. (2007), "Modern Statistics for Spatial Point Processes", *Scandinavian Journal of Statistics*, 34, 643–684.

Schoenberg, F. P. (2005), "Consistent Parametric Estimation of the Intensity of a Spatial-temporal Point Process", *Journal of Statistical Planning and Inference*, 128(1), 79–93.

Seidler, T. G. and Plotkin, J. B. (2006), "Seed Dispersal and Spatial Pattern in Tropical Trees", *PLoS Biology*, 4, 2132–2137.

Silverman, B. W. (1998), *Density Estimation for Statistics and Data Analysis*, New York: Chapman & Hall.

Stoyan, D. and Stoyan, H. (1994), *Fractals, Random Shapes and Point Fields*, New York: Wiley.

Stoyan, D. and Stoyan, H. (1998), "Non-Homogeneous Gibbs Process Models for Forestry – A Case Study", *Biometrical Journal*, 40, 521–531.

Waagepetersen, R. P. (2007), "An Estimating Function Approach to Inference for Inhomogeneous Neyman-Scott Processes", *Biometrics*, 62, 252–258.

Waagepetersen, R. P. and Guan, Y. (2008), "Two-Step Estimation for Inhomogeneous Spatial Point Processes", *Journal of the Royal Statistical Society, Ser. B*, to appear.

Yongtao Guan
Division of Biostatistics
Yale School of Public Health
Yale University
New Haven, CT 06520-8034
E-mail address: yongtao.guan@yale.edu