

# Variance Estimation for Statistics Computed from Inhomogeneous Spatial Point Processes

Yongtao Guan

April 14, 2007

## Abstract

This paper introduces a new approach to estimate the variance of statistics that are computed from an inhomogeneous spatial point process. The proposed approach is based on the assumption that the observed point process can be thinned to be a second-order stationary point process, where the thinning probability depends only on the first-order intensity function of the (unthinned) original process. The resulting variance estimator is proved to be asymptotically consistent for the target parameter under some very mild conditions. The use of the proposed approach is demonstrated in two important applications of modeling inhomogeneous spatial point processes: 1) residual diagnostics of a fitted model, and 2) inference on the unknown regression coefficients. A simulation study and an application to a real data example are used to demonstrate the efficacy of the proposed approach.

---

Yongtao Guan is Assistant Professor, Division of Biostatistics, Yale School of Public Health, Yale University, New Haven, CT 06520.8034, e-mail [yongtao.guan@yale.edu](mailto:yongtao.guan@yale.edu). The author would like to thank Rasmus Waagepetersen for helpful discussions on the *Beilschmiedia pendula Lauraceae* data, and the Editor, an Associate Editor and two referees for helpful comments that have improved the paper.

KEY WORDS: Inhomogeneous Spatial Point Process, Second-order Reweighted Stationarity, Thinning.

## 1 Introduction

A key interest in modeling inhomogeneous spatial point patterns is to model the first-order intensity function (FOIF; see Section 2 for its definition) of the underlying process that generated the point patterns. This is generally done by assuming a parametric structure for the FOIF in terms of some known covariates. For example, in a forestry study, the observed point patterns can be the locations of trees, whereas the covariates may be measurements about the local environmental factors such as soil and watering conditions. The main question is if and how these factors affect the distribution of trees. In what follows, let  $N$  denote the spatial point process of interest that is observed over a domain of interest, say  $D$ , and let  $\lambda(\mathbf{s}; \theta)$  be the FOIF of  $N$  at  $\mathbf{s}$ , depending on a vector of unknown regression coefficients  $\theta$ . The main interest often is to estimate and make inference on  $\theta$ . To obtain an estimate for  $\theta$ , the following likelihood criterion:

$$U(\theta) = \sum_{\mathbf{x} \in D \cap N} \log[\lambda(\mathbf{x}; \theta)] - \int_D \lambda(\mathbf{s}; \theta) d\mathbf{s}, \quad (1)$$

can be used. The maximizer of (1) is an estimator of  $\theta$ . It is referred to as the Poisson maximum likelihood estimator (PMLE; Schoenberg, 2004) since (1) coincides with the true maximum likelihood when the underlying process  $N$  is an inhomogeneous Poisson process. If  $N$  is not Poisson, the PMLE is still consistent for  $\theta$  under some mild conditions (Schoenberg, 2004) and is asymptotically normal for a class of Neyman-Scott processes (Waagepetersen, 2006).

Once an estimate for  $\theta$  is obtained, it is then of interest to assess the goodness-of-fit of the fitted model. Furthermore, inference on  $\theta$  is often desired. For these analyses, knowledge on second-order structure of the process, specifically the second-order intensity function (SOIF, see Section 2 for its

definition), is generally needed. In practice, the SOIF is often retracted from a fitted parametric model for the whole process. For example, to model the locations of tropical rainforest trees, an inhomogeneous Neyman-Scott process (INSP) model was used in Waagepetersen (2006). The model attributed the clustering of trees to a one-round seed dispersal. Waagepetersen (2006) carefully pointed out that the model was only a “crude model” since the clustering in reality might be due to seed dispersal over several generations. Despite the question on the validity of the model, it should be noted that inference on  $\theta$  depends only on how well the resulting SOIF from the fitted model approximates the true SOIF of the process. Thus even if the fitted parametric model for the whole process is incorrect, it does not necessarily invalidate the correctness of inference. Nevertheless it would be nice if this parametric assumption can be relaxed.

The goal of this paper is to introduce a novel method for the estimation of the variance of a general statistic taking the following form:

$$T(B) = \sum_{\mathbf{x} \in B \cap N} f(\mathbf{x}), \quad (2)$$

where  $B$  is a given Borel set and  $f$  is a known function. Statistics in the form of (2) and their variance estimates play an important role in both diagnostics and inferences in the modeling of inhomogeneous spatial point patterns, as explained further in Section 3. Since it may be difficult to judge if an estimated SOIF is close enough to the true SOIF, it is desirable to assume more flexible conditions when obtaining the variance estimates for  $T(B)$ . In this paper, the spatial point process  $N$  is assumed to be only second-order reweighted-stationary (SORWS; see Section 2 for its definition). Under this assumption, a thinned second-order stationary spatial point process can then be formed, where the probability to keep an event from the original process in the thinned process is proportional to the inverse of the FOIF at the location of the event. The main idea is that for each thinned process, a new statistic  $T^*(B)$  can be defined, where the first two moments of  $T^*(B)$  are

translation invariant, i.e.  $T^*(B)$  and  $T^*(B + \mathbf{u})$  have the same mean and variance. As a result, the variance of  $T^*(B)$  can then be estimated by using the sample variance of its “shifted” copies, i.e.  $T^*(B + \mathbf{u})$  such that  $B + \mathbf{u} \in D$ . It is shown in Section 2 that the variance of  $T^*(B)$  has a simple relationship with that of  $T(B)$ . By solving this relationship, an estimate for the variance of  $T(B)$  can then be easily obtained. Note that no parametric structure on the high-order structure of the process is required, which is a main advantage of the proposed method over other currently available methods.

The rest of the article is organized as follows. Section 2 introduces the proposed method. Section 3 considers its application in model diagnostics and inference on the regression parameters. Section 4 contains results from a simulation study that was used to assess the finite sample performance of the proposed method. An application to a real data example is given in Section 5.

## 2 The Proposed Method

### 2.1 A thinning algorithm

Consider a two-dimensional spatial point process,  $N$ . For any Borel set  $B \subset \mathbb{R}^2$ , let  $|B|$  denote the area of  $B$ , and  $N(B)$  denote the number of events of  $N$  in  $B$ . For any  $\mathbf{s} \in \mathbb{R}^2$ , let  $ds$  be an infinitesimal region containing  $\mathbf{s}$ . Following Diggle (2003), the  $k$ th-order intensity function of  $N$  is defined as:

$$\lambda_k(\mathbf{s}_1, \dots, \mathbf{s}_k) = \lim_{|ds_i| \rightarrow 0} \left\{ \frac{\mathbf{E}[N(ds_1) \cdots N(ds_k)]}{|ds_1| \cdots |ds_k|} \right\}, \quad i = 1, \dots, k.$$

To avoid confusion,  $ds_i$  here are assumed to be compact and convex sets that converge to zero at the same rate. For a more rigorous definition of the intensity functions, see Daley and Vere-Jones (1988) under the name of factorial moment density. Note that in the above definition, the dependence of the intensity functions on the unknown parameter vector  $\theta$  is omitted. This omission is used merely

for the convenience of notation.

A point process is said to be SORWS in this paper if  $\lambda_2(\mathbf{s}_1, \mathbf{s}_2) = \lambda(\mathbf{s}_1)\lambda(\mathbf{s}_2)g(\mathbf{s}_1 - \mathbf{s}_2)$  for some function  $g$ , where  $g$  is often referred to as the pair correlation function (PCF; e.g. Stoyan and Stoyan, 1994). Note that if the PCF exists for a point process, then the definition of the second-order reweighted stationarity being considered here is same as the original definition given by Baddeley et al (2000). Examples of SORWS processes include, but are not limited to, the INSP, the log Gaussian Cox process (Møller et al, 1998) with a stationary correlation function for the Gaussian random field that generates the intensities, and any point process obtained by thinning a SORWS point process (e.g. Baddeley et al, 2000).

For a SORWS point process, some simple derivations by using the Campbell's theorem (e.g. Stoyan and Stoyan 1994) yield the following expression for the variance of the statistic  $T(B)$  defined in (2):

$$Var [T(B)] = \int [f(\mathbf{s})]^2 \lambda(\mathbf{s}) d\mathbf{s} + \iint f(\mathbf{s}_1) f(\mathbf{s}_2) \lambda(\mathbf{s}_1) \lambda(\mathbf{s}_2) [g(\mathbf{s}_1 - \mathbf{s}_2) - 1] d\mathbf{s}_1 d\mathbf{s}_2, \quad (3)$$

where all integrals in the above are over  $B$ . Now consider the following thinned process on  $D$ :

$$N_i^* = \left\{ \mathbf{x} : \mathbf{x} \in N \cap D, P(\mathbf{x} \text{ is not thinned}) = \frac{\min_{\mathbf{s} \in D} \lambda(\mathbf{s})}{\lambda(\mathbf{x})} \right\}. \quad (4)$$

The subindex  $i$  in the above is due to the fact that multiple thinned realizations can be created from a single realization of  $N$ . Note that all  $N_i^*$  are second-order stationary on  $D$ . In particular, the first- and second-order intensity functions are given as:

$$\lambda^* = \min_{\mathbf{s} \in D} \lambda(\mathbf{s}) \quad \text{and} \quad \lambda_2^*(\mathbf{s}_1, \mathbf{s}_2) = (\lambda^*)^2 g(\mathbf{s}_1 - \mathbf{s}_2), \quad \text{where } \mathbf{s}_1, \mathbf{s}_2 \in D. \quad (5)$$

Let  $B(\mathbf{u})$  be the Borel set that is obtained by shifting  $B$  by  $\mathbf{u}$ . Consider the following statistic defined on  $B(\mathbf{u})$ :

$$T_i^*(\mathbf{u}) = \sum_{\mathbf{x} \in B(\mathbf{u}) \cap N_i^*} f(\mathbf{x} - \mathbf{u}) \lambda(\mathbf{x} - \mathbf{u}). \quad (6)$$

In view of the results in (5), the variance of  $T_i^*(\mathbf{u})$  is given by:

$$\begin{aligned} \text{Var} [T_i^*(\mathbf{u})] &= \lambda^* \int [f(\mathbf{s})\lambda(\mathbf{s})]^2 ds \\ &+ (\lambda^*)^2 \iint f(\mathbf{s}_1)f(\mathbf{s}_2)\lambda(\mathbf{s}_1)\lambda(\mathbf{s}_2) [g(\mathbf{s}_1 - \mathbf{s}_2) - 1] d\mathbf{s}_1 d\mathbf{s}_2, \end{aligned} \quad (7)$$

where all integrals in (7) are over  $B$ . Interestingly, (7) does not depend on the shifting constant  $\mathbf{u}$ .

Furthermore, it can be seen by comparing (3) and (7) that

$$\text{Var} [T(B)] = \frac{1}{(\lambda^*)^2} \text{Var} [T_i^*(\mathbf{u})] - \frac{1}{\lambda^*} \int [f(\mathbf{s})\lambda(\mathbf{s})]^2 ds + \int [f(\mathbf{s})]^2 \lambda(\mathbf{s}) ds. \quad (8)$$

The last two terms in (8) can be calculated or numerically approximated by using the FOIF. Thus only  $\text{Var} [T_i^*(\mathbf{u})]$  needs be estimated in order to estimate  $\text{Var}[T(B)]$ . Since  $\text{Var} [T_i^*(\mathbf{u})]$  is independent of  $\mathbf{u}$ , it is natural to consider the following method-of-moment estimator:

$$\hat{V}^* = [b|D^*|]^{-1} \sum_{i=1}^b \int_{D^*} \{T_i^*(\mathbf{u}) - E[T^*(\mathbf{u})]\}^2 d\mathbf{u}, \quad (9)$$

where  $b$  is a predefined positive integer,  $D^* = \{\mathbf{u} : \mathbf{u} \in D \text{ and } B(\mathbf{u}) \subset D\}$ , and  $E[T^*(\mathbf{u})] = \lambda^* \int_B f(\mathbf{s})\lambda(\mathbf{s}) ds$ . This leads to the following estimator for  $\text{Var}[T(B)]$ :

$$\hat{V} = \frac{1}{(\lambda^*)^2} \hat{V}^* - \frac{1}{\lambda^*} \int [f(\mathbf{s})\lambda(\mathbf{s})]^2 ds + \int [f(\mathbf{s})]^2 \lambda(\mathbf{s}) ds. \quad (10)$$

In practice, one has to decide the number of thinned realizations to be considered (i.e.  $b$ ). Our experience is that a larger  $b$  typically reduces the variability of  $\hat{V}^*$  (and thus that of  $\hat{V}$ ). Therefore, a large  $b$  should be preferred over a small  $b$ . However, the value of  $\hat{V}^*$  often becomes nearly constant if  $b$  is large enough. One way to select  $b$  is thus to assess the trace plot of  $\hat{V}^*$  against  $b$  and set  $b$  to the value after which a nearly constant  $\hat{V}^*$  can be obtained.

## 2.2 Large sample properties

The large sample properties of the variance estimator in (10) will be studied under an increasing-domain set up. Specifically, let  $D_n$  be a ‘‘convex averaging sequence’’ of subsets of  $\mathbb{R}^2$ , i.e.  $D_n$  is

bounded and convex,  $D_n \subset D_{n+1}$  for any  $n$ , and the radius of the largest circle contained in  $D_n$  goes to infinity as  $n \rightarrow \infty$ . Furthermore, let  $B_n$  be a sequence of Borel sets and let  $T(B_n)$  be  $T(B)$  calculated on  $B_n$ . Assume that  $|B_n|/|D_n| \rightarrow 0$  as  $n \rightarrow \infty$ . In another word,  $B_n$  may become increasingly large but should always be relatively small compared to  $D_n$  as  $n$  increases. A special example of  $B_n$  is  $B_n = B$  for a given Borel set  $B$  for all  $n$ .

Let  $\hat{V}_n$  be  $\hat{V}$  calculated on  $D_n$ . To show the consistency of  $\hat{V}_n$ , the dependence strength of  $N$  needs be quantified. This can be done by using the cumulant function of  $N$ . Let  $\text{Cum}(Y_1, \dots, Y_k)$  be the coefficient of  $i^k t_1 \dots t_k$  in the Taylor series expansion of  $\log[\mathbb{E}\{\exp(i \sum_{j=1}^k Y_j t_j)\}]$  about the origin for some given random variables  $Y_1, \dots, Y_k$  (e.g. Brillinger, 1975). The  $k$ th-order cumulant function of  $N$  is defined as:

$$Q_k(\mathbf{s}_1, \dots, \mathbf{s}_k) = \lim_{|d\mathbf{s}_i| \rightarrow 0} \left\{ \frac{\text{Cum}[N(d\mathbf{s}_1), \dots, N(d\mathbf{s}_k)]}{|d\mathbf{s}_1| \dots |d\mathbf{s}_k|} \right\}, \quad i = 1, \dots, k.$$

The cumulant functions can be linked to the intensity functions defined in Section 2.1 by using the well-known relationship between cumulants and moments (e.g. McCullagh, 1987). Specifically, the  $k$ -th order cumulant function can be expressed as a function of the intensity functions up to the  $k$ -th order and the reverse is also true. For example,  $Q_1(\mathbf{s}) = \lambda(\mathbf{s})$  and  $Q_2(\mathbf{s}_1, \mathbf{s}_2) = \lambda_2(\mathbf{s}_1, \mathbf{s}_2) - \lambda(\mathbf{s}_1)\lambda(\mathbf{s}_2)$ . See Daley and Vere-Jones (1988, p.147) for expressions in the more general case.

Typically, near-zero values of  $Q_k(\mathbf{s}_1, \dots, \mathbf{s}_k)$  for  $k \geq 2$  correspond to a weak dependence. In the special case that  $N$  is an inhomogeneous Poisson process, i.e. events occurring at disjoint sets are completely independent, then  $Q_k(\mathbf{s}_1, \dots, \mathbf{s}_k) = 0$  if any  $\mathbf{s}_i - \mathbf{s}_j \neq \mathbf{0}$ ,  $i, j = 1, \dots, k$  for all  $k \geq 2$ . In this paper, the following regularity conditions on  $N$  are assumed:

$$C_1 < \lambda(\cdot) < C_2, \text{ where } 0 < C_1 < C_2 < \infty, \quad (11)$$

$$\sup_{\mathbf{s}_1 \in \mathbb{R}^2} \int_{\mathbb{R}^2} \dots \int_{\mathbb{R}^2} |Q_k(\mathbf{s}_1, \dots, \mathbf{s}_k)| d\mathbf{s}_2 \dots d\mathbf{s}_k < C \text{ for } k = 2, 3, 4, \text{ where } 0 < C < \infty. \quad (12)$$

Condition (11) requires that events from  $N$  are distributed “evenly” across the whole region, i.e. not unreasonably too dense or too sparse in some regions. Condition (12) requires the dependence in  $N$  is weak. This is a very mild condition and is satisfied by many processes such as the INSP, the log Gaussian Cox process, and any point process obtained by thinning a stationary point process that satisfies this condition (e.g. Heinrich, 1985).

**Theorem 1.** *If (11) and (12) hold, and  $f(\cdot)$  is bounded, then  $\hat{V}_n/\text{Var}[T(B)] \xrightarrow{P} 1$ .*

In practice, the true FOIF  $\lambda(\cdot)$  is typically unknown and thus must be estimated. Let  $\hat{\theta}_n$  be an estimate defined on  $D_n$  for the true but unknown parameter vector  $\theta_0$ , and  $\lambda(\cdot; \hat{\theta}_n)$  be an estimate of  $\lambda(\cdot)$ . Let  $\lambda_n^*(\hat{\theta}_n) = \min_{\mathbf{s} \in D_n} \lambda(\mathbf{s}; \hat{\theta}_n)$ . Then the analogue for the thinned process defined in (4) becomes

$$N_i^*(\hat{\theta}_n) = \left\{ \mathbf{x} : \mathbf{x} \in N \cap D_n, P(\mathbf{x} \text{ is not thinned}) = \frac{\lambda_n^*(\hat{\theta}_n)}{\lambda(\mathbf{x}; \hat{\theta}_n)} \right\}.$$

Based on  $N_i^*(\hat{\theta}_n)$ , the counterpart of (9) is

$$\hat{V}_n^*(\hat{\theta}_n) = [b|D_n^*|]^{-1} \sum_{i=1}^b \int_{D_n^*} [T_i^*(\mathbf{u}; \hat{\theta}_n) - \mu_n(\hat{\theta}_n)]^2 d\mathbf{u},$$

where

$$T_i^*(\mathbf{u}; \hat{\theta}_n) = \sum_{\mathbf{x} \in B_n(\mathbf{u}) \cap N_i^*(\hat{\theta}_n)} f(\mathbf{x} - \mathbf{u}) \lambda(\mathbf{x} - \mathbf{u}; \hat{\theta}_n),$$

$$\mu_n(\hat{\theta}_n) = \lambda_n^*(\hat{\theta}_n) \int_{B_n} f(\mathbf{s}) \lambda(\mathbf{s}; \hat{\theta}_n) d\mathbf{s}, \quad \text{or} \quad [b|D_n^*|]^{-1} \sum_{i=1}^b \int_{D_n^*} T_i^*(\mathbf{u}; \hat{\theta}_n) d\mathbf{u}. \quad (13)$$

This leads further to the following estimator for  $\text{Var}[T(B)]$ :

$$\hat{V}_n(\hat{\theta}_n) = \frac{1}{[\lambda_n^*(\hat{\theta}_n)]^2} \hat{V}_n^*(\hat{\theta}_n) - \frac{1}{\lambda_n^*(\hat{\theta}_n)} \int [f(\mathbf{s}) \lambda(\mathbf{s}; \hat{\theta}_n)]^2 d\mathbf{s} + \int [f(\mathbf{s})]^2 \lambda(\mathbf{s}; \hat{\theta}_n) d\mathbf{s}. \quad (14)$$

A question of interest then is if  $\hat{V}_n(\hat{\theta}_n)$  is a consistent estimator for  $\text{Var}[T(B)]$ . To answer this question, knowledge on how fast  $\hat{\theta}_n$  converges to  $\theta_0$  is needed. Due to a recent result in Guan and



Loh (2006), it appears to be reasonable to assume that

$$\sqrt{|D_n|}(\hat{\theta}_n - \theta_0) = O_p(1). \quad (15)$$

The following corollary gives the asymptotic consistency of  $\hat{V}_n(\hat{\theta}_n)$  for  $Var[T(B)]$ .

**Corollary 1.** *If (11), (12), (15) hold,  $f(\cdot)$  is bounded and  $\lambda(\cdot, \theta)$  has a bounded first-order derivative with respect to  $\theta$ , then  $\hat{V}_n(\hat{\theta}_n)/Var[T(B)] \xrightarrow{p} 1$ .*

The next section considers the applications of the proposed method in two important occasions: residual diagnostics of a fitted model and inference on the unknown regression parameters.

### 3 Applications

#### 3.1 Residual diagnostics

After a model is fitted, the next step of the analysis is often to assess the goodness-of-fit of the fitted model. Baddeley et al (2005) proposed a novel residual analysis approach that can be used for this purpose. Specifically, let  $\lambda_p(\mathbf{s}, N)$  denote the Papangelou conditional intensity (PCI; Papangelou, 1974). The residual is then defined as:

$$R_p(B) = N(B) - \int_B \lambda_p(\mathbf{s}, N) ds. \quad (16)$$

For many Gibbs point processes, the PCI is available in closed forms. Thus  $R_p(B)$  defined in (16) can be calculated without difficulty. For many clustered point processes such as the INSP and the Cox process, however, the PCI is often difficult to obtain. Instead, Waagepetersen (2005) suggested to define the residuals in terms of the FOIF. One possibility is to consider

$$R(B) = N(B) - \int_B \lambda(\mathbf{s}) ds. \quad (17)$$

More generally, the following “scaled” residual

$$R(B, f) = \sum_{\mathbf{x} \in B \cap N} f(\mathbf{x}) - \int_B f(\mathbf{s})\lambda(\mathbf{s})d\mathbf{s} \quad (18)$$

may be used. If  $f(\mathbf{x}) = 1$ , then (18) corresponds to  $R(B)$  in (17); if  $f(\mathbf{x}) = 1/\lambda(\mathbf{x})$ , then (18) is the inverse residual.

A useful diagnostic tool based on (18) is the “smoothed residual field” (SRF; Baddeley et al, 2005). Let  $K(\cdot)$  be a kernel function defined on a compact set  $B$  (e.g. a circle or square), where  $B$  is centered at the origin. For an arbitrary location  $\mathbf{u}$ , the “smoothed residual” (SR) at  $\mathbf{u}$  is defined as:

$$R(\mathbf{u}, B, f; \hat{\theta}) = \sum_{\mathbf{x} \in B(\mathbf{u}) \cap N} K(\mathbf{x} - \mathbf{u})f(\mathbf{x}) - \int_{B(\mathbf{u})} K(\mathbf{s} - \mathbf{u})f(\mathbf{s})\lambda(\mathbf{s}; \hat{\theta})d\mathbf{s}. \quad (19)$$

If the fitted model is close to the true model, then the SR  $R(\mathbf{u}, B, f; \hat{\theta})$  given in (19) should be close to zero for all  $\mathbf{u}$ . A systematic deviation from zero may be either due to a lack-of-fit or the existence of outliers (i.e. regions with extremely dense or sparse events). To determine if a deviation is significant, the variance of the SR has to be taken into consideration. To obtain an estimate for the variance, note that  $B$  is typically small relative to  $D$ . Furthermore by condition (15), the second term on the right hand side of the equality sign in (19) converges in probability to the expected value of the first term if the fitted model is correct. Thus for data with a large sample size, it is often sufficient to consider the variance of only the first term, which is in the form of (2). The proposed variance estimation procedure can then be directly applied.

The residual defined in (18) can also be used to help assess if a new covariate should be included in the model. For this purpose, let  $Z(\mathbf{s})$  be a real-valued spatial covariate defined at each location  $\mathbf{s} \in D$ , and  $K(\cdot, h)$  be a kernel function defined on  $|\cdot| \leq h$  for some preselected bandwidth  $h$ . For an arbitrary location  $\mathbf{u} \in D$ , define  $B(\mathbf{u}, h) = \{\mathbf{s} : |Z(\mathbf{s}) - Z(\mathbf{u})| \leq h\}$ . The SR defined by (19)

can be extended for covariate  $Z(\cdot)$  as:

$$R(\mathbf{u}, h, f; \hat{\theta}) = \sum_{\mathbf{x} \in B(\mathbf{u}, h) \cap N} K[Z(\mathbf{x}) - Z(\mathbf{u}), h] f(\mathbf{x}) - \int_{B(\mathbf{u}, h)} K[Z(\mathbf{s}) - Z(\mathbf{u})] f(\mathbf{s}) \lambda(\mathbf{s}; \hat{\theta}) d\mathbf{s}. \quad (20)$$

Based on the SRs for  $Z(\cdot)$ , a lurking variable plot can be created by plotting  $R(\mathbf{u}, h, f; \hat{\theta})$  against  $Z(\mathbf{u})$  for all  $\mathbf{u} \in D$  (Baddeley et al, 2005). To understand the motivation of using (20), consider the simple case with  $f(\cdot) = 1$ . The two terms on the right side of the equation are simply a nonparametric kernel smoothing estimator and a kernel-smoothed version of the parametric estimator of the FOIF, respectively. These two estimates of intensity should be approximately equal if the fitted model is correct. A systematic deviation from zero indicates the necessity to include  $Z(\cdot)$  in the model or to revise its functional form (if it has already been included). To account for the variability of  $R(\mathbf{s}, h, f; \hat{\theta})$ , the proposed variance estimation procedure can again be applied.

### 3.2 Inference on regression coefficients

An important task in modeling inhomogeneous spatial point patterns is to make inference on the regression coefficients. For this purpose, it is convenient to assume that the statistic,  $T_n = \sqrt{|D_n|}(\hat{\theta}_n - \theta_0)/\hat{v}_n$ , has a limiting distribution, say  $J(\cdot)$ , where  $(\hat{v}_n)^2$  is a consistent estimator for  $|D_n|Var(\hat{\theta}_n)$ , e.g. the block bootstrap estimator proposed recently by Guan and Loh (2006). In practice, the form of  $J(\cdot)$  may be unknown and thus must be estimated. Politis and Romano (1994) and Sherman and Carlstein (1996) independently proposed a subsampling procedure which can be used to estimate  $J(\cdot)$ . The estimated distribution, say  $\hat{J}_n$ , can in turn be used for inference on the unknown parameter(s).

A similar approach can be obtained for spatial point processes. To see this, let  $D_{l(n)}$  be a predefined shape, where  $|D_{l(n)}| \rightarrow \infty$  and  $|D_{l(n)}|/|D_n| \rightarrow 0$  as  $n \rightarrow \infty$ , i.e.  $D_{l(n)}$  are increasingly large but still relatively small compared to  $D_n$ . Let  $D_{l(n)}(\mathbf{u})$  be  $D_{l(n)}$  shifted by  $\mathbf{u}$

and define  $D_n^s = \{\mathbf{u} : \mathbf{u} \in D_n, D_{l(n)}(\mathbf{u}) \subset D_n\}$ . For each  $D_{l(n)}(\mathbf{u})$ , let  $\hat{\theta}_{l(n)}(\mathbf{u})$  be  $\hat{\theta}_n$  obtained on  $D_{l(n)}(\mathbf{u})$ , and  $[\hat{v}_{l(n)}(\mathbf{u})]^2$  be an estimate for  $[v_{l(n)}(\mathbf{u})]^2 = |D_{l(n)}| \text{Var}[\hat{\theta}_{l(n)}(\mathbf{u})]$ . Define  $T_{l(n)}(\mathbf{u}) = \sqrt{|D_{l(n)}|} [\hat{\theta}_{l(n)}(\mathbf{u}) - \theta_0] / \hat{v}_{l(n)}(\mathbf{u})$ . Let  $I(\cdot)$  be an indicator function. Then  $J(\cdot)$  can be estimated by:

$$\hat{J}_n(y) = \frac{\int_{D_n^s} I[T_{l(n)}(\mathbf{u}) \leq y] d\mathbf{u}}{|D_n^s|}.$$

A key task in order to obtain  $\hat{J}_n(y)$  is to calculate  $\hat{v}_{l(n)}(\mathbf{u})$ . For this purpose, note the recent result in Guan and Loh (2006) that under some regularity conditions,

$$\sqrt{|D_{l(n)}|} [\hat{\theta}_{l(n)}(\mathbf{u}) - \theta_0] \sim \left[ \frac{1}{|D_{l(n)}|} \int_{D_{l(n)}(\mathbf{u})} \frac{[\lambda^{(1)}(\mathbf{s}; \theta_0)]^2}{\lambda(\mathbf{s}; \theta_0)} d\mathbf{s} \right]^{-1} \frac{U_{l(n)}^{(1)}(\theta_0, \mathbf{u})}{\sqrt{|D_{l(n)}|}},$$

where  $U_{l(n)}^{(1)}(\theta, \mathbf{u})$  is the first derivative of  $U(\theta)$  in (1) that is obtained on  $D_{l(n)}(\mathbf{u})$ , and the notation  $a \sim b$  means that  $a$  and  $b$  have the same limiting distribution. To calculate  $\hat{v}_{l(n)}(\mathbf{u})$ , it is necessary to estimate the variance of  $U_{l(n)}^{(1)}(\theta_0, \mathbf{u})$ . Note that

$$U_{l(n)}^{(1)}(\theta_0, \mathbf{u}) = \sum_{\mathbf{x} \in D_{l(n)}(\mathbf{u}) \cap N} \frac{\lambda^{(1)}(\mathbf{x}; \theta_0)}{\lambda(\mathbf{x}; \theta_0)} - \int_{D_{l(n)}(\mathbf{u})} \lambda^{(1)}(\mathbf{s}; \theta_0) d\mathbf{s}. \quad (21)$$

Clearly, the first term on the right hand side of the equality sign in (21) is in the form of  $T(B)$  in (2). Thus the proposed variance estimation procedure can again be applied.

## 4 A Simulation Study

To evaluate the performance of the proposed variance estimation procedure, a simulation study was conducted. The model considered here was the INSP model, where the FOIF of the process, at any  $\mathbf{s} = (x, y)$ , was given by  $\lambda(\mathbf{s}) = \alpha \exp(\beta x)$  for some predefined  $\alpha$  and  $\beta$ . Note that  $\beta$  controls the level of inhomogeneity of the model. In particular, a  $\beta$  with a larger absolute value means that the resulting process is more inhomogeneous. In the simulation,  $\beta$  was assigned to be 1 and 2, which both led to an increased FOIF as  $x$  increased. The different  $\beta$  values allowed us to assess

the performance of the proposed procedure under different level of inhomogeneity. Note that the model being considered here is rather simple with just one smooth covariate. As pointed out by one referee, the results here may provide a too optimistic picture of the performance of the method. The accuracy of the variance estimate is likely to be worse in case of a more complicated form of inhomogeneity. A unit square was used as the domain of interest throughout the simulation.

To simulate a realization from the specified INSP model, a homogeneous Poisson process was first simulated as the parent process. For each parent, a Poisson number of offspring was then generated and the position of each offspring relative to its parent was defined as a radially symmetric Gaussian random variable (see, e.g. Diggle, 2003). Each offspring was thinned as in Waagepetersen (2006) where the probability to retain an offspring was equal to the intensity at the offspring location divided by the maximum intensity in the unit square, i.e.  $\alpha \exp(\beta)$ . In what follows, let  $\rho$ ,  $\mu$  and  $\sigma$  denote the intensity of the parent process, the expected number of offspring per parent, and the standard deviation of the Gaussian dispersion variable, respectively. In the simulation,  $\rho$  was 100 and 400,  $\mu$  was 3.15 for  $\beta = 1$  but 4.66 for  $\beta = 2$ , and  $\sigma$  was .02 and .04. Note that a smaller  $\sigma$  led to a stronger clustering. The resulting mean number of events per realization (denoted by  $\lambda$ ) was 200 and 800 for  $\rho = 100, 400$ , respectively. The different  $\lambda$  values allowed us to study the effect of sample size on the performance of the proposed procedure.

For each realization of the process, the unknown parameters  $\alpha$  and  $\beta$  were first estimated by maximizing (1) and using  $x$  as the covariate. Based on the estimated  $\alpha$  and  $\beta$ , the SR defined in (19) was then obtained over four  $0.1 \times 0.1$  squares that were centered at  $x = 0.2, 0.4, 0.6, 0.8$  and  $y = 0.5$ . The variance estimator in (14) was calculated to estimate the variance of each SR.  $\mu_n(\hat{\theta}_n)$  in (14) was obtained by using the second formula in (13). Results based on the first formula were almost identical to those based on the second and thus are omitted from presentation.

Another method to estimate the variance of the statistics was to first fit a parametric model for

the PCF and then plug the estimated PCF into (3). For this two classes of parametric models were considered: a Poisson cluster process model with a Gaussian dispersion and a log Gaussian Cox process model with an exponential correlation function for the Gaussian random field generating the intensity of the process. The first model was the correct model whereas the second was an incorrect but nevertheless comparable model. To estimate the unknown parameters in each of the models, the “minimum contrast estimation” procedure (e.g. Møller and Waagepetersen, 2003) was used. Specifically, the empirical  $K$ -function was contrasted with the theoretical  $K$ -function for lags less than or equal to  $4\sigma$ . For the ease of presentation, the three variance estimators used in the simulation are referred to as estimator 1, 2 and 3, respectively.

Table 1 reports the estimated biases and standard deviations (STDs) for estimators 1-3 for  $\beta = 1$ . For bias, the sample variance of the SR from 10,000 simulations was treated as the true variance. For estimator 1, both the bias and the STD typically became smaller when the sample size increased, i.e. the accuracy of the estimator improved. This provided support for the large sample theory in Section 2. The bias was quite small for both estimators 1 and 2 compared to the STD, but could get very large for estimator 3. One exception was when  $\lambda = 200$  and  $\sigma = .04$ . In this case the bias for estimator 3 was also very small. The small bias for estimators 1 and 2 was expected. Figure 1 plots the true  $K$ -function for the simulated processes and the  $K$ -function using the average of the estimated parameters for estimator 3. Clearly the true  $K$ -function was approximated well when  $\lambda = 200$  and  $\sigma = .04$  but not in the other three cases. This indicated that the magnitude of the bias depended on how well the true second-order structure of the process was approximated by the fitted model. For STD, estimator 1 generally had a slightly larger STD than estimator 2 when  $\sigma = .02$  and/or  $\lambda = 200$ , but a smaller STD than estimator 3 when  $\sigma = .02$ . When  $\lambda = 200$  and  $\sigma = .04$ , the STD of estimator 3 was even smaller than that of estimator 2. When  $\lambda = 800$  and  $\sigma = .04$ , the STDs of all three estimators were comparable.

Tables 2 reports the mean squared errors (MSEs) for estimators 1-3 for both  $\beta = 1$  and  $\beta = 2$ . Generally, the MSE became larger when  $\beta$  increased for all three estimators except when  $x = .02$ . This increased error was due to the increased variability in the estimation of the FOIF parameters. For  $\sigma = .02$ , estimator 2 always had the smallest MSE, whereas estimator 3 had the largest. For  $\lambda = 200$  and  $\sigma = .04$ , estimator 3 often had the smallest MSE. This was expected in view of the observations on the biases and the STDs in this case. For  $\lambda = 800$  and  $\sigma = .04$ , estimator 1 surprisingly even outperformed estimator 2. This was likely due to the large variability in the estimation of the PCF in this case. In summary, the proposed variance estimation procedure performed only slightly worse than the plug-in method when the correct PCF model was used but often much better than the plug-in method when an incorrect but comparable model was used. In some situations, the proposed procedure performed better even when the correct PCF model was used for the plug-in method. In practice, it is almost impossible to determine what the correct PCF model is. Thus the proposed procedure appears attractive in real applications, especially when there is doubt over whether a fitted PCF model describes the second-order structure of the process adequately.

## 5 An Application to Real Data

The data considered in this section consist of 3605 tree locations of the species *Beilschmiedia pendula Lauraceae* (abbreviated by *Beilschmiedia* hereinafter) in a  $1000 \times 500$  meters plot in Barro Colorado Island (see Figure 2). *Beilschmiedia* is a perennial tree species that belongs to the *Lauraceae* Family. These locations were extracted from a much larger data set obtained in a 1995 census, which contains measurements for over 300 species existing in the same plot. Waagepetersen (2006) fit an INSP model to the data, using elevation and gradient as covariates. Specifically, the

FOIF of the model was defined as:

$$\lambda(\mathbf{s}) = \exp[\beta_0 + \beta_1 E(\mathbf{s}) + \beta_2 G(\mathbf{s})], \quad (22)$$

where  $E(\mathbf{s})$  and  $G(\mathbf{s})$  were the (estimated) elevation and gradient at location  $\mathbf{s}$ , respectively. In his subsequent analysis, Waagepetersen concluded a significant effect for gradient but not for elevation. Furthermore, the empirical  $K$ -function that he obtained agreed well with the fitted theoretical  $K$ -function, which indicated an acceptable fit.

The empirical  $K$ -function also suggested that the *Beilschmiedia* trees were clustered. From a biological point of view, the clustering might be caused by seed dispersal, which mostly likely had occurred over multiple generations, or by correlation among other important environmental factors (i.e. not elevation and gradient), which presumably also affected the distribution of *Beilschmiedia*. The INSP model is used to model clustering due to the former. However, it is only a “crude model” even in this sense since it considers seed dispersal only in one generation. As a result, the INSP model may not be appropriate for the data. Although this does not necessarily invalidate Waagepetersen’s conclusion, it would be of interest to further investigate these results under more flexible conditions.

The data were analyzed here for three purposes: 1) to assess the goodness-of-fit of the fitted model in Waagepetersen (2006), 2) to investigate any spatial trend in terms of the spatial coordinates and 3) to make inference on the regression parameters  $\beta_1$  and  $\beta_2$ . The first two purposes were not considered by using the fitted FOIF directly in Waagepetersen (2006). For all analyses, (22) was assumed to be the correct FOIF model but no specific assumption was made on the high-order structure of the process other than it was SORWS. The residual diagnostics discussed in Section 3.1 was then used for the first two purposes, whereas the inferential procedure using subsampling discussed in Section 3.2 was used for the third purpose.



Figure 3 plots the standardized SRF that was based on the (unstandardized) SR in (19). The set  $B$  in (19) was chosen to be a  $80 \times 80$  meters square region, and the kernel function  $K(\cdot)$  was a uniform kernel over  $B$ . The variance for each SR was estimated by the proposed variance estimation procedure, which was subsequently used to produce the standardized SRF. From Figure 3, the overall fit appeared to be good, except for one region that is centered around  $(300, 450)$  meters. The standardized SRs in this region were much larger than as expected by the model. This suggested that *Beilschmiedia* were more abundant in this region than could be attributed to elevation and gradient alone. It appears sensible to suspect that other factors that were not included in the model also affected the locations of *Beilschmiedia*. To identify these factors, it is necessary to collect more covariates within this region. Furthermore, there appeared to be a possible problem of overestimation of the FOIF for a large region in the center of the plot. This should also be investigated in future studies.

Figure 4 plots the standardized SRFs by using each of the two coordinates as a covariate, where the (unstandardized) SR used in each case to obtain the standardized SR was defined in (20). The bandwidth  $h$  in (20) was set equal to 10 meters and the kernel function  $K(\cdot)$  was a uniform kernel over  $B(\cdot, h)$ . Again, the proposed variance procedure was applied to estimate the variance of each SR so as to produce the standardized SRF. Figure 4 did not show a strong spatial trend in either of the two coordinates. However, the plot against the  $x$  coordinate indicated one outlier around  $x = 300$ . This provided further evidence for the conclusion that the outlying region suggested by Figure 3 should be further investigated.

To make inference on  $\beta_1$  and  $\beta_2$ , their 95% confidence intervals (CIs) were calculated by using the inferential procedure in Section 3.2. The subblock size was  $100 \times 50$  meters. Two other subblock sizes ( $120 \times 60$  and  $140 \times 70$  meters) were also considered and they both yielded similar results. The initial variance estimate was equal to 0.017 for  $\hat{\beta}_1$  and 2.119 for  $\hat{\beta}_2$  by using the resampling

method in Guan and Loh (2006). The resulting CI was  $(-.015, .042)$  for  $\beta_1$  and  $(1.161, 10.225)$  for  $\beta_2$ . These results confirmed the conclusion of Waagepetersen (2006) that  $\beta_2$  was significant but  $\beta_1$  was not. Note that this was achieved under a much weaker assumption on the underlying process.

## Appendix A: Proof of Theorem 1

For the ease of presentation, define

$$F(\mathbf{x}) = f(\mathbf{x})\lambda(\mathbf{x}), \quad T_{n,i}^*(\mathbf{u}) = \sum_{\mathbf{x} \in B_n(\mathbf{u}) \cap N_i^*} F(\mathbf{x} - \mathbf{u}), \quad \mu_n = \int_{B_n} F(\mathbf{s}) d\mathbf{s}, \quad \text{and} \quad V_n^* = \text{Var} [T_{n,i}^*(\mathbf{u})].$$

Consider the case  $b = 1$ . Note that  $\hat{V}_n^* = |D_n^*|^{-1} \int_{D_n^*} [T_{n,1}^*(\mathbf{u}) - \mu_n]^2 d\mathbf{u}$  is unbiased for  $V_n^*$ . To show that  $\hat{V}_n^* \xrightarrow{p} V_n^*$ , it is only necessary to show that

$$C_n(\mathbf{u}_1, \mathbf{u}_2) = \frac{1}{|B_n|^2} \text{Cov} \left\{ [T_{n,1}^*(\mathbf{u}_1) - \mu_n]^2, [T_{n,1}^*(\mathbf{u}_2) - \mu_n]^2 \right\} \rightarrow 0 \text{ as } \|\mathbf{u}_1 - \mathbf{u}_2\| \rightarrow \infty \text{ and } n \rightarrow \infty.$$

Let  $\mu_n^2 = E \left\{ [T_{n,1}^*(\mathbf{u})]^2 \right\}$ . Note that

$$\begin{aligned} |B_n|^2 C_n(\mathbf{u}_1, \mathbf{u}_2) &= E \left\{ [T_{n,1}^*(\mathbf{u}_1)]^2 [T_{n,1}^*(\mathbf{u}_2)]^2 \right\} - 4\mu_n E \left\{ T_{n,1}^*(\mathbf{u}_1) [T_{n,1}^*(\mathbf{u}_2)]^2 \right\} \\ &\quad + 4(\mu_n)^2 E [T_{n,1}^*(\mathbf{u}_1) T_{n,1}^*(\mathbf{u}_2)] - (\mu_n^2)^2 + 4(\mu_n)^2 \mu_n^2 - 4(\mu_n)^4. \end{aligned}$$

Note further that

$$\begin{aligned} [T_{n,1}^*(\mathbf{u})]^2 &= \sum_{\mathbf{x}_1, \mathbf{x}_2 \in B_n(\mathbf{u})} F(\mathbf{x}_1 - \mathbf{u}) F(\mathbf{x}_2 - \mathbf{u}), \\ T_{n,1}^*(\mathbf{u}_1) T_{n,1}^*(\mathbf{u}_2) &= \sum_{\mathbf{x}_1 \in B_n(\mathbf{u}_1), \mathbf{x}_2 \in B_n(\mathbf{u}_2)} F(\mathbf{x}_1 - \mathbf{u}_1) F(\mathbf{x}_2 - \mathbf{u}_2), \\ T_{n,1}^*(\mathbf{u}_1) [T_{n,1}^*(\mathbf{u}_2)]^2 &= \sum_{\mathbf{x}_1 \in B_n(\mathbf{u}_1)} \sum_{\mathbf{x}_2, \mathbf{x}_3 \in B_n(\mathbf{u}_2)} F(\mathbf{x}_1 - \mathbf{u}_1) F(\mathbf{x}_2 - \mathbf{u}_2) F(\mathbf{x}_3 - \mathbf{u}_2), \\ [T_{n,1}^*(\mathbf{u}_1)]^2 [T_{n,1}^*(\mathbf{u}_2)]^2 &= \sum_{\mathbf{x}_1, \mathbf{x}_2 \in B_n(\mathbf{u}_1)} \sum_{\mathbf{x}_3, \mathbf{x}_4 \in B_n(\mathbf{u}_2)} F(\mathbf{x}_1 - \mathbf{u}_1) F(\mathbf{x}_2 - \mathbf{u}_1) F(\mathbf{x}_3 - \mathbf{u}_2) F(\mathbf{x}_4 - \mathbf{u}_2). \end{aligned}$$

Thus by ignoring some multiplicative constants,  $|B_n|^2 C_n(\mathbf{u}_1, \mathbf{u}_2)$  can be written as the sum of the following nine terms

$$\int_{B_n(\mathbf{u}_1) \cap B_n(\mathbf{u}_2)} [F(\mathbf{x} - \mathbf{u}_1)]^2 [F(\mathbf{x} - \mathbf{u}_2)]^2 d\mathbf{x},$$

$$\begin{aligned}
& \int_{B_n(\mathbf{u}_1)} \int_{B_n(\mathbf{u}_2)} [F(\mathbf{x}_1 - \mathbf{u}_1)]^2 [F(\mathbf{x}_2 - \mathbf{u}_2)]^2 g(\mathbf{x}_1 - \mathbf{x}_2) d\mathbf{x}_1 d\mathbf{x}_2, \\
& \int_{B_n(\mathbf{u}_1)} \int_{B_n(\mathbf{u}_1) \cap B_n(\mathbf{u}_2)} F(\mathbf{x}_1 - \mathbf{u}_1) F(\mathbf{x}_2 - \mathbf{u}_1) [F(\mathbf{x}_2 - \mathbf{u}_2)]^2 d\mathbf{x}_1 d\mathbf{x}_2, \\
& \int \int_{B_n(\mathbf{u}_1) \cap B_n(\mathbf{u}_2)} F(\mathbf{x}_1 - \mathbf{u}_1) F(\mathbf{x}_1 - \mathbf{u}_2) F(\mathbf{x}_2 - \mathbf{u}_1) F(\mathbf{x}_2 - \mathbf{u}_2) Q_2(\mathbf{x}_1, \mathbf{x}_2) d\mathbf{x}_1 d\mathbf{x}_2, \\
& \int \int \int_{B_n(\mathbf{u}_1) B_n(\mathbf{u}_2)} [F(\mathbf{x}_1 - \mathbf{u}_1)]^2 F(\mathbf{x}_2 - \mathbf{u}_2) F(\mathbf{x}_3 - \mathbf{u}_2) Q_3(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3) d\mathbf{x}_1 d\mathbf{x}_2 d\mathbf{x}_3, \\
& \int_{B_n(\mathbf{u}_1) \cap B_n(\mathbf{u}_2)} \int_{B_n(\mathbf{u}_1)} \int_{B_n(\mathbf{u}_2)} F(\mathbf{x}_1 - \mathbf{u}_1) F(\mathbf{x}_1 - \mathbf{u}_2) F(\mathbf{x}_2 - \mathbf{u}_1) F(\mathbf{x}_3 - \mathbf{u}_2) Q_2(\mathbf{x}_2, \mathbf{x}_3) d\mathbf{x}_1 d\mathbf{x}_2 d\mathbf{x}_3, \\
& \int_{B_n(\mathbf{u}_1) \cap B_n(\mathbf{u}_2)} \int_{B_n(\mathbf{u}_1)} \int_{B_n(\mathbf{u}_2)} F(\mathbf{x}_1 - \mathbf{u}_1) F(\mathbf{x}_1 - \mathbf{u}_2) F(\mathbf{x}_2 - \mathbf{u}_1) F(\mathbf{x}_3 - \mathbf{u}_2) Q_3(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3) d\mathbf{x}_1 d\mathbf{x}_2 d\mathbf{x}_3, \\
& \int \int \int \int_{B_n(\mathbf{u}_1) B_n(\mathbf{u}_2)} F(\mathbf{x}_1 - \mathbf{u}_1) F(\mathbf{x}_2 - \mathbf{u}_1) F(\mathbf{x}_3 - \mathbf{u}_2) F(\mathbf{x}_4 - \mathbf{u}_2) Q_2(\mathbf{x}_1, \mathbf{x}_3) Q_2(\mathbf{x}_2, \mathbf{x}_4) d\mathbf{x}_1 d\mathbf{x}_2 d\mathbf{x}_3 d\mathbf{x}_4, \\
& \int \int \int \int_{B_n(\mathbf{u}_1) B_n(\mathbf{u}_2)} F(\mathbf{x}_1 - \mathbf{u}_1) F(\mathbf{x}_2 - \mathbf{u}_1) F(\mathbf{x}_3 - \mathbf{u}_2) F(\mathbf{x}_4 - \mathbf{u}_2) Q_4(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4) d\mathbf{x}_1 d\mathbf{x}_2 d\mathbf{x}_3 d\mathbf{x}_4,
\end{aligned}$$

which are of order  $o(|B_n|^2)$  as  $\|\mathbf{u}_1 - \mathbf{u}_2\| \rightarrow \infty$  and  $n \rightarrow \infty$  due to conditions (11) and (12). Thus

Theorem 1 is proved.

## Appendix B: Proof of Corollary 1

Here consider only the case that  $\mu_n(\hat{\theta}_n) = \lambda_n^*(\hat{\theta}_n) \int_{B_n} f(\mathbf{s}) \lambda(\mathbf{s}; \hat{\theta}_n) d\mathbf{s}$ . The proof for the case of  $\mu_n(\hat{\theta}_n) = [b|D_n^*|]^{-1} \sum_{i=1}^b \int_{D_n^*} T_i^*(\mathbf{u}; \hat{\theta}_n) d\mathbf{u}$  follows similarly and thus is omitted. For the ease of presentation, let  $F(\mathbf{x}; \theta) = f(\mathbf{x}) \lambda(\mathbf{x}; \theta)$ ,  $\lambda_n^*(\theta) = \min_{\mathbf{s} \in D_n} \lambda(\mathbf{s}; \theta)$ ,  $p_n(\mathbf{x}; \theta) = \lambda_n^*(\theta) / \lambda(\mathbf{x}; \theta)$ , and  $N_1^*(\theta)$  be a thinned realization of  $N_1^*$  defined in (4), where  $\lambda(\cdot) = \lambda(\cdot; \theta)$  is used to produce the thinning probabilities.

Let  $r(\mathbf{x})$  be a uniform random variable in  $[0, 1]$ . If  $r(\mathbf{x}) \leq p_n(\mathbf{x}; \theta)$ , when  $\mathbf{x} \in (N \cap D_n)$ , then  $\mathbf{x}$  will be retained in  $N_1^*(\theta)$ . Based on the set  $\{r(\mathbf{x}) : \mathbf{x} \in (N \cap D_n)\}$ , we determine  $N_1^*(\theta_0)$  and  $N_1^*(\hat{\theta}_n)$ . Define  $N_a^* = \{\mathbf{x} \in N_1^*(\theta_0), \mathbf{x} \notin N_1^*(\hat{\theta}_n)\}$  and  $N_b^* = \{\mathbf{x} \notin N_1^*(\theta_0), \mathbf{x} \in N_1^*(\hat{\theta}_n)\}$ . Note that  $P(\mathbf{x} \in N_a^* \cup N_b^* | \mathbf{x} \in N) \leq |p_n(\mathbf{x}; \hat{\theta}_n) - p_n(\mathbf{x}; \theta_0)|$ .

Define

$$\begin{aligned}\mu_n(\theta) &= \lambda_n^*(\theta) \int_{B_n} F(\mathbf{s}; \theta) d\mathbf{s}, \\ T_{n,1}^*(\mathbf{u}; \theta) &= \sum_{\mathbf{x} \in B_n(\mathbf{u}) \cap N_1^*(\hat{\theta}_n)} F(\mathbf{x} - \mathbf{u}; \theta), \\ T_{n,0}^*(\mathbf{u}; \theta) &= \sum_{\mathbf{x} \in B_n(\mathbf{u}) \cap N_1^*(\theta_0)} F(\mathbf{x} - \mathbf{u}; \theta).\end{aligned}$$

To prove Corollary 1, first note that

$$\begin{aligned}& \hat{V}_n^*(\hat{\theta}_n) - \frac{1}{|D_n^*|} \int_{D_n^*} [T_{n,0}^*(\mathbf{u}; \theta_0) - \mu_n(\theta_0)]^2 d\mathbf{u} \\ &= \frac{1}{|D_n^*|} \int_{D_n^*} [T_{n,1}^*(\mathbf{u}; \hat{\theta}_n) - T_{n,1}^*(\mathbf{u}; \theta_0) + T_{n,1}^*(\mathbf{u}; \theta_0) - T_{n,0}^*(\mathbf{u}; \theta_0) + \mu_n(\theta_0) - \mu_n(\hat{\theta}_n)]^2 d\mathbf{u} \\ &+ \frac{2}{|D_n^*|} \int_{D_n^*} [T_{n,0}^*(\mathbf{u}; \theta_0) - \mu_n(\theta_0)] [T_{n,1}^*(\mathbf{u}; \hat{\theta}_n) - T_{n,1}^*(\mathbf{u}; \theta_0) + T_{n,1}^*(\mathbf{u}; \theta_0) - T_{n,0}^*(\mathbf{u}; \theta_0) \\ &\quad + \mu_n(\theta_0) - \mu_n(\hat{\theta}_n)] d\mathbf{u}.\end{aligned}$$

The above is of order  $o(|B_n|)$  in probability (i.p.) if the following are true:

- (a)  $\frac{1}{|D_n^*||B_n|} \int_{D_n^*} [T_{n,1}^*(\mathbf{u}; \hat{\theta}_n) - T_{n,1}^*(\mathbf{u}; \theta_0)]^2 d\mathbf{u} \xrightarrow{p} 0,$
- (b)  $\frac{1}{|D_n^*||B_n|} \int_{D_n^*} [T_{n,1}^*(\mathbf{u}; \theta_0) - T_{n,0}^*(\mathbf{u}; \theta_0)]^2 d\mathbf{u} \xrightarrow{p} 0,$
- (c)  $\frac{1}{|B_n|} [\mu_n(\theta_0) - \mu_n(\hat{\theta}_n)]^2 \xrightarrow{p} 0$

Condition (15) and the assumptions that  $f(\cdot)$  is finite and  $\lambda(\cdot, \theta)$  has a bounded first-order derivative

imply that for all  $\delta > 0$  and  $0 < \alpha < 1$ , there exists  $C < \infty$  such that

$$(a) \leq \frac{C\delta}{|D_n^*|^{\alpha+1}|B_n|} \int_{D_n^*} \left[ \sum_{N \cap B_n(\mathbf{u})} 1 \right]^2 d\mathbf{u} \text{ with probability 1.}$$

The above converges to zero i.p. due to conditions (11) and (12).

To show (b), note that

$$T_{n,1}^*(\mathbf{u}; \theta_0) - T_{n,0}^*(\mathbf{u}; \theta_0) = \sum_{N_b^* \cap B_n(\mathbf{u})} F(\mathbf{x} - \mathbf{u}; \theta_0) - \sum_{N_a^* \cap B_n(\mathbf{u})} F(\mathbf{x} - \mathbf{u}; \theta_0).$$

Thus there exists  $C < \infty$  such that

$$(b) \leq \frac{C}{|D_n^*||B_n|} \int_{D_n^*} \left[ \sum_{N_a^* \cap B_n(\mathbf{u})} 1 \right]^2 + \left[ \sum_{N_b^* \cap B_n(\mathbf{u})} 1 \right]^2 d\mathbf{u}.$$

The above converges to zero i.p. due to the fact that  $P(\mathbf{x} \in N_a^* \cup N_b^* | \mathbf{x} \in N) \leq \frac{C}{|D_n^*|^{\alpha/2}}$  i.p. The latter is due to condition (15) and the smoothness condition on the FOIF assumed in Corollary 1.

Finally, note that (c)  $\xrightarrow{p} 0$  due to condition (15). Thus Corollary 1 is proved.

## References

- 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.
- Baddeley, A. J., Turner, R., Møller, J. and Hazelton, M. (2005) “Residual analysis for spatial point processes (with discussion)” *Journal of the Royal Statistical Society, Series B*, 67, 617–666
- Brillinger, D. R. (1975), *Time Series: Data Analysis and Theory*, Holt, Rinehart & Winston.
- Daley, D. J. and Vere-Jones, D. (1988), *An Introduction to the Theory of Point Processes*, New York: Springer-Verlag.
- Diggle, P. J. (2003), *Statistical Analysis of Spatial Point Patterns*, New York: Oxford University Press Inc.
- Guan, Y. and Loh, J. M. (2006), “A block bootstrap procedure in modeling inhomogeneous spatial point patterns”, *submitted*.
- Heinrich, L. (1985), “Normal convergence of multidimensional shot noise and rates of this convergence”, *Advances in Applied Probability*, 17, 709–730.
- McCullagh, P. (1987), *Tensor Methods in Statistics*, New York: Chapman and Hall.

- Møller, J., Syversveen, A. R. and Waagepetersen, R. P. (1998), “Log Gaussian Cox processes”, *Scandinavian Journal of Statistics*, 25, 451–482.
- Møller, J. and Waagepetersen, R. P. (2003), *Statistical Inference and Simulation for Spatial Point Processes*, New York: Chapman & Hall.
- Papangelou, F. (1974), “The conditional intensity of general point processes and an application to line processes”, *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 28, 207–226.
- Politis, D. N. and Romano, J. P. (1994), “Large sample confidence regions based on subsamples under minimal assumptions”, *The Annals of Statistics*, 22, 2031–2050.
- Schoenberg, F. P. (2004), “Consistent parametric estimation of the intensity of a spatial-temporal point process”, *Journal of Statistical Planning and Inference*, 128(1), 79–93.
- Sherman, M. and Carlstein, E. (1996), “Replicate histograms”, *Journal of the American Statistical Association*, 91, 566–576.
- Stoyan, D. and Stoyan, H. (1994), *Fractals, Random Shapes and Point Fields*, New York: Wiley.
- Waagepetersen, R. P. (2005), “Discussion of “Residual analysis for spatial point processes””, *Journal of the Royal Statistical Society, Series B*, 67, 662.
- Waagepetersen, R. P. (2006), “An estimating function approach to inference for inhomogeneous Neyman-Scott processes”, *Biometrics*, to appear.

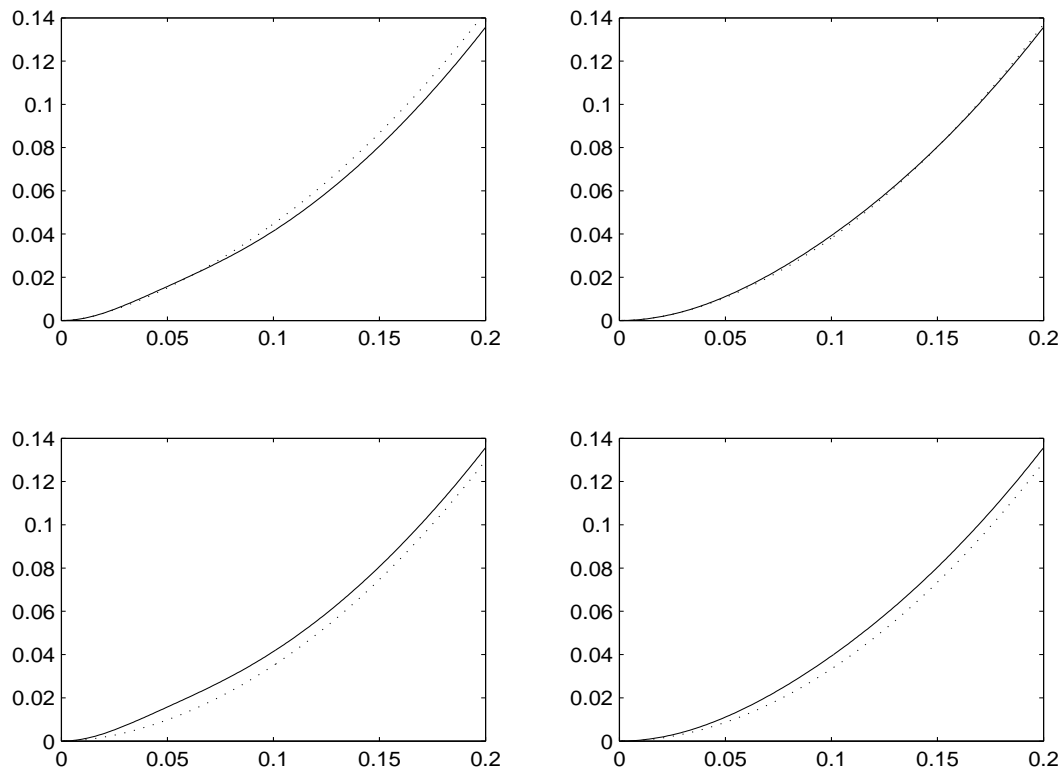


Figure 1: Plots of the true and estimated  $K$ -functions. Top to bottom  $\lambda = 200, 800$  and left to right  $\sigma = .02, .04$ . The solid line is the true function and the dotted line is the estimated function.

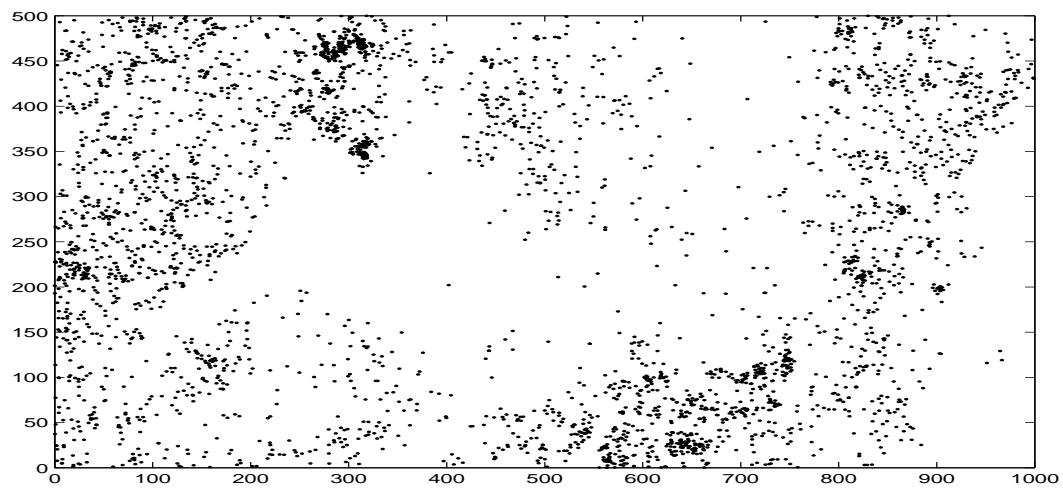


Figure 2: Locations of *Beilschmiedia pendula* Lauraceae trees.

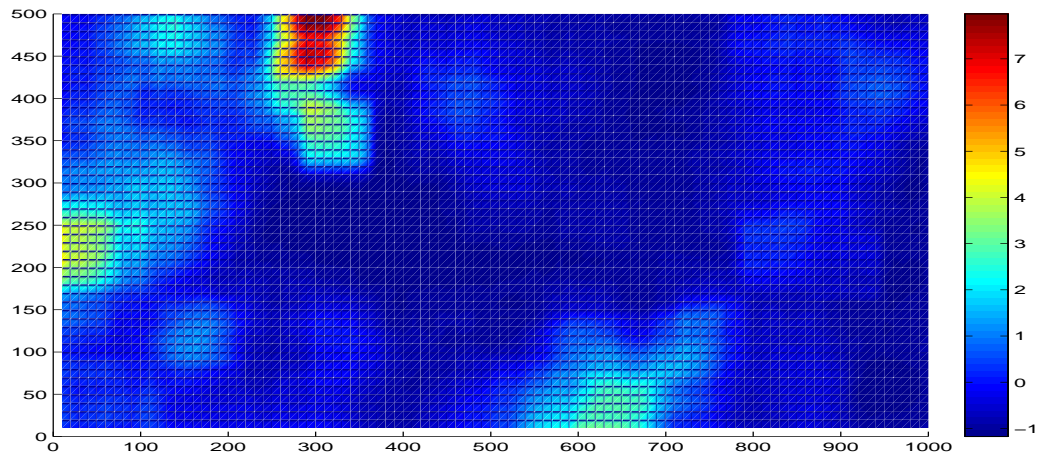


Figure 3: Smoothed residual field for the fitted model to the *Beilschmiedia pendula Lauraceae* data.

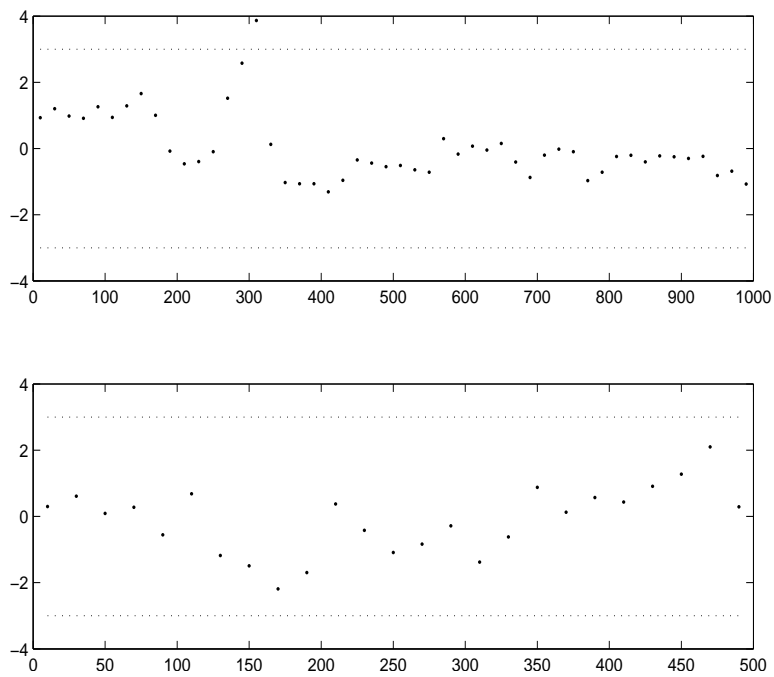


Figure 4: Plots to detect trend for the fitted model to the *Beilschmiedia pendula Lauraceae* data.

The covariates being considered are the two coordinates, i.e.  $x$  (top) and  $y$  (bottom).



Table 1: Bias and standard deviation (STD) of the variance estimator for the smoothed residual defined in (19). Each bias and STD was divided by the estimated true variance of the residual. Methods 1, 2, 3 are the proposed method, the plug-in method using the true PCF and the plug-in method using an incorrect PCF, respectively.

			$\sigma = .02$				$\sigma = .04$				
	$\lambda$	method	$x = .02$	.04	.06	.08	.02	.04	.06	.08	
BIAS	200	1	.027	-.075	-.030	.040	.047	.037	-.010	.048	
		2	.037	-.066	-.019	.055	.035	.028	-.022	.035	
		3	.102	-.002	.056	.143	.035	.028	-.023	.035	
	800	1	.010	-.027	-.034	.017	.018	.021	.004	.019	
		2	.018	-.017	-.024	.029	.011	.011	-.003	.011	
		3	.128	.101	.105	.177	.055	.061	.053	.074	
	STD	200	1	.296	.205	.214	.300	.271	.226	.228	.309
			2	.293	.196	.198	.295	.261	.216	.213	.290
			3	.318	.215	.219	.327	.260	.214	.210	.285
800		1	.169	.144	.151	.190	.164	.157	.171	.209	
		2	.158	.133	.138	.183	.161	.155	.172	.217	
		3	.181	.152	.159	.211	.164	.154	.170	.218	

Table 2: Mean squared error (MSE) of the variance estimator for the smoothed residual defined in (19). Each MSE was divided by the squared value of the estimated true variance of the residual. Methods 1, 2, 3 are the proposed method, the plug-in method using the true PCF and the plug-in method using an incorrect PCF, respectively.

			$\sigma = .02$				$\sigma = .04$				
$\beta$	$\lambda$	method	$x = .02$	.04	.06	.08	.02	.04	.06	.08	
1	200	1	.088	.048	.047	.092	.076	.053	.052	.098	
		2	.087	.043	.040	.090	.070	.047	.046	.085	
		3	.112	.046	.051	.127	.069	.046	.045	.083	
	800	1	.029	.021	.024	.036	.027	.025	.029	.044	
		2	.025	.018	.020	.034	.026	.024	.030	.047	
		3	.049	.033	.036	.076	.030	.028	.032	.053	
	2	200	1	.083	.051	.059	.107	.072	.054	.072	.162
			2	.076	.044	.050	.102	.062	.045	.057	.121
			3	.089	.053	.067	.134	.062	.043	.054	.115
800		1	.027	.024	.031	.055	.023	.024	.035	.064	
		2	.025	.021	.030	.053	.022	.027	.048	.100	
		3	.041	.039	.057	.098	.022	.026	.048	.106	