A Thinned Block Bootstrap Variance Estimation Procedure for Inhomogeneous Spatial Point Patterns

May 22, 2007

Abstract

When modeling inhomogeneous spatial point patterns, it is of interest to fit a parametric model for the first order intensity function (FOIF) of the process in terms of some measured covariates. Estimates for the regression coefficients, say $\hat{\beta}$, can be obtained by maximizing a Poisson maximum likelihood criterion (Schoenberg, 2004). Little work has been done on the asymptotic distribution of $\hat{\beta}$ except in some special cases. In this article, we show that $\hat{\beta}$ is asymptotically normal for a general class of mixing processes. To estimate the variance of $\hat{\beta}$, we propose a novel thinned block bootstrap procedure, which assumes that the point process is second-order reweigthed stationary. To apply this procedure, only the FOIF but not any high-order terms of the process needs to be estimated. We establish the consistency of the resulting variance estimator, and demonstrate its efficacy through simulations and an application to a real data example.

KEY WORDS: Block Bootstrap, Inhomogeneous Spatial Point Process, Thinning.

1 Introduction

A main interest when analyzing spatial point pattern data is to model the first-order intensity function (FOIF) of the underlying process that has generated the spatial point pattern. Heuristically, the FOIF is a function that describes the likelihood for an event of the process to occur at a given location (see Section 2 for its formal definition). We say a process is homogenous if its FOIF is a constant and inhomogeneous otherwise. In practice, we often wish to model the FOIF in relation to some measured covariates. For example, for the *Beilschmiedia pendula Lauraceae* (BPL) data given in Section 5, we would like to model the FOIF in terms of two important variables of landscape features: elevation and gradient. Since the FOIF characterizes the probability finding a BPL tree at a given location with the associated elevation and gradient values, the study of this function can yield valuable insight on how these landscape features affect the spatial distribution of BPL trees. If a significant relation can be established, then this will provide evidence in support of the niche assembly theory, which states that different species benefit from different habitats determined by local environmental features (e.g. Waagepetersen, 2007).

To model the FOIF, we assume that it can be expressed as a parametric function of the available covariates. Specifically, let N denote a spatial point process defined over $\mathbb{R}^2$, with the FOIF at $\mathbf{s} \in \mathbb{R}^2$ given by $\lambda(\mathbf{s}; \beta)$, where β is a $p \times 1$ vector of unknown regression coefficients associated with the covariates, and let D be the region where a realization of N has been observed. Our goal is then to estimate and make inference on the regression coefficients β.

To estimate β, we consider the following maximum likelihood criterion:

$$U(\beta) = \frac{1}{|D|} \sum_{\mathbf{x} \in D \cap N} \log \lambda(\mathbf{x}; \beta) - \frac{1}{|D|} \int_D \lambda(\mathbf{s}; \beta)\, d\mathbf{s}. \tag{1}$$

The maximizer of (1) is taken as an estimator for β (denoted by $\hat{\beta}$ throughout this section). Note that (1) is proportional to the true maximum likelihood if N is an inhomogeneous Poisson process, i.e. when the events of the process occurring in disjoint sets are completely independent. For the BPL example, however, the locations of the BPL trees are likely to be clustered, possibly due to seed dispersal, and/or correlation among environmental factors that have not been accounted for by the model. Schoenberg (2004) showed that under some mild conditions, $\hat{\beta}$ obtained by maximizing (1) is still consistent for β for a class of spatial-temporal point process models, even if the process is not Poisson. However, he did not provide the asymptotic distribution of $\hat{\beta}$. Waagepetersen (2007) significantly extended the scope of the method by deriving asymptotic properties of $\hat{\beta}$ including asymptotic normality for a wide class of spatial cluster processes.

To make inference on β, we will need information on the distributional properties of $\hat{\beta}$. One standard approach is to derive the limiting distribution of $\hat{\beta}$ under an appropriate asymptotic framework, and then use it as an approximation for the distribution of $\hat{\beta}$ in a finite sample setting. We note that currently available asymptotics for inhomogeneous spatial point processes are inadequate since they either assume complete spatial independence, i.e. the process is Poisson (e.g. Rathbun and Cressie, 1994), or use a parametric model for the dependence of the process (e.g. Waagepetersen, 2007). For data arising from biological studies such as the BPL data, however, the underlying biological process generating the spatial point patterns is rather complex and often not well understood. Thus, the use

of a specific form for the dependence may be debatable and could lead to incorrect inference on the regression parameters β (since the distribution of $\hat{\beta}$ depends on the dependence structure of the process).

In this paper, we study the distributional properties of $\hat{\beta}$ under an increasing domain setting. To quantify the dependence of the process, we use a more flexible, model-free mixing condition (Section 2), but do not assume any specific parametric structure on the dependence. Our main result shows that under some mild conditions, the standardized distribution of $\hat{\beta}$ is asymptotically normal. If the variance of $\hat{\beta}$ is known, approximate confidence intervals for β can then be obtained so the inference on β becomes straightforward. Thus our result further extends the scope of applications of (1) beyond Schoenberg (2004) and Waagepetersen (2007).

One complication in practice is that the variance of $\hat{\beta}$ is unknown and thus must be estimated. From Theorem 1 in Section 2, we see that the variance of $\hat{\beta}$ depends on the second-order cumulant function (SOCF) of the process, a function that is related to the dependence structure of the process. To avoid specifying the SOCF, we develop a nonparametric thinned block bootstrap procedure to estimate the variance of $\hat{\beta}$ by using a combination of a thinning algorithm and block bootstrap. Specifically, in the thinning step, we retain (i.e. do not thin) an observed event from N with a probability that is proportional to the inverse of the estimated FOIF at the location of the event. If N is second-order reweighted stationary (see Section 2) and if $\hat{\beta}$ is close to β, then the thinned process should resemble a second-order stationary (SOS) process. We show in Section 3 that the variance of $\hat{\beta}$ can be written in terms of the variance of a statistic S_n, where S_n is computed from the

thinned process and is expressed in terms of the intensity function only. The task of estimating the variance of $\hat{\beta}$ is thus translated into estimating the variance of a statistic defined on a SOS process, which we can accomplish by using block bootstrap (see Section 3 for details). We prove in Section 3 that the resulting variance estimator is L_2 consistent for the target variance, and perform a simulation study in Section 4 to investigate the performance of the proposed procedure. To illustrate its use in a practical setting, we also apply the proposed procedure to the BPL data in Section 5.

Before proceeding to the next section, we note that resampling methods including block bootstrap have been applied extensively in spatial statistics (e.g. Politis *et al.*, 1999; Lahiri, 2003). Most of these methods, however, were developed for stationary, quantitative spatial processes that are observed on a regularly-spaced grid. In the regression setting also for quantitative processes, Cressie (1993, Section 7.3.2) discussed both semiparametric and parametric bootstrap methods to resample the residuals whereas Sherman (1997) proposed a subsampling approach. Politis and Sherman (2001) and McElroy and Politis (2007) considered using subsampling for marked point processes. The former assumed the point process to be stationary, whereas the latter assumed it to be Poisson. None of the aforementioned resampling procedures can be used for inhomogeneous spatial point processes due to the unique feature of the data. Note that an inhomogeneous spatial point process by nature is not quantitative, observed at random locations, nonstationary and can be non-Poisson.

2 Notation and Preliminary Asymptotic Results

Let N be a two-dimensional spatial point process observed over a domain of interest D. For a Borel set $B \subset \mathbb{R}^2$, let $|B|$ denote the area of B, and $N(B)$ denote the number of events from N that fall in B. We define the kth-order intensity and cumulant functions of N as:

$$\lambda_k(\mathbf{s}_1, \cdots, \mathbf{s}_k) = \lim_{|d\mathbf{s}_i| \to 0} \left\{ \frac{\mathrm{E}[N(d\mathbf{s}_1) \cdots N(d\mathbf{s}_k)]}{|d\mathbf{s}_1| \cdots |d\mathbf{s}_k|} \right\}, \quad i = 1, \cdots, k,$$

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

respectively. Here $d\mathbf{s}$ is an infinitesimal region containing $\mathbf{s}$ and $\mathrm{Cum}(Y_1, \cdots, Y_k)$ is the coefficient of $i^k t_1 \cdots t_k$ in the Taylor series expansion of $\log\{\mathrm{E}[\exp(i \sum_{j=1}^k Y_j t_j)]\}$ about the origin (e.g. Brillinger, 1975). For the intensity function, $\lambda_k(\mathbf{s}_1, \cdots, \mathbf{s}_k)|d\mathbf{s}_1| \cdots |d\mathbf{s}_k|$ is the approximate probability for $d\mathbf{s}_1, \cdots, d\mathbf{s}_k$ to each contain an event. For the cumulant function, $Q_k(\mathbf{s}_1, \cdots, \mathbf{s}_k)$ describes the dependence among sites $\mathbf{s}_1, \cdots, \mathbf{s}_k$, where a close-to-zero value indicates near independence. Specifically, if N is Poisson, then $Q_k(\mathbf{s}_1, \cdots, \mathbf{s}_k) = 0$ if at least two of $\mathbf{s}_1, \cdots, \mathbf{s}_k$ are different. The k-th order cumulant (intensity) function can be expressed as a function of the intensity (cumulant) functions up to the k-th order. See Daley and Vere-Jones (1988, p.147) for details.

We study the large sample behavior of $\hat{\beta}$ under an increasing-domain setting, where $\hat{\beta}$ is obtained by maximizing (1). Specifically, consider a sequence of regions, D_n. Let ∂D_n denote the boundary of D_n, $|\partial D_n|$ denote the length of ∂D_n, and $\hat{\beta}_n$ denote $\hat{\beta}$ obtained over D_n. We assume

$$C_1 n^2 \leq |D_n| \leq C_2 n^2, C_1 n \leq |\partial D_n| \leq C_2 n \text{ for some } C_1 \leq C_2 < \infty. \tag{2}$$

Condition (2) requires that D_n must become large in all directions (i.e. the data are truly spatial) and that the boundary is not too irregular. Many commonly used domain sequences satisfy this condition. To see an example, let $A \subset (0,1] \times (0,1]$ be the interior of a simple closed curve with nonempty interior. If we define D_n as A inflated by a factor n, then D_n satisfies condition (2). Note that this formulation incorporates a wide variety of shapes, e.g. rectangular and elliptical shapes.

To formally state the large sample distributional properties of $\hat{\beta}_n$, it is necessary to quantify the dependence in N. We do so by using the model-free strong mixing coefficient (Rosenblatt, 1956), which is defined as follows:

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

In the above, the supremum is taken over all compact and convex subsets $E_1 \subset \mathbb{R}^2$, and over all $\mathbf{s} \in \mathbb{R}^2$ such that $d(E_1, E_2) \geq k$, where $d(E_1, E_2)$ is the maximal distance between E_1 and E_2 (e.g. Guan *et al.*, 2006), and $\mathcal{F}(E)$ is the σ-algebra generated by the random events of N that are in E. We assume the following mixing condition on N:

$$\sup_p \frac{\alpha(p;k)}{p} = \mathrm{O}(k^{-\epsilon}) \text{ for some } \epsilon > 2. \tag{3}$$

Condition (3) states that for any two fixed sets, the dependence between them must decay to zero at a polynomial rate of the inter-sets distance, k. The speed in which the dependence decays to zero also depends on the size of the sets (i.e. p). In particular, for a fixed k, condition (3) allows the dependence to increase as p increases. Any point process with a finite dependence range, e.g. the Matérn cluster process (Stoyan and Stoyan, 1994),

satisfies this condition. Furthermore, it is also satisfied by the log Gaussian Cox process (LGCP), which is very flexible in modeling environmental data (e.g. Møller *et al.*, 1998), if the correlation of the underlying Gaussian random field decays at a polynomial rate faster than $2 + \epsilon$ and has a spectral density bounded below from zero. This is due to Corollary 2 of Doukhan (1994, p.59).

In addition to conditions (2) and (3), we also need some mild conditions on the intensity and cumulant functions of N. In what follows, let $f^{(i)}(\beta)$ denote the ith derivative of $f(\beta)$. We assume

$$\lambda(\mathbf{s};\beta) \text{ is bounded below from zero,} \tag{4}$$

$$\lambda^{(2)}(\mathbf{s};\beta) \text{ is bounded and continuous with respect to } \beta, \tag{5}$$

$$\sup_{\mathbf{s}_1} \int \cdots \int |Q_k(\mathbf{s}_1, \cdots, \mathbf{s}_k)| d\mathbf{s}_2 \cdots d\mathbf{s}_k < C \text{ for } k = 2, 3, 4. \tag{6}$$

Conditions (4) and (5) are straightforward conditions that can be checked directly for a proposed FOIF model. Condition (6) is a fairly weak condition. It also requires the process N to be weakly dependent, but from a perspective that is different to (3). In the homogeneous case, (6) is implied by Brillinger mixing, which holds for many commonly used point process models such as Poisson cluster process, a class of doubly stochastic Poisson process (i.e. Cox process), and certain renewal process (e.g. Heinrich, 1985). In the inhomogeneous case, $Q_k(\mathbf{s}_1, \cdots, \mathbf{s}_k)$ can often be written as $\lambda(\mathbf{s}_1) \cdots \lambda(\mathbf{s}_k)\varphi_k(\mathbf{s}_2 - \mathbf{s}_1, \cdots, \mathbf{s}_k - \mathbf{s}_1)$, where $\varphi_k(\cdot)$ is a cumulant function for some homogeneous process. Then (6) holds if $\lambda(\mathbf{s})$ is bounded and $\int \cdots \int |\varphi_k(\mathbf{u}_2, \cdots, \mathbf{u}_k)| d\mathbf{u}_2 \cdots d\mathbf{u}_k < C$. Processes satisfying this condition include, but are not limited to, the LGCP, the inhomogeneous Neyman-Scott process (INSP; Waagepetersen, 2007), and any inhomogeneous process that is obtained by thinning

a homogeneous process satisfying this condition.

Theorem 1. *Assume that conditions (2)-(6) hold, and that $\hat{\beta}_n$ converges to β_0 in probability, where β_0 is the true parameter vector. Then*

$$|D_n|^{1/2}(\Sigma_n)^{-1/2}(\hat{\beta}_n - \beta_0) \xrightarrow{d} N(0, I_p),$$

where I_p is a $p \times p$ identity matrix,

$$\Sigma_n = |D_n|(A_n)^{-1}B_n(A_n)^{-1}$$

$$A_n = \int_{D_n} \frac{\lambda^{(1)}(\mathbf{s};\beta_0)[\lambda^{(1)}(\mathbf{s};\beta_0)]'}{\lambda(\mathbf{s};\beta_0)} d\mathbf{s}$$

$$B_n = A_n + \iint_{D_n} \frac{\lambda^{(1)}(\mathbf{s}_1;\beta_0)[\lambda^{(1)}(\mathbf{s}_2;\beta_0)]'}{\lambda(\mathbf{s}_1;\beta_0)\lambda(\mathbf{s}_2;\beta_0)} Q_2(\mathbf{s}_1,\mathbf{s}_2) d\mathbf{s}_1 d\mathbf{s}_2.$$

Proof. See Appendix A. □

We assume in Theorem 1 that $\hat{\beta}_n$ is a consistent estimator for β_0. Schoenberg (2004) established the consistency of $\hat{\beta}_n$ for a class of spatial-temporal processes under some mild conditions. He assumed the spatial domain was fixed but the time domain increased to infinity. His results can be directly extended to our case. Therefore, we simply assume consistency in Theorem 1. In connection with previous results, our asymptotic result coincides with that of Rathbun and Cressie (1994) in the inhomogeneous Poisson process case and Waagepetersen (2007) in the inhomogeneous Neyman-Scott process case, respectively. Furthermore, note that Σ_n depends only on the first- and second-order properties of the process.

3 Thinned Block Bootstrap

3.1 The Proposed Method

Theorem 1 states that $|D_n|^{1/2}(\Sigma_n)^{-1/2}(\hat{\beta}_n - \beta_0)$ converges to a standard multivariate normal distribution as n increases. This provides the theoretical foundation for making inference on β. In practice, however, Σ_n is typically unknown and thus must be estimated. From the definition of Σ_n in Theorem 1, we see that two quantities, i.e. A_n and B_n, need to be estimated in order to estimate Σ_n. Note that A_n depends only on the FOIF and thus can be estimated easily once a FOIF model has been fitted to the data. The quantity B_n, however, also depends on the SOCF $Q_2(\cdot)$. Unless an explicit parametric model is assumed for $Q_2(\cdot)$, which, as argued in Section 1, may be unrealistic, it is difficult to directly estimate B_n. In this section, we propose a thinned block bootstrap approach to first estimate a term that is related to B_n, and then use it to produce an estimate for B_n.

From now on, we assume that the point process N is second-order reweighted stationary (SORWS), i.e. there exists a function $g(\cdot)$ defined in $\mathbb{R}^2$ such that $\lambda_2(\mathbf{s}_1, \mathbf{s}_2) = \lambda(\mathbf{s}_1)\lambda(\mathbf{s}_2)g(\mathbf{s}_1 - \mathbf{s}_2)$. The function $g(\cdot)$ is often referred to as the pair correlation function (PCF; e.g. Stoyan and Stoyan, 1994). Note that our definition of second-order reweighted stationarity is a special case of the original definition given by Baddeley *et al.* (2000). Specifically, these two definitions coincide if the PCF exists. Examples of SORWS processes include the INSP, the LGCP, and any point process obtained by thinning a SORWS point process (e.g. Baddeley *et al.*, 2000). We also assume that β_0 is known. Thus the dependence of the FOIF on β will be suppressed in this section. In practice, the proposed

algorithm can be applied by simply replacing β_0 with its estimate $\hat{\beta}$. The theoretical justification for using $\hat{\beta}$ is given in Appendix C.

The essence of our approach is based on the fact that a SORWS point process can be thinned to be SOS by applying some proper thinning weights to the events. Specifically, we consider the following thinned process:

$$\Psi = \Big\{\mathbf{x} : \mathbf{x} \in N \cap D_n, P(\mathbf{x} \text{ is retained}) = \frac{\min_{\mathbf{s}\in D_n} \lambda(\mathbf{s})}{\lambda(\mathbf{x})}\Big\}. \tag{7}$$

Clearly Ψ is SOS since its first- and second-order intensity functions can be written as:

$$\lambda_n = \min_{\mathbf{s}\in D_n} \lambda(\mathbf{s}) \text{ and } \lambda_{2,n}(\mathbf{s}_1, \mathbf{s}_2) = (\lambda_n)^2 g(\mathbf{s}_1 - \mathbf{s}_2),$$

respectively, where $\mathbf{s}_1, \mathbf{s}_2 \in D_n$. Based on Ψ, we then define the following statistic:

$$S_n = \sum_{\mathbf{x}\in\Psi\cap D_n} \lambda^{(1)}(\mathbf{x}). \tag{8}$$

Note that $Q_2(\mathbf{s}_1, \mathbf{s}_2) = \lambda(\mathbf{s}_1)\lambda(\mathbf{s}_2)\left[g(\mathbf{s}_1 - \mathbf{s}_2) - 1\right]$ since N is SORWS. Thus the covariance matrix of S_n, where S_n is defined in (8), is given by

$$Cov(S_n) = \lambda_n \int_{D_n} \lambda^{(1)}(\mathbf{s})[\lambda^{(1)}(\mathbf{s})]' d\mathbf{s} + (\lambda_n)^2 \iint_{D_n} \frac{\lambda^{(1)}(\mathbf{s}_1)[\lambda^{(1)}(\mathbf{s}_2)]'}{\lambda(\mathbf{s}_1)\lambda(\mathbf{s}_2)} Q_2(\mathbf{s}_1, \mathbf{s}_2) d\mathbf{s}_1 d\mathbf{s}_2. \tag{9}$$

The first term on the right hand side of (9) depends only on the FOIF, and thus can be estimated once a FOIF model has been fitted. The second term, interestingly, is $(\lambda_n)^2(B_n - A_n)$. These facts suggest that if we can estimate $Cov(S_n)$, then we can use the relationship between $Cov(S_n)$ and B_n to estimate B_n in an obvious way. Note that S_n is a statistic defined on $\Psi \cap D_n$, where Ψ is a second-order stationary process. Thus the problem of estimating B_n becomes a problem of estimating the variance of a statistic calculated from a SOS process, which we can accomplish by using the block bootstrap procedure.

Toward that end, we divide D_n into k_n subblocks of the same shape and size. Let $D^i_{l(n)}$, $i = 1, \cdots, k_n$ denote these subblocks, where $l(n) = cn^\alpha$ for some $c > 0$ and $\alpha \in (0, 1)$ signifies the subblock size. For each $D^i_{l(n)}$, let $\mathbf{c}_i$ denote the "center" of the subblock, where the "centers" are defined in some consistent way across all $D^i_{l(n)}$. We then resample with replacement k_n subblocks from all the available subblocks and "glue" them together so as to create a new region that will be used to approximate D_n. We do so for a large number, say B times. For the bth collection of random subblocks, let J_b be the set of k_n random indices sampled from $\{1, \cdots, k_n\}$ that are associated with the selected subblocks. For each event $\mathbf{x} \in D^{J_b(i)}_{l(n)}$, we replace it with a new value $\mathbf{x} - \mathbf{c}_{J_b(i)} + \mathbf{c}_i$, so that all events in $D^{J_b(i)}_{l(n)}$ are properly shifted into $D^i_{l(n)}$. In other words, the ith sampled subblock, $D^{J_b(i)}_{l(n)}$, will be "glued" at where $D^i_{l(n)}$ was in the original data. We then define

$$S^b_n = \sum_{i=1}^{k_n} \sum_{\mathbf{x} \in \Psi \cap D^{J_b(i)}_{l(n)}} \lambda^{(1)}(\mathbf{x} - \mathbf{c}_{J_b(i)} + \mathbf{c}_i), \tag{10}$$

where the sum $\sum_{\mathbf{x} \in \Psi \cap D^{J_b(i)}_{l(n)}}$ is over all events in block $D^{J_b(i)}_{l(n)}$ translated into block $D^i_{l(n)}$. Based on the obtained $S^b_n, b = 1, \ldots, B$, we can calculate the sample covariance of S^b_n and use that as an estimate for $Cov(S_n)$. Note that for a given realization of N on D_n, the application of the thinning algorithm given in (7) can yield multiple thinned realizations. This will in turn lead to multiple sample covariance matrices by applying the proposed block bootstrap procedure. One possible solution is to simply average all the obtained sample covariance matrices so as to produce a single estimate for $Cov(S_n)$. To do so, let K denote the total number of thinned processes, and $S^b_{n,j}$ and $\bar{S}_{n,j}$ denote S^b_n and the bootstrapped mean of S^b_n, $b = 1, \cdots, B$ for the jth thinned process, respectively, where S^b_n

is defined in (10). We thus obtain the following estimator for the covariance matrix of S_n:

$$\widehat{Cov(S_n)} = \frac{1}{K}\sum_{j=1}^{K}\sum_{b=1}^{B}\frac{(S_{n,j}^b - \bar{S}_{n,j})(S_{n,j}^b - \bar{S}_{n,j})'}{B-1} \tag{11}$$

3.2 Consistency and Practical Issues

To study the asymptotic properties of $\widehat{Cov(S_n)}$, we also assume the following condition:

$$\frac{1}{|D_n|^{1/2}}\int_{D_n}\int_{\mathbb{R}^2\backslash D_n}|g(\mathbf{s}_1-\mathbf{s}_2)-1|d\mathbf{s}_1 d\mathbf{s}_2 < C, \tag{12}$$

where $\mathbb{R}^2\backslash D_n = \{\mathbf{s} : \mathbf{s} \in \mathbb{R}^2 \text{ but } \mathbf{s} \notin D_n\}$. Condition (12) is only slightly stronger than $\int_{\mathbb{R}^2}|g(\mathbf{s})-1|d\mathbf{s} < C$, which is implied by condition (6). In the case that D_n are regular enough (e.g. a sequence of $n \times n$ squares) such that $|(D_n+\mathbf{s})\cap\mathbb{R}^2\backslash D_n| < C||\mathbf{s}||||D_n|^{1/2}$, then condition (12) is implied by the condition $\int_{\mathbb{R}^2}||\mathbf{s}||||g(\mathbf{s})-1|d\mathbf{s} < C$, which holds for any process that has a finite range of dependence. Furthermore, it holds for the LGCP, if the correlation function of the underlying Gaussian random field generating the intensities decays to zero at a rate faster than $||\mathbf{s}||^{-3}$.

Theorem 2. *Assume that conditions (2), (4), (5) and (6) hold. Then*

$$\frac{1}{|D_n|}\widehat{Cov(S_n)} \xrightarrow{L_2} \frac{1}{|D_n|}Cov(S_n).$$

If we further assume (12), then $\frac{1}{|D_n|^2}E\{[\widehat{Cov(S_n)} - Cov(S_n)]^2\} \leq C[1/k_n + 1/|D_{l(n)}|]$ *for some* $C < \infty$.

Proof. See Appendix B. □

Theorem 2 establishes the L_2 consistency of the proposed covariance matrix estimator for a wide range of values for the number of thinned processes and the subblock size. To

apply the proposed method in practice, however, these values must be determined. For the former, one can in principle apply the proposed thinning algorithm as many times as necessary until a stable estimator can be obtained. For the latter, Theorem 2 suggests that the "optimal" rate for the subblock size is $|D_n|^{1/2}$, where the word "optimal" is in the sense of minimizing the mean squared error. This result agrees with the findings for the "optimal" subblock size for other resampling variance estimators obtained from quantitative processes (e.g. Sherman, 1996). If we assume further that the best rate can be approximated by $c_n = c|D_n|^{1/2}$, where c is an unknown constant, then we can adapt the algorithm in Hall and Jing (1996) to estimate c_n and thus to determine the "optimal" subblock size.

Specifically, let D_m^i, $i = 1, \cdots, k_n'$, be a set of subblocks contained in D_n that have the same size and shape, $S_{m,j}^i$ be S_n in (8) obtained from the jth thinned process on D_m^i, and $\bar{S}_{m,j}$ be the mean of $S_{m,j}^i$. We then define the sample variance-covariance matrix of $S_{m,j}^i$ averaged over the K replicate thinned point processes:

$$\hat{\theta}_m = \frac{1}{K} \sum_{j=1}^{K} \sum_{i=1}^{k_n'} \frac{(S_{m,j}^i - \bar{S}_{m,j})(S_{m,j}^i - \bar{S}_{m,j})'}{k_n' - 1}.$$

If D_m^i is relatively small when compared to D_n, then the above should be a good estimator for the true covariance matrix of S_m, where S_m is S_n in (8) defined on D_m. For each D_m^i, we then apply (11) to estimate $Cov(S_m)$ over a fine set of candidate subblock sizes, say $c_m' = c'|D_m^i|^{1/2}$. Let $\hat{\theta}_m(j,k)$ and $[\hat{\theta}_m^i(j,k)]'$ be the (j,k)th element of $\hat{\theta}_m$ and the resulting estimator for using c_m', respectively. We define the best c_m as the one that has the smallest value for the following criterion:

$$M(c_m') = \frac{1}{k_n'} \sum_{i=1}^{k_n'} \sum_{j=1}^{p} \sum_{k=1}^{p} \{[\hat{\theta}_m^i(j,k)]' - \hat{\theta}_m(j,k)\}^2.$$

The "optimal" subblock size for D_n, i.e. c_n, can then be estimated easily by using the relationship $c_n = c_m(|D_n|/|D_m|)^{1/2}$.

4 Simulations

4.1 Log Gaussian Cox Process

We performed a simulation study to test the results of our theoretical findings. On an observation region D of size $L \times L$, where $L = 1, 2, 4$, we simulated 1000 realizations of an inhomogeneous Poisson process with intensity function $\lambda(\mathbf{s})$ given by

$$\log \lambda(\mathbf{s}) \quad = \quad \alpha + \beta X(\mathbf{s}) + G(\mathbf{s}), \tag{13}$$

where G is a mean zero Gaussian random field. This is the LGCP model, with inhomogeneity due to the covariate X.

For values of X, we used a single realization from a different mean zero Gaussian random field. This realization was kept fixed throughout the whole simulation study, i.e. the same X was used for each of the 1000 point process realizations. Figure 1 shows a surface plot of X for $L = 4$. The values of X for $L = 1$ and 2 are the lower left corner of those shown in the figure. For G, we used the exponential model, $C(h) = \sigma^2 \exp(-h/\rho)$, for its covariance function, where $\rho = 0.1, 0.2$ and $\sigma = 0.1$. We set α and β to be 7.02 and 2, respectively. The point process patterns were generated using the *rpoispp* function in the *spatstat* R package (Baddeley and Turner, 2005). The average number of points simulated were about 1200, 4500 and 16700 for $L = 1, 2, 4$, respectively.

For each of the 1000 point patterns, the estimates $\hat{\alpha}$ and $\hat{\beta}$ were obtained by maximizing

the Poisson likelihood (using the *ppm* function in *spatstat*). The point pattern was next thinned according to the criterion described in Section 3. Twenty independent thinned processes were obtained for each realization. The average number of points in the thinned patterns was roughly 490, 1590 and 5400 respectively for $L = 1, 2$ and 4, so about 30-40% of the points were retained. The bootstrap procedure was then applied to each of the thinned processes to get the variance estimates for $\hat{\alpha}$ and $\hat{\beta}$. From (13), we find that the quantities to be computed with each bootstrap sample are $\sum \hat{\lambda}(\mathbf{s}^*)$ and $\sum X(\mathbf{s}^*)\hat{\lambda}(\mathbf{s}^*)$ for α and β respectively, where $\mathbf{s}^*$ represents the locations of the points in a bootstrap sample. We used non-overlapping square blocks in the bootstrap procedure. For $L = 1$, we used blocks with side length 1/4, 1/3 and 1/2, while for $L = 2, 4$, we used blocks of side length 1/4, 1/3, 1/2 and 1. The number of bootstrap samples used was 499.

The variances of $\hat{\alpha}$ and $\hat{\beta}$ were computed from the bootstrap variances using Theorem 1, (9) and the relation of $Cov(S_n)$ to B_n. The variance estimates were averaged across the independent thinnings. Finally, nominal 95% confidence intervals were constructed, yielding a total of 1000 confidence intervals for α and β.

Table 1 shows the empirical coverage of the confidence intervals for β. The coverage for α (not shown) is similar but slightly lower. For both values of ρ, we find that the empirical coverage increases towards the nominal level as the observation region size increases. This increase in coverage agrees with what we expect from the theory. The empirical coverage for $\rho = 0.1$ is typically closer to the nominal level than for $\rho = 0.2$. Note that $\rho = 0.1$ corresponds to a weaker dependence. Thus we expect our method to work better if the dependence in the data is relatively weaker. It's interesting to note that for $L = 2, 4$, the

best coverage for $\rho = 0.2$ is achieved by using a block that is larger than its counterpart for $\rho = 0.1$. Thus for errors with longer range dependence, it appears that the block bootstrap procedure requires larger blocks to work well.

Table 2 shows the bootstrap estimates of the standard errors of $\hat{\beta}$ for $\rho = 0.2$, averaged over the 1000 realizations. The standard deviations of these estimates over the 1000 realizations are included in brackets. The "true" standard errors computed using the estimated regression parameters from the independent realizations are also shown. The table shows that the bootstrap estimates have a negative bias, becoming proportionally closer to the "true" standard errors as the region size L increases.

In summary, we find that our method works reasonably well when the regression parameters are estimated. We find that there is a slight undercoverage in each instance, but the undercoverage becomes smaller as the sample size increases. We also estimated the standard errors by using the true regression parameters in our thinning step. The coverage increases only slightly.

4.2 Inhomogeneous Neyman-Scott Process

Theorem 1 gives the variance in terms of the FOIF and the SOCF. Another approach for the variance estimation is to use the plug-in method. As point out by one referee, we may simply estimate the PCF or the K-function using the original data (Baddeley *et al.*, 2000). This will yield an estimate of the SOCF which can in turn be plugged into the expression for B_n in Theorem 1 in order to obtain an estimate for the asymptotic variance. Waagepetersen (2007) studied the performance of this method in the INSP case.

To compare the performance of our method with the plug-in method, we also simulated data from the INSP model given in Waagepetersen (2007). To do so, we first simulated a homogeneous Poisson process with intensity $\kappa = 50$ as the parent process. For each parent, we then generated a Poisson number of offspring. We defined the position of each offspring relative to its parent by a radially symmetric Gaussian random variable (e.g. Diggle, 2003). Let ω denote the standard deviation of the variable. We used $\omega = 0.02, 0.04$, representing relatively strong and weak clustering. Finally, we thinned the offspring as in Waagepetersen (2007) by setting the probability to retain an offspring equal to the intensity at the offspring location divided by the maximum intensity in the study region. For the intensity, we used $\log \lambda(\mathbf{s}) = \alpha + \beta X(\mathbf{s})$, where α, β and $X(\mathbf{s})$ were as defined in the LGCP case.

We simulated 1000 realizations of the process on both a 1×1 and a 2×2 square. For each realization, we applied the thinned block bootstrap and the plug-in method in Waagepetersen (2007) to estimate the standard errors of $\hat{\alpha}$ and $\hat{\beta}$ and to further obtain their respective 95% confidence intervals. For the plug-in method, we estimated the second-order parameters, κ and ω, by a minimum contrast estimation procedure. Specifically, we obtained these estimates by minimizing

$$\int_0^a [\hat{K}(t) - K(t; \kappa, \omega)]^2 dt, \tag{14}$$

with respect to (κ, ω) for a specified a, where $\hat{K}(t)$ and $K(t; \kappa, \omega)$ are the empirical and theoretical K-functions, respectively. We used $a = 4\omega$ following the recommendation of **** (2007). The estimated values of κ and ω were then used to estimate the standard errors of $\hat{\alpha}$ and $\hat{\beta}$. To do so, we used the *inhom.thomas.asympcov* function in the *InhomCluster* R package, which is available on Waagepetersen's website at

http://www.math.aau.dk/∼rw/sppcode/.

Table 3 shows the empirical coverage of the confidence intervals for β. As in the LGCP case, there is a slight undercoverage, but the undercoverage appears to be less serious. In particular, for $L = 2$ and $\omega = 0.02$, the coverage is very close to the nominal level for all block sizes being used. Furthermore, the coverage is better for $\omega = 0.02$ than for $\omega = 0.04$. This is because the former yields a process with a weaker dependence than the latter. For the plug-in method, we find that the coverage are all very close to the nominal level. Thus our method does not perform as well as the plug-in method when the true SOCF is used. However, the difference in coverage tends to diminish as the sample size increases. As pointed out by Waagepetersen (2007), the performance of the plug-in method is affected by the choice of the tuning parameter a in (14). In the simulation, we also considered using the default value of a given by *spatstat*. Results not shown here suggested that the coverage often were worse than those from our method.

5 An Application

We applied the proposed thinned block bootstrap procedure to a tropical rain forest data set that was collected in a 1000×500 meters plot in Barro Colorado Island. The data contain measurements for over 300 species existing in the same plot in multiple years, as well as information on elevation and gradient recorded on a 5×5 meters grid within the same region. We were particularly interested in modeling the locations of 3605 *Beilschmiedia pendula Lauraceae* (BPL) trees (see Figure 2) recorded in a 1995 census by using elevation and gradient as covariates. The same data set was analyzed by Waagepetersen (2007), in

which he considered the following FOIF model:

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

In the above, $E(\mathbf{s})$ and $G(\mathbf{s})$ are the (estimated) elevation and gradient at location $\mathbf{s}$, respectively. Waagepetersen (2007) fitted the model by maximizing the Poisson maximum likelihood criterion given in (1). Specifically, he obtained the estimates $\hat{\beta}_1 = 0.02$ and $\hat{\beta}_2 = 5.84$. To estimate the standard errors associated with these estimates, he assumed further that the locations of the BPL trees were generated by an INSP. Based on the estimated standard errors, he concluded that β_2 was significant but β_1 was not.

We reanalyzed the data since the INSP model was only a "crude" model for the data, as pointed out by Waagepetersen (2007). The INSP model attributed the possible clustering among the BPL trees to a one-round seed dispersal process. In reality, however, the clustering might be due to many different factors, e.g. important environmental factors other than elevation and gradient, which had not been included in the model. Even if the clustering was due to seed dispersal alone, it was probably due to seed dispersal that had happened in multiple generations. As a result, the validity of the model, and the subsequent estimated standard errors and inference on the regression parameters, should be further investigated. Our proposed thinned blocking bootstrap method, on the other hand, required much weaker conditions, and thus might yield more reasonable estimates for the standard errors and more reliable inference on the regression parameters.

To estimate the standard errors for $\hat{\beta}_1$ and $\hat{\beta}_2$, we applied (12) with $K = 100$ and $B = 999$, and used 200×100 meters subblocks. We chose $K = 100$ since the trace plots for the estimated standard errors became fairly stable if 100 and more thinned realizations

were used. The subblock size was selected by using the data-driven procedure discussed at the end of Section 3. Table 4 gives the estimated standard errors for $\hat{\beta}_1$ and $\hat{\beta}_2$ by using the thinned block bootstrap and the plug-in method in Waagepetersen (2007), and the respective 95% confidence intervals based on these estimated standard errors. Note that the intervals based on the thinned block bootstrap are slightly shorter than those from the plug-in method, but nevertheless lead to the same conclusion as the plug-in method. Specifically, they suggest that β_2 is significant but β_1 is not. From a biological point of view, this means that the BPL trees prefer to live on slopes but do not really favor either high or low altitudes. Thus our results formally confirmed the findings in Waagepetersen (2007) by using less restrictive assumptions.

Appendix A: Proof of Theorem 1

Let β_0 stand for the true regression parameter. For the ease of presentation but without losing generality, we assume that $\hat{\beta}$ and β_0 are both scalars, i.e. $p = 1$. In the case of $p > 1$, the proof can be easily generalized by the application of the Cramer-Wold device.

Lemma 1. *Assume that conditions (4)-(6) hold. Then* $|D_n|^2 E\{[U_n^{(1)}(\beta_0)]^4\} < C$ *for some* $C < \infty$.

Proof. By repeatedly using the Campbell's theorem (e.g. Stoyan and Stoyan, 1994) and the relationship between moments and cumulants, and by using the fact that $\lambda^{(1)}(\mathbf{s}; \beta_0)/\lambda(\mathbf{s}; \beta_0)$ is bounded due to conditions (4) and (5), we can derive that $|D_n|^4 E\{[U_n^{(1)}(\beta_0)]^4\}$ is bounded

by the following (if ignoring some multiplicative constants):

$$\begin{aligned}
& \iiiint |Q_4(\mathbf{s}_1,\mathbf{s}_2,\mathbf{s}_3,\mathbf{s}_4)| d\mathbf{s}_1 d\mathbf{s}_2 d\mathbf{s}_3 d\mathbf{s}_4 + \Big[\iint |Q_2(\mathbf{s}_1,\mathbf{s}_2)| d\mathbf{s}_1 d\mathbf{s}_2\Big]^2 \\
+ & \iiint |Q_3(\mathbf{s}_1,\mathbf{s}_2,\mathbf{s}_3)| d\mathbf{s}_1 d\mathbf{s}_2 d\mathbf{s}_3 + |D_n| \iint |Q_2(\mathbf{s}_1,\mathbf{s}_2)| d\mathbf{s}_1 d\mathbf{s}_2 \\
+ & \iint |Q_2(\mathbf{s}_1,\mathbf{s}_2)| d\mathbf{s}_1 d\mathbf{s}_2,
\end{aligned}$$

where all the integrals in the above are over D_n. It then follows from condition (6) that the above sum is of order $|D_n|^2$. So the lemma is proved. □

Proof of Theorem 1: First note that

$$U_n^{(1)}(\beta) = \frac{1}{|D_n|}\Big[\sum_{\mathbf{x}\in N\cap D_n} \frac{\lambda^{(1)}(\mathbf{x};\beta)}{\lambda(\mathbf{x};\beta)} - \int \lambda^{(1)}(\mathbf{s};\beta) d\mathbf{s}\Big],$$

$$\begin{aligned}
U_n^{(2)}(\beta) &= \frac{1}{|D_n|}\Big[\sum_{\mathbf{x}\in N\cap D_n} \frac{\lambda^{(2)}(\mathbf{x};\beta)}{\lambda(\mathbf{x};\beta)} - \int \lambda^{(2)}(\mathbf{s};\beta) d\mathbf{s}\Big] - \frac{1}{|D_n|}\sum_{\mathbf{x}\in N\cap D_n}\Big[\frac{\lambda^{(1)}(\mathbf{x};\beta)}{\lambda(\mathbf{x};\beta)}\Big]^2 \\
&:= U_{n,a}^{(2)}(\beta) - U_{n,b}^{(2)}(\beta).
\end{aligned}$$

By using the Taylor expansion, we can obtain

$$|D_n|^{1/2}(\hat\beta_n - \hat\beta_0) = -[U_n^{(2)}(\tilde\beta_n)]^{-1}|D_n|^{1/2}U_n^{(1)}(\beta_0),$$

where $\tilde\beta_n$ is between $\hat\beta_n$ and β_0. We need to show that

$$U_{n,a}^{(2)}(\beta_0) \xrightarrow{p} 0,$$

$$U_{n,b}^{(2)}(\beta_0) \xrightarrow{p} \frac{1}{|D_n|}\int \frac{[\lambda^{(1)}(\mathbf{s};\beta_0)]^2}{\lambda(\mathbf{s};\beta_0)} d\mathbf{s}.$$

To show the first, note that $E[U_{n,a}^{(2)}(\beta_0)] = 0$. Thus, we only need look at the variance term:

$$\begin{aligned}
Var\left[U_{n,a}^{(2)}(\beta_0)\right] &= \frac{1}{|D_n|^2}\iint \frac{\lambda^{(2)}(\mathbf{s}_1;\beta_0)\lambda^{(2)}(\mathbf{s}_2;\beta_0)}{\lambda(\mathbf{s}_1;\beta_0)\lambda(\mathbf{s}_2;\beta_0)} Q_2(\mathbf{s}_1,\mathbf{s}_2) d\mathbf{s}_1 d\mathbf{s}_2 \\
&+ \frac{1}{|D_n|^2}\int \frac{[\lambda^{(2)}(\mathbf{s};\beta_0)]^2}{\lambda(\mathbf{s};\beta_0)} d\mathbf{s},
\end{aligned}$$

which converges to zero due to conditions (4)-(6). Thus $U_{n,a}^{(2)}(\beta_0) \xrightarrow{p} 0$.

To show the second, note that $E\left[U_{n,b}^{(2)}(\beta_0)\right] = \frac{1}{|D_n|}\int \frac{[\lambda^{(1)}(\mathbf{s};\beta_0)]^2}{\lambda(\mathbf{s};\beta_0)} d\mathbf{s}$. Thus, we again only need to consider the variance term:

$$\begin{aligned} Var\left[U_{n,b}^{(2)}(\beta_0)\right] &= \frac{1}{|D_n|^2}\iint \frac{[\lambda^{(1)}(\mathbf{s}_1;\beta_0)\lambda^{(1)}(\mathbf{s}_2;\beta_0)]^2}{\lambda(\mathbf{s}_1;\beta_0)\lambda(\mathbf{s}_2;\beta_0)} Q_2(\mathbf{s}_1,\mathbf{s}_2) d\mathbf{s}_1 d\mathbf{s}_2 \\ &+ \frac{1}{|D_n|^2}\int \frac{[\lambda^{(1)}(\mathbf{s};\beta_0)]^4}{\lambda(\mathbf{s};\beta_0)} d\mathbf{s}, \end{aligned}$$

which converges to zero due to conditions (4)-(6). Thus $U_{n,b}^{(2)}(\beta_0) \xrightarrow{p} \frac{1}{|D_n|}\int \frac{[\lambda^{(1)}(\mathbf{s};\beta_0)]^2}{\lambda(\mathbf{s};\beta_0)} d\mathbf{s}$.

Now we want to show that $|D_n|^{1/2}U_n^{(1)}(\beta_0)$ converges to a normal distribution. Firstly, we derive the mean and variance of $|D_n|^{1/2}U_n^{(1)}(\beta_0)$. Clearly $E\left[|D_n|^{1/2}U_n^{(1)}(\beta_0)\right] = 0$. For the variance,

$$\begin{aligned} Var\left[|D_n|^{1/2}U_n^{(1)}(\beta_0)\right] &= \frac{1}{|D_n|}\iint \frac{\lambda^{(1)}(\mathbf{s}_1;\beta_0)\lambda^{(1)}(\mathbf{s}_2;\beta_0)}{\lambda(\mathbf{s}_1;\beta_0)\lambda(\mathbf{s}_2;\beta_0)} Q_2(\mathbf{s}_1,\mathbf{s}_2) d\mathbf{s}_1 d\mathbf{s}_2 \\ &+ \frac{1}{|D_n|}\int \frac{[\lambda^{(1)}(\mathbf{s};\beta_0)]^2}{\lambda(\mathbf{s};\beta_0)} d\mathbf{s}. \end{aligned}$$

Now consider a new sequence of regions D_n^*, where $D_n^* \subset D_n$ and $|D_n^*|/|D_n| \to 1$ as $n \to \infty$. Let $U_{n,*}^{(1)}(\beta_0)$ be the estimating function for the realization of the point process on D_n^*. Based on the expression of the variance, we can deduce that

$$Var\left[|D_n|^{1/2}U_n^{(1)}(\beta_0) - |D_n^*|^{1/2}U_{n,*}^{(1)}(\beta_0)\right] \to 0 \text{ as } n \to \infty.$$

Thus,

$$|D_n|^{1/2}U_n^{(1)}(\beta_0) \sim |D_n^*|^{1/2}U_{n,*}^{(1)}(\beta_0),$$

where the notation $a_n \sim b_n$ means that a_n and b_n have the same limiting distribution. Thus we only need to show that $|D_n^*|^{1/2}U_{n,*}^{(1)}(\beta_0)$ converges to a normal distribution for some properly defined D_n^*.

To obtain a sequence of D_n^*, we apply the blocking technique used in Guan *et al.* (2004). Let $l(n) = n^\alpha$, $m(n) = n^\alpha - n^\eta$ for $4/(2+\epsilon) < \eta < \alpha < 1$, where ϵ is defined as in condition (3). We first divide the original domain D_n into some nonoverlapping $l(n) \times l(n)$ subsquares, $D^i_{l(n)}$, $i = 1, \cdots, k_n$; within each subsquare, we further obtain $D^i_{m(n)}$, an $m(n) \times m(n)$ square sharing the same center with $D^i_{l(n)}$. Note that $d(D^i_{m(n)}, D^j_{m(n)}) \geq n^\eta$ for $i \neq j$. Now define

$$D_n^* = \cup_{i=1}^{k_n} D^i_{m(n)}.$$

Condition (2) implies that $|D_n^*|/|D_n| \to 1$. So we only need to show $|D_n^*|^{1/2} U^{(1)}_{n,*}(\beta_0)$ converges to a normal distribution. This is true due to the mixing condition, the result in Lemma 1, and the application of the Lyapunov's central limit theorem. The proof similar to that in Guan *et al.* (2004). We thus omit the details.

Appendix B: Proof of Theorem 2

To simplify notation, let $\varphi(\mathbf{u}) = g(\mathbf{u}) - 1$, $\varphi_k(\mathbf{s}_1, \cdots, \mathbf{s}_k) = Q_k(\mathbf{s}_1, \cdots, \mathbf{s}_k)/[\lambda(\mathbf{s}_1) \cdots \lambda(\mathbf{s}_k)]$, $g_k(\mathbf{s}_1, \cdots, \mathbf{s}_k) = \lambda_k(\mathbf{s}_1, \cdots, \mathbf{s}_k)/[\lambda(\mathbf{s}_1) \cdots \lambda(\mathbf{s}_k)]$ for $k = 3, 4$, and $Z(\mathbf{s}) = \lambda^{(1)}(\mathbf{s})$. As in the proof of Theorem 1, we assume that $\hat{\beta}$ and β_0 are both scalars, i.e. $p = 1$. As a result, $Z(\mathbf{s})$ is a scalar so that $Cov(S_n) = Var(S_n)$. Furthermore, we only consider the case $K = 1$. Thus the covariance estimator defined in (11) becomes:

$$\widehat{Var(S_n)} = \sum_{b=1}^{B} \frac{(S_n^b - \bar{S}_n)^2}{B-1}.$$

As B goes to infinity, we see that the above converges to

$$Var(S_n^b | \Psi \cap D_n)$$

$$
\begin{aligned}
&= \frac{1}{k_n}\sum_{i=1}^{k_n}\sum_{j=1}^{k_n}\sum_{D^j_{l(n)}}\sum_{D^j_{l(n)}} Z(\mathbf{x}_1-\mathbf{c}_j+\mathbf{c}_i)Z(\mathbf{x}_2-\mathbf{c}_j+\mathbf{c}_i)\\
&- \frac{1}{k_n^2}\sum_{i=1}^{k_n}\sum_{j_1=1}^{k_n}\sum_{j_2=1}^{k_n}\sum_{D^{j_1}_{l(n)}}\sum_{D^{j_2}_{l(n)}} Z(\mathbf{x}_1-\mathbf{c}_{j_1}+\mathbf{c}_i)Z(\mathbf{x}_2-\mathbf{c}_{j_2}+\mathbf{c}_i),
\end{aligned}
$$

where the notation $\sum_D$ is short for $\sum_{D\cap\Psi}$. We wish to show $Var(S_n^b|\Psi\cap D_n)/|D_n|$ converges in L_2 to

$$
\frac{1}{|D_n|}Var(S_n) = \frac{(\lambda_n)^2}{|D_n|}\int_{D_n}\int_{D_n} Z(\mathbf{s}_1)Z(\mathbf{s}_2)\varphi(\mathbf{s}_1-\mathbf{s}_2)d\mathbf{s}_1 d\mathbf{s}_2 + \frac{\lambda_n}{|D_n|}\int_{D_n}[Z(\mathbf{s})]^2 d\mathbf{s}.
$$

To show this, we need to show the following:

$$
\frac{1}{|D_n|}E[Var(S_n^b|\Psi\cap D_n)] \to \frac{1}{|D_n|}Var(S_n), \tag{15}
$$

$$
\frac{1}{|D_n|^2}Var[Var(S_n^b|\Psi\cap D_n)] \to 0. \tag{16}
$$

To show (15), note that

$$
\begin{aligned}
&E[Var(S_n^b|\Psi\cap D_n)]\\
&= (\lambda_n)^2\sum_{i=1}^{k_n}\iint_{D^i_{l(n)}} Z(\mathbf{s}_1)Z(\mathbf{s}_2)g(\mathbf{s}_1-\mathbf{s}_2)d\mathbf{s}_1 d\mathbf{s}_2 + \frac{\lambda_n(k_n-1)}{k_n}\int_{D_n}[Z(\mathbf{s})]^2 d\mathbf{s}\\
&- \frac{(\lambda_n)^2}{k_n^2}\sum_{i=1}^{k_n}\sum_{j_1=1}^{k_n}\sum_{j_2=1}^{k_n}\iint_{D^i_{l(n)}} Z(\mathbf{s}_1)Z(\mathbf{s}_2)g(\mathbf{s}_1-\mathbf{s}_2+\mathbf{c}_{j_1}-\mathbf{c}_{j_2})d\mathbf{s}_1 d\mathbf{s}_2.
\end{aligned}
$$

Thus,

$$
\begin{aligned}
&\frac{1}{|D_n|}\left\{E[Var(S_n^b|\Psi\cap D_n)] - Var(S_n)\right\}\\
&= -\frac{\lambda_n}{k_n|D_n|}\int_{D_n}[Z(\mathbf{s})]^2 d\mathbf{s} - \frac{(\lambda_n)^2}{|D_n|}\sum\sum_{i\neq j}\int_{D^i_{l(n)}}\int_{D^j_{l(n)}} Z(\mathbf{s}_1)Z(\mathbf{s}_2)\varphi(\mathbf{s}_1-\mathbf{s}_2)d\mathbf{s}_1 d\mathbf{s}_2\\
&- \frac{(\lambda_n)^2}{|D_n|}\frac{1}{k_n^2}\sum_{i=1}^{k_n}\sum_{j_1=1}^{k_n}\sum_{j_2=1}^{k_n}\iint_{D^i_{l(n)}} Z(\mathbf{s}_1)Z(\mathbf{s}_2)\varphi(\mathbf{s}_1-\mathbf{s}_2+\mathbf{c}_{j_1}-\mathbf{c}_{j_2})d\mathbf{s_1} d\mathbf{s}_2\\
&:= -T_n^1 - T_n^2 - T_n^3.
\end{aligned}
$$

T_n^1 goes to zero due to conditions (2) and (5). For T_n^2 and T_n^3, lengthy yet elementary derivations yield:

$$\begin{aligned} |T_n^2| &\leq \frac{C(\lambda_n)^2}{|k_n|}\sum_{i=1}^{k_n}\Big[\frac{1}{|D_n|}\int_{D_n-D_n}|D_n\cap(D_n-\mathbf{s})||\varphi(\mathbf{s})|d\mathbf{s} \\ &\qquad -\frac{1}{|D^i_{l(n)}|}\int_{D^i_{l(n)}-D^i_{l(n)}}|D^i_{l(n)}\cap(D^i_{l(n)}-\mathbf{s})||\varphi(\mathbf{s})|d\mathbf{s}\Big], \end{aligned}$$

$$T_n^3 \leq \frac{C(\lambda_n)^2}{k_n}\frac{1}{|D_n|}\int_{D_n}\int_{D_n}|\varphi(\mathbf{s}_1-\mathbf{s}_2)|d\mathbf{s}_1 d\mathbf{s}_2.$$

Both terms converge to zero due to conditions (2), (4), (5) and (6). Thus (15) is proved. Furthermore, we can see that the order of T_n^1 and T_n^3 are both $1/k_n$, while that of T_n^2 is $1/|D_{l(n)}|^{1/2}$ due to condition (12).

To prove (16), we first note that $[Var(S_n^b|\Psi\cap D_n)]^2$ is equal to the sum of the following three terms:

$$\frac{1}{k_n^2}\sum_{i_1=1}^{k_n}\sum_{i_2=1}^{k_n}\sum_{j_1=1}^{k_n}\sum_{j_2=1}^{k_n}\sum_{D^{j_1}_{l(n)}}\sum_{D^{j_1}_{l(n)}}\sum_{D^{j_2}_{l(n)}}\sum_{D^{j_2}_{l(n)}}$$
$$\Big[Z(\mathbf{x}_1-\mathbf{c}_{j_1}+\mathbf{c}_{i_1})Z(\mathbf{x}_2-\mathbf{c}_{j_1}+\mathbf{c}_{i_1})Z(\mathbf{x}_3-\mathbf{c}_{j_2}+\mathbf{c}_{i_2})Z(\mathbf{x}_4-\mathbf{c}_{j_2}+\mathbf{c}_{i_2})\Big]$$

$$-\frac{2}{k_n^3}\sum_{i_1=1}^{k_n}\sum_{i_2=1}^{k_n}\sum_{j_1=1}^{k_n}\sum_{j_2=1}^{k_n}\sum_{j_3=1}^{k_n}\sum_{D^{j_1}_{l(n)}}\sum_{D^{j_1}_{l(n)}}\sum_{D^{j_2}_{l(n)}}\sum_{D^{j_3}_{l(n)}}$$
$$\Big[Z(\mathbf{x}_1-\mathbf{c}_{j_1}+\mathbf{c}_{i_1})Z(\mathbf{x}_2-\mathbf{c}_{j_1}+\mathbf{c}_{i_1})Z(\mathbf{x}_3-\mathbf{c}_{j_2}+\mathbf{c}_{i_2})Z(\mathbf{x}_4-\mathbf{c}_{j_3}+\mathbf{c}_{i_2})\Big]$$

$$\frac{1}{k_n^4}\sum_{i_1=1}^{k_n}\sum_{i_2=1}^{k_n}\sum_{j_1=1}^{k_n}\sum_{j_2=1}^{k_n}\sum_{j_3=1}^{k_n}\sum_{j_4=1}^{k_n}\sum_{D^{j_1}_{l(n)}}\sum_{D^{j_2}_{l(n)}}\sum_{D^{j_3}_{l(n)}}\sum_{D^{j_4}_{l(n)}}$$
$$\Big[Z(\mathbf{x}_1-\mathbf{c}_{j_1}+\mathbf{c}_{i_1})Z(\mathbf{x}_2-\mathbf{c}_{j_2}+\mathbf{c}_{i_1})Z(\mathbf{x}_3-\mathbf{c}_{j_3}+\mathbf{c}_{i_2})Z(\mathbf{x}_4-\mathbf{c}_{j_4}+\mathbf{c}_{i_2})\Big]$$

Furthermore, the following relationships are possible for $\mathbf{x}_1,\cdots,\mathbf{x}_4$:

a) $\mathbf{x}_1=\mathbf{x}_2=\mathbf{x}_3=\mathbf{x}_4$

b) $\mathbf{x}_1 = \mathbf{x}_2$ and $\mathbf{x}_3 = \mathbf{x}_4$ but $\mathbf{x}_1 \neq \mathbf{x}_3$, or $\mathbf{x}_1 = \mathbf{x}_3$ and $\mathbf{x}_2 = \mathbf{x}_4$ but $\mathbf{x}_1 \neq \mathbf{x}_2$

c) $\mathbf{x}_1 \neq \mathbf{x}_2$ but $\mathbf{x}_2 = \mathbf{x}_3 = \mathbf{x}_4$, or $\mathbf{x}_3 \neq \mathbf{x}_4$ but $\mathbf{x}_1 = \mathbf{x}_2 = \mathbf{x}_3$

d) $\mathbf{x}_1 \neq \mathbf{x}_2 \neq \mathbf{x}_3$ but $\mathbf{x}_3 = \mathbf{x}_4$, or $\mathbf{x}_1 \neq \mathbf{x}_2 \neq \mathbf{x}_3$ but $\mathbf{x}_1 = \mathbf{x}_4$,

or $\mathbf{x}_2 \neq \mathbf{x}_3 \neq \mathbf{x}_4$ but $\mathbf{x}_1 = \mathbf{x}_2$

e) $\mathbf{x}_1 \neq \mathbf{x}_2 \neq \mathbf{x}_3 \neq \mathbf{x}_4$

When calculating the variance of $\frac{1}{|D_n|}Var(S_n^b|\Psi \cap D_n)$, the above relationships, combined with the expressions for the squared value of $\frac{1}{|D_n|}E[Var(S_n^b|\Psi \cap D_n)]$, in turn lead to integrals of different complexity. To save space, we present only the integral from term e). The remaining integrals can be shown to have order no higher than $1/k_n$ in a similar manner.

Let $F_1(\mathbf{s}_1, \mathbf{s}_2, \mathbf{s}_3, \mathbf{s}_4) = g_4(\mathbf{s}_1, \mathbf{s}_2, \mathbf{s}_3, \mathbf{s}_4) - g(\mathbf{s}_1, \mathbf{s}_2)g(\mathbf{s}_3, \mathbf{s}_4)$. Term e) leads to the following integral:

$$
\begin{aligned}
&\frac{(\lambda_n)^4}{k_n^2|D_n|^2}\sum_{i_1=1}^{k_n}\sum_{i_2=1}^{k_n}\sum_{j_1=1}^{k_n}\int_{D_{l(n)}^{j_1}} Z(\mathbf{s}_1 - \mathbf{c}_{j_1} + \mathbf{c}_{i_1}) \\
&\Big[\sum_{j_2=1}^{k_n}\int_{D_{l(n)}^{j_1}}\int_{D_{l(n)}^{j_2}}\int_{D_{l(n)}^{j_2}} Z(\mathbf{s}_2 - \mathbf{c}_{j_1} + \mathbf{c}_{i_1})Z(\mathbf{s}_3 - \mathbf{c}_{j_2} + \mathbf{c}_{i_2})Z(\mathbf{s}_4 - \mathbf{c}_{j_2} + \mathbf{c}_{i_2}) \\
&\qquad F_1(\mathbf{s}_1, \mathbf{s}_2, \mathbf{s}_3, \mathbf{s}_4)d\mathbf{s}_2 d\mathbf{s}_3 d\mathbf{s}_4 \\
&-\frac{2}{k_n}\sum_{j_2=1}^{k_n}\sum_{j_3=1}^{k_n}\int_{D_{l(n)}^{j_1}}\int_{D_{l(n)}^{j_2}}\int_{D_{l(n)}^{j_3}} Z(\mathbf{s}_2 - \mathbf{c}_{j_1} + \mathbf{c}_{i_1})Z(\mathbf{s}_3 - \mathbf{c}_{j_2} + \mathbf{c}_{i_2})Z(\mathbf{s}_4 - \mathbf{c}_{j_3} + \mathbf{c}_{i_2}) \\
&\qquad F_1(\mathbf{s}_1, \mathbf{s}_2, \mathbf{s}_3, \mathbf{s}_4)d\mathbf{s}_2 d\mathbf{s}_3 d\mathbf{s}_4 \\
&+\frac{1}{k_n^2}\sum_{j_2=1}^{k_n}\sum_{j_3=1}^{k_n}\sum_{j_4=1}^{k_n}\int_{D_{l(n)}^{j_2}}\int_{D_{l(n)}^{j_3}}\int_{D_{l(n)}^{j_4}} Z(\mathbf{s}_2 - \mathbf{c}_{j_2} + \mathbf{c}_{i_1})Z(\mathbf{s}_3 - \mathbf{c}_{j_3} + \mathbf{c}_{i_2})Z(\mathbf{s}_4 - \mathbf{c}_{j_4} + \mathbf{c}_{i_2}) \\
&\qquad F_1(\mathbf{s}_1, \mathbf{s}_2, \mathbf{s}_3, \mathbf{s}_4)d\mathbf{s}_2 d\mathbf{s}_3 d\mathbf{s}_4\Big]d\mathbf{s}_1.
\end{aligned}
$$

Let $F_2(\mathbf{s}_1, \mathbf{s}_2, \mathbf{s}_3, \mathbf{s}_4) = \varphi_4(\mathbf{s}_1, \mathbf{s}_2, \mathbf{s}_3, \mathbf{s}_4) + 2\varphi_3(\mathbf{s}_1, \mathbf{s}_2, \mathbf{s}_3) + 2\varphi_3(\mathbf{s}_1, \mathbf{s}_3, \mathbf{s}_4) + 2\varphi(\mathbf{s}_1 - \mathbf{s}_3)\varphi(\mathbf{s}_2 - \mathbf{s}_4)$. By noting that $Z(\cdot)$ is bounded, we can derive from lengthy yet elementary algebra

that the above is bounded by:

$$\frac{2C(\lambda_n)^4}{k_n|D_n|^2}\sum_{j_1=1}^{k_n}\int_{D_{l(n)}^{j_1}}\int_{D_{l(n)}^{j_1}}\int_{D_n}\int_{D_n}|F_2(\mathbf{s}_1,\mathbf{s}_2,\mathbf{s}_3,\mathbf{s}_4)|d\mathbf{s}_1d\mathbf{s}_2d\mathbf{s}_3d\mathbf{s}_4$$
$$+\ \frac{C(\lambda_n)^4}{|D_n|^2}\sum_{j_1=1}^{k_n}\sum_{j_2=1}^{k_n}\int_{D_{l(n)}^{j_1}}\int_{D_{l(n)}^{j_1}}\int_{D_{l(n)}^{j_2}}\int_{D_{l(n)}^{j_2}}|F_1(\mathbf{s}_1,\mathbf{s}_2,\mathbf{s}_3,\mathbf{s}_4)|d\mathbf{s}_1d\mathbf{s}_2d\mathbf{s}_3d\mathbf{s}_4.$$

The first integral in the above is of order $1/k_n$ due to conditions (4), (5) and (6). The second integral is bounded by

$$\frac{C(\lambda_n)^4}{|D_n|^2}\sum_{j_1=1}^{k_n}\sum_{j_2=1}^{k_n}\int_{D_{l(n)}^{j_1}}\int_{D_{l(n)}^{j_1}}\int_{D_{l(n)}^{j_2}}\int_{D_{l(n)}^{j_2}}|\varphi_4(\mathbf{s}_1,\mathbf{s}_2,\mathbf{s}_3,\mathbf{s}_4)|d\mathbf{s}_1d\mathbf{s}_2d\mathbf{s}_3d\mathbf{s}_4$$
$$+\ \frac{4C(\lambda_n)^4}{k_n|D_n|}\sum_{j_2=1}^{k_n}\int_{D_n}\int_{D_{l(n)}^{j_2}}\int_{D_{l(n)}^{j_2}}|\varphi_3(\mathbf{s}_1,\mathbf{s}_3,\mathbf{s}_4)|d\mathbf{s}_1d\mathbf{s}_3d\mathbf{s}_4$$
$$+\ \frac{2C(\lambda_n)^4}{|D_n|^2}\sum_{j_1=1}^{k_n}\sum_{j_2=1}^{k_n}\Big[\int_{D_{l(n)}^{j_1}}\int_{D_{l(n)}^{j_2}}|\varphi(\mathbf{s}_1,\mathbf{s}_3)|d\mathbf{s}_1d\mathbf{s}_3\Big]^2.$$

The above is of order $1/k_n + a_n$ due to conditions (4), (5), (6) and (12), where the order of a_n is no higher than $1/|D_{l(n)}|$.

Appendix C: Justification for Using $\hat{\beta}_n$ in the Thinning Step

Let $p_n(\mathbf{x};\theta) = \min_{\mathbf{s}\in D_n}\lambda(\mathbf{s};\theta)/\lambda(\mathbf{x};\theta)$ and $r(\mathbf{x})$ be a uniform random variable in $[0,1]$. If $r(\mathbf{x}) \le p_n(\mathbf{x};\theta)$, where $\mathbf{x}\in(N\cap D_n)$, then $\mathbf{x}$ will be retained in the thinned process, say $\Psi(\theta)$. Based on the same set $\{r(\mathbf{x}) : \mathbf{x}\in(N\cap D_n)\}$, we can determine $\Psi(\theta_0)$ and $\Psi(\hat{\theta}_n)$. Note that $\Psi(\theta_0)$ is Ψ defined in (7) using the true FOIF. Define $\Psi_a = \{\mathbf{x}\in\Psi(\theta_0), \mathbf{x}\notin\Psi(\hat{\theta}_n)\}$ and $\Psi_b = \{\mathbf{x}\notin\Psi(\theta_0), \mathbf{x}\in\Psi(\hat{\theta}_n)\}$. Note that $P(\mathbf{x}\in\Psi_a\cup\Psi_b|\mathbf{x}\in N) \le |p_n(\mathbf{x};\hat{\theta}_n) - p_n(\mathbf{x};\theta_0)|$. We need to show that

$$\frac{1}{|D_n|}\Big\{Var[S_n^b|\Psi(\theta_0)\cap D_n] - Var[S_n^b|\Psi(\hat{\theta}_n)\cap D_n]\Big\} \xrightarrow{p} 0,$$

where $Var[S_n^b|\Psi(\theta_0)\cap D_n]$ is $Var(S_n^b|\Psi\cap N)$ defined in Appendix B. $Var[S_n^b|\Psi(\hat{\theta}_n)\cap D_n]$ is defined analogously.

Let $\sum_D$ and $\sum\sum_D^{\neq}$ denote $\sum_{D\cap(\Psi_a\cup\Psi_b)}$ and $\sum\sum_{\mathbf{x}_1,\mathbf{x}_2\in D\cap(\Psi_a\cup\Psi_b),\mathbf{x}_1\neq\mathbf{x}_2}$. Note that

$$\begin{aligned}
& \frac{1}{|D_n|}\Big|Var[S_n^b|\Psi(\theta_0)\cap D_n]-Var[S_n^b|\Psi(\hat{\theta}_n)\cap D_n]\Big| \\
< \; & \frac{C}{|D_n|k_n}\sum_{i=1}^{k_n}\sum_{j=1}^{k_n}\sum_{D_{l(n)}^j}\sum_{D_{l(n)}^j}1+\frac{C}{|D_n|k_n^2}\sum_{i=1}^{k_n}\sum_{j_1=1}^{k_n}\sum_{j_2=1}^{k_n}\sum_{D_{l(n)}^{j_1}}\sum_{D_{l(n)}^{j_2}}1 \\
= \; & \frac{C}{|D_n|}\sum_{j=1}^{k_n}\sum\sum_{D_{l(n)}^j}^{\neq}1+\frac{C}{|D_n|}\sum_{D_n}1+\frac{C}{|D_n|k_n}\Big[\sum_{D_n}1\Big]^2.
\end{aligned}$$

Note that $P[\mathbf{x}\in(\Psi_a\cup\Psi_b)|\mathbf{x}\in N]\leq\frac{C}{|D_n|^{\alpha/2}}$ with probability 1 for $0<\alpha<1$ due to Theorem 1. Thus all three terms in the above converge to zero in probability. This in turn leads to the consistency of the variance estimator.

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. and Turner, R. (2005), "Spatstat: an R Package for Analyzing Spatial Point Patterns", *Journal of Statistical Software*, 12, 1–42.

Brillinger, D. R. (1975), *Time Series: Data Analysis and Theory*, New York: Holt, Rinehart & Winston.

Cressie, N. A. C. (1993), *Statistics for Spatial Data*, New York: Wiley.

Daley, D. J. and Vere-Jones, D. (1988), *An Introduction to the Theory of Point Processes*, New York: Springer-Verlag.

Doukhan, P. (1994), *Mixing: Properties and Examples*, New York: Springer-Verlag.

Guan, Y., Sherman, M. and Calvin, J. A. (2004), "A Nonparametric Test for Spatial Isotropy Using Subsampling", *Journal of the American Statistical Assocation*, 99, 810–821.

Guan, Y., Sherman, M. and Calvin, J. A. (2006), "Assessing Isotropy for Spatial Point Processes", *Biometrics*, 62, 119–125.

Hall, P. and Jing, B. (1996), "On Sample Reuse Methods for Dependent Data", *Journal of the Royal Statistical Society, Series B*, 58, 727–737.

Heinrich, L. (1985), "Normal Convergence of Multidimensional Shot Noise and Rates of This Convergence", *Advances in Applied Probability*, 17, 709–730.

Lahiri, S. N. (2003), *Resampling Methods for Dependent data*, New York: Springer-Verlag.

McElroy, T. and Politis, D. N. (2007), "Stable marked point processes", *Annals of Statistics*, to appear.

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

Politis, D. N. (1999), *Subsampling*, New York: Springer-Verlag.

Rathbun, S. L. and Cressie, N. (1994), "Asymptotic Properties of Estimators for the Parameters of Spatial Inhomogeneous Poisson Point Processes", *Advances in Applied Probability*, 26, 122–154.

Rosenblatt, M. (1956), “A Central Limit Theorem and a Strong Mixing Condition”, *Proceedings of the National Academy of Sciences*, 42, 43–47.

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. (1996), “Variance Estimation for Statistics Computed from Spatial Lattice Data”, *Journal of the Royal Statistical Society, Series B*, 58, 509–523.

Sherman, M. (1997), “Subseries Methods in Regression”, *Journal of the American Statistical Association*, 92, 1041–1048.

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

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

****, *. (2007), “**********”, submitted.

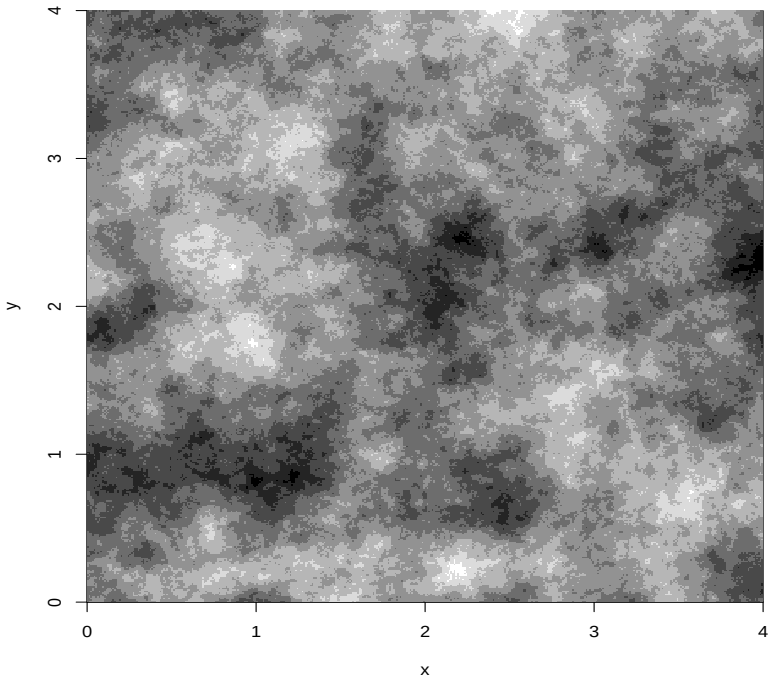

Figure 1: Image of the covariate X used in the simulation. The range of X is between -0.6066 and 0.5303, where darker colors represent larger values of X.

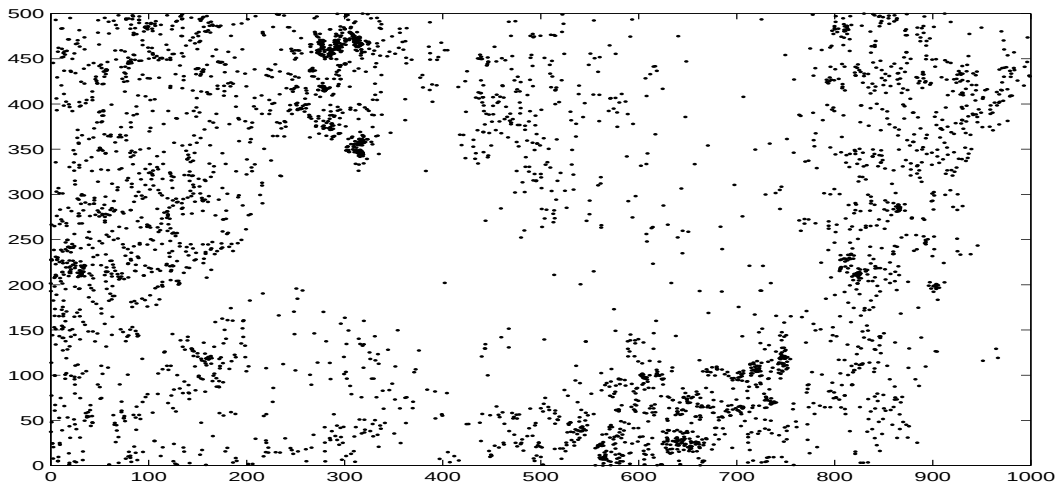

Figure 2: Locations of *Beilschmiedia pendula Lauraceae* trees.

$\rho = .1$	Block size				$\rho = .2$	Block size			
Region size	1/4	1/3	1/2	1	Region size	1/4	1/3	1/2	1
1	.85	.77	.49	-	1	.84	.76	.50	-
2	.91	.90	.88	.49	2	.87	.88	.87	.50
4	.93	.93	.93	.90	4	.88	.89	.90	.89

Table 1: Empirical coverage of 95% nominal confidence intervals of β in the LGCP case.

$\rho = .2$		Block size			
Region size	true SE	1/4	1/3	1/2	1
1	.24	.17 (.04)	.16 (.05)	.13 (.06)	-
2	.12	.10 (.02)	.10 (.02)	.10 (.03)	.08 (.03)
4	.07	.056 (.003)	.057 (.004)	.060 (.006)	.060 (.008)

Table 2: Estimates of the standard errors of $\hat{\beta}$ for $\rho = .2$ in the LGCP case.

$\omega = .02$		Block size			$\omega = .04$		Block size		
Region size	Plug-in	1/4	1/3	1/2	Region size	Plug-in	1/4	1/3	1/2
1	.96	.91	.89	.81	1	.93	.87	.87	.77
2	.96	.94	.94	.94	2	.95	.90	.90	.91

Table 3: Empirical coverage of 95% nominal confidence intervals of β obtained by the thinned block bootstrap method and the plug-in method in the INSP case.

		Plug-in method		Thinned block bootrap	
	EST	STD	95% CI	STD	95% CI
β_1	.02	.02	(-.02,.06)	.017	(-.014,.054)
β_2	5.84	2.53	(.89,10.80)	2.12	(1.69,9.99)

Table 4: Results for for the *Beilschmiedia pendula Lauraceae* data. The abbreviations EST, STD and CI denote the point estimate of the regression parameters by maximizing (1), standard deviation and confidence interval, respectively.