# Fast Block Variance Estimation Procedures for Inhomogeneous Spatial Point Processes 

By<br>YONGTAO GUAN,<br>Division of Biostatistics, Yale University, New Haven CT 06520-8034, U.S.A.<br>yongtao.guan@yale.edu<br>Summary

We introduce two new variance estimation procedures by using non-overlapping and overlapping blocks, respectively. The non-overlapping block (NOB) estimator can be viewed as the limit of the thinned block bootstrap (TBB) estimator recently proposed in Guan and Loh (2007), by letting the number of thinned processes and bootstrap samples therein both increase to infinity. Compared to the latter, the NOB estimator can be obtained much faster since it does not require any thinning or bootstrap steps, and is more stable since it is the limit of the latter by using an infinite number of thinnings and bootstrap samples. The overlapping block estimator further improves the performance of the NOB with a modest increase in computational time. A simulation study demonstrates the superiority of the proposed estimators over the TBB estimator. Some key words: Block Variance Estimator; Inhomogeneous Spatial Poisson Process; Thinning.
Short Title. Block Variance Estimator for Inhomogeneous Point Processes.

## 1. Introduction

Let $N$ be a two-dimensional spatial point process that is observed on a domain of interest $D \subset \mathbb{R}^{2}$. Let $\lambda(s)$ and $\lambda\left(s_{1}, s_{2}\right)$ denote the first- and second-order intensity functions (e.g., Diggle, 2003) of the process, respectively. In this paper, we will focus on a flexible class of second-order reweighted stationary processes (SORWS; Baddeley et al., 2000). Specifically, we assume $\lambda\left(s_{1}, s_{2}\right)=\lambda\left(s_{1}\right) \lambda\left(s_{2}\right) g\left(s_{1}-s_{2}\right)$ for some function $g(\cdot)$, where $g(\cdot)$ is called the pair correlation function (PCF; e.g., Møller and Waagepetersen, 2003). In the special case when $\lambda(s)=\lambda$ for some constant $\lambda>0$ for all $s \in \mathbb{R}^{2}$, then the process is further said to be second-order stationary (SOS).

A common interest in practice is to model the first-order intensity function (FOIF) of the process. For this, we assume that $\lambda(\cdot)$ can be written as a parametric function of some observed covariates associated with the process, where the function is completely determined by a $p \times 1$ vector of unknown regression parameters, $\beta$. We thus rewrite $\lambda(\cdot)$ as $\lambda(\cdot ; \beta)$. Our main goal is to estimate and make inference on $\beta$.

Let $|D|$ denote the area of $D$. To estimate $\beta$, the following Poisson maximum likelihood criterion (Schoenberg, 2004) is often used:

$$
\begin{equation*}
U(\beta)=\frac{1}{|D|} \sum_{x \in D \cap N} \log \lambda(x ; \beta)-\frac{1}{|D|} \int_{D} \lambda(s ; \beta) d s \tag{1}
\end{equation*}
$$

Let $\hat{\beta}$ be the maximizer of (1). Schoenberg (2004) showed that $\hat{\beta}$ is consistent for $\beta$ for a wide class of spatial-temporal point process models, even if the process is not Poisson. Waagepetersen (2007) and Guan and Loh (2007) established asymptotic normality for $\hat{\beta}$ for a class of inhomogeneous Neyman-Scott processes and a class of mixing point processes, respectively.

To make inference on $\beta$, the variance of $\hat{\beta}$ needs to be estimated. Let $D_{n}$ be a sequence of domains that approach to $\mathbb{R}^{2}$ in all directions as $n$ increases, $\hat{\beta}_{n}$ and $\beta_{0}$ be the estimated and the true parameter vectors, respectively, $\lambda^{(i)}(\cdot ; \beta)$ be the $i$ th derivative of $\lambda(\cdot ; \beta)$ with respect to $\beta$, and ${ }^{\top}$ be the matrix transpose operator. Guan and Loh (2007) showed that under some suitable conditions,

$$
\begin{equation*}
\Sigma_{n}=\left|D_{n}\right| \operatorname{Cov}\left(\hat{\beta}_{n}\right) \approx\left|D_{n}\right|\left(A_{n}\right)^{-1} B_{n}\left(A_{n}\right)^{-1} \tag{2}
\end{equation*}
$$

where

$$
\begin{gather*}
A_{n}=\int_{D_{n}} \frac{\lambda^{(1)}\left(s ; \beta_{0}\right)\left[\lambda^{(1)}\left(s ; \beta_{0}\right)\right]^{\top}}{\lambda\left(s ; \beta_{0}\right)} d s  \tag{3}\\
B_{n}=A_{n}+\iint_{D_{n}} \lambda^{(1)}\left(u ; \beta_{0}\right)\left[\lambda^{(1)}\left(v ; \beta_{0}\right)\right]^{\top}[g(u-v)-1] d u d v . \tag{4}
\end{gather*}
$$

From (3), it can be seen that $A_{n}$ depends only on the FOIF and thus can be calculated easily once the FOIF has been estimated. From (4), it can be seen that $B_{n}$, however, depends also on the PCF. Often a parametric model for the PCF is first fitted by using, say a minimum contrast estimation procedure (e.g., Møller and Waagepetersen, 2003),
and the estimated PCF is then plugged back into (4) in order to estimate $B_{n}$. To avoid a parametric assumption on the PCF , which can be restrictive in some applications, Guan and Loh (2007) proposed a thinned block bootstrap (TBB) estimator for $B_{n}$. Their procedure involves repeatedly thinning the original point process and then bootstrapping each thinned realization of the process. The TBB procedure assumes that the process is SORWS but does not require any specific parametric form for the PCF. A drawback of this procedure is that it can be very time consuming due to the repeated thinning and bootstrap steps. The goal of this paper is to propose two alternative variance estimation procedures that are built upon the TBB estimator but can be performed much more quickly. Furthermore, it will become clear that the proposed procedures also outperforms the TBB approach by having both smaller bias and variance.

## 2. Background on the TBB procedure

The TBB procedure makes use of the fact that any SORWS process can be thinned to be SOS by applying proper thinning weights. For example, Guan and Loh (2007) considered the following thinned process:

$$
\begin{equation*}
\Psi_{n}=\left\{x: x \in N, P(x \text { is retained })=\min _{s \in D_{n}} \lambda\left(s ; \beta_{0}\right) / \lambda\left(x ; \beta_{0}\right)\right\} . \tag{5}
\end{equation*}
$$

Clearly $\Psi_{n}$ is SOS on $D_{n}$ since its first- and second-order intensity functions can be respectively written as:

$$
\lambda_{n}=\min _{s \in D_{n}} \lambda\left(s ; \beta_{0}\right) \text { and } \lambda_{2, n}\left(s_{1}, s_{2}\right)=\left(\lambda_{n}\right)^{2} g\left(s_{1}-s_{2}\right) .
$$

For each thinned process, $\Psi_{n}$, Guan and Loh (2007) defined the following statistic:

$$
S_{n}=\sum_{x \in \Psi_{n} \cap D_{n}} \lambda^{(1)}\left(x ; \beta_{0}\right) .
$$

By using the fact that $\Psi_{n}$ is SOS, it can be seen that

$$
\begin{equation*}
\operatorname{Cov}\left(S_{n}\right)=\lambda_{n} \int_{D_{n}} \lambda^{(1)}\left(s ; \beta_{0}\right)\left[\lambda^{(1)}\left(s ; \beta_{0}\right)\right]^{\top} d s+\left(\lambda_{n}\right)^{2}\left(B_{n}-A_{n}\right), \tag{6}
\end{equation*}
$$

where $A_{n}$ and $B_{n}$ are given as in (3) and (4), respectively. In order to estimate $B_{n}$, it is sufficient to estimate the covariance matrix of $S_{n}$. For this Guan and Loh (2007) proposed the following block bootstrap algorithm:

1. Obtain a thinned realization of the process as in (5) with $\beta_{0}$ being replaced by $\hat{\beta}_{n}$.
2. Divide $D_{n}$ into $k_{n}$ nonoverlapping subblocks, $D_{l_{n}}^{i}, i=1, \cdots, k_{n}$, where $l_{n}$ signifies the size of each subblock. For each $D_{l_{n}}^{i}$, let $c_{i}$ denote the "center" of the subblock. For each thinned process, resample with replacement for $B$ times $k_{n}$ subblocks from $D_{l_{n}}^{i}, i=1, \cdots, k_{n}$. For the $b$ th collection of the resampled random subblocks, let $J_{b}$ be the set of $k_{n}$ random indices sampled from $\left\{1, \cdots, k_{n}\right\}$ that are associated with the selected subblocks. Define

$$
S_{n}^{b}=\sum_{i=1}^{k_{n}} \sum_{x \in \Psi_{n} \cap D_{l(n)}^{J_{l}(i)}} \lambda^{(1)}\left(x-c_{J_{b}(i)}+c_{i} ; \hat{\beta}_{n}\right),
$$

Obtain the sample covariance matrix for $S_{n}^{b}, b=1, \cdots, B$.
3. Repeat Steps 1 and 2 for $M$ times and use the average of the resulting sample covariance matrices as the estimate for the covariance matrix given in (6).

Let $\widehat{\operatorname{Cov}\left(S_{n}\right)}$ be the obtained estimator from the above algorithm. Then (6) implies the following estimator for $B_{n}$ :

$$
\begin{equation*}
\left.\hat{B}_{n}=\widehat{\operatorname{Cov}\left(S_{n}\right.}\right) / \hat{\lambda}_{n}^{2}-\int_{D_{n}} \lambda^{(1)}\left(s ; \hat{\beta}_{n}\right)\left[\lambda^{(1)}\left(s ; \hat{\beta}_{n}\right)\right]^{\top} d s / \hat{\lambda}_{n}+\hat{A}_{n} \tag{7}
\end{equation*}
$$

where

$$
\begin{gathered}
\hat{\lambda}_{n}=\min _{s \in D_{n}} \lambda\left(s ; \hat{\beta}_{n}\right), \\
\hat{A}_{n}=\int_{D_{n}} \frac{\lambda^{(1)}\left(s ; \hat{\beta}_{n}\right)\left[\lambda^{(1)}\left(s ; \hat{\beta}_{n}\right)\right]^{\top}}{\lambda\left(s ; \hat{\beta}_{n}\right)} d s .
\end{gathered}
$$

## 3. The Proposed Variance Estimation Procedures

## 3•1 The algorithms

Throughout this section, let $\lambda(\cdot)$ and $\lambda^{(1)}(\cdot)$ denote $\lambda\left(\cdot ; \hat{\beta}_{n}\right)$ and $\lambda^{(1)}\left(\cdot ; \hat{\beta}_{n}\right)$, respectively. Let $\Psi_{n}^{m}$ denote the $m$ th thinned realization of the process. For a fixed $M$ and by
letting $B \rightarrow \infty$, some simple derivations suggest that $\widehat{\operatorname{Cov}\left(S_{n}\right)}$ converges to

$$
\begin{aligned}
& \operatorname{Cov}\left(S_{n}^{b} \mid \Psi_{n} \cap D_{n}\right) \\
= & \frac{1}{k_{n} M} \sum_{m=1}^{M} \sum_{i=1}^{k_{n}} \sum_{j=1}^{k_{n}} \sum_{x \in D_{l_{n}}^{j} \cap \Psi_{n}^{m}} \sum_{y \in D_{l_{n}}^{j} \cap \Psi_{n}^{m}} \lambda^{(1)}\left(x-c_{j}+c_{i}\right)\left[\lambda^{(1)}\left(y-c_{j}+c_{i}\right)\right]^{\top} \\
- & \frac{1}{k_{n}^{2}} \sum_{i=1}^{k_{n}} \sum_{j_{1}=1}^{k_{n}} \sum_{j_{2}=1}^{k_{n}} \sum_{x \in D_{l_{n}}^{j_{1}} \cap \Psi_{n}^{m}} \sum_{y \in D_{l_{n}}^{j_{2}} \cap \Psi_{n}^{m}} \lambda^{(1)}\left(x-c_{j_{1}}+c_{i}\right)\left[\lambda^{(1)}\left(y-c_{j_{2}}+c_{i}\right)\right]^{\top} .
\end{aligned}
$$

Let $\sum_{x} \sum_{y}^{\neq}$denote summation over $x$ and $y$ such that $x \neq y$. Now let $M \rightarrow \infty$, then the above further converges to

$$
\begin{equation*}
\mathbb{E}\left[\operatorname{Cov}\left(S_{n}^{b} \mid \Psi_{n} \cap D_{n}\right) \mid N\right]=\lambda_{n}^{2} \hat{V}_{n, 1}-\lambda_{n}^{2} \hat{V}_{n, 2}+\lambda_{n} \hat{V}_{n, 3}-\lambda_{n} \hat{V}_{n, 3} / k_{n} \tag{8}
\end{equation*}
$$

where

$$
\begin{gather*}
\hat{V}_{n, 1}=\frac{1}{k_{n}} \sum_{i=1}^{k_{n}} \sum_{j=1}^{k_{n}} \sum_{x \in D_{l_{n}}^{j} \cap N} \sum_{y \in D_{l_{n}}^{j} \cap N}^{\neq} \frac{\lambda^{(1)}\left(x-c_{j}+c_{i}\right)\left[\lambda^{(1)}\left(y-c_{j}+c_{i}\right)\right]^{\top}}{\lambda(x) \lambda(y)},  \tag{9}\\
\hat{V}_{n, 2}=\frac{1}{k_{n}^{2}} \sum_{i=1}^{k_{n}} \sum_{j_{1}=1}^{k_{n}} \sum_{j_{2}=1}^{k_{n}} \sum_{x \in D_{l_{n}}^{j_{1}} \cap N} \sum_{y \in D_{l_{n}}^{j_{2}} \cap N}^{\neq} \frac{\lambda^{(1)}\left(x-c_{j_{1}}+c_{i}\right)\left[\lambda^{(1)}\left(y-c_{j_{2}}+c_{i}\right)\right]^{\top}}{\lambda(x) \lambda(y)},  \tag{10}\\
\hat{V}_{n, 3}=\frac{1}{k_{n}} \sum_{i=1}^{k_{n}} \sum_{j=1}^{k_{n}} \sum_{x \in D_{l_{n}}^{j} \cap N} \frac{\lambda^{(1)}\left(x-c_{j}+c_{i}\right)\left[\lambda^{(1)}\left(x-c_{j}+c_{i}\right)\right]^{\top}}{\lambda(x)} . \tag{11}
\end{gather*}
$$

Note that $\mathbb{E}\left(\hat{V}_{n, 3}\right) \approx \int_{D_{n}} \lambda^{(1)}(s)\left[\lambda^{(1)}(s)\right]^{\top} d s$ if $\hat{\beta}_{n} \approx \beta_{0}$. In view of (7)-(11), it is natural to consider the following estimator for $B_{n}$ :

$$
\begin{equation*}
\hat{B}_{n}=\hat{V}_{n, 1}-\hat{V}_{n, 2}+\hat{A}_{n} \tag{12}
\end{equation*}
$$

where $\hat{V}_{n, 1}$ and $\hat{V}_{n, 2}$ are defined as in (9) and (10), respectively. Note that the term $-\lambda_{n} \hat{V}_{n, 3} / k_{n}$ in (8) is not included in (12) since it's ignorable due to $k_{n} \rightarrow \infty . V_{n, 1}$ and $\hat{V}_{n, 2}$ here can be calculated directly without incurring any thinning or bootstrapping steps. Therefore, significant computational gains can be achieved. See Appendix A for the computational details. Furthermore, note that the new estimator can be intuitively regarded as the limiting version of the TBB estimator by letting both $B$ and $M$ increase
to infinity. Thus it is reasonable to expect (12) to be more stable than any TBB estimator using a fixed $B$ and/or $M$.

From (9) and (10), it can be seen that $\hat{V}_{n, 1}$ and $\hat{V}_{n, 2}$ are both defined in terms of the point process $N$ observed on the nonoverlapping blocks $D_{l_{n}}^{i}, i=1, \cdots, k_{n}$. A direct extension of the above procedure is to use overlapping blocks. For this let $D_{l_{n}}$ be the subblock centered at the origin and let $D_{n}^{*}=\left\{s: D_{l_{n}}+s \subset D_{n}\right\}$. Define $D_{l_{n}}^{s}=D_{l_{n}}+s$. The new versions of (9) and (10) using overlapping blocks are given as follows:

$$
\begin{gather*}
\hat{V}_{n, 1}=\frac{1}{\left|D_{n}^{*}\right|} \sum_{i=1}^{k_{n}} \int_{D_{n}^{*}} \sum_{x \in D_{l_{n}}^{s} \cap N} \sum_{y \in D_{l_{n}}^{s} \cap N}^{\neq} \frac{\lambda^{(1)}\left(x-c_{j}+c_{i}\right)\left[\lambda^{(1)}\left(y-c_{j}+c_{i}\right)\right]^{\top}}{\lambda(x) \lambda(y)} \mathrm{d} s,  \tag{13}\\
\hat{V}_{n, 2}=\frac{1}{\left|D_{n}^{*}\right|} \sum_{i=1}^{k_{n}} \int_{D_{l_{n}}^{u}} \int_{D_{l_{n}}^{v}} \sum_{x \in D_{l_{n}}^{u} \cap N} \sum_{y \in D_{l_{n}}^{v} \cap N}^{\neq} \frac{\lambda^{(1)}\left(x-u+c_{i}\right)\left[\lambda^{(1)}\left(y-v+c_{i}\right)\right]^{\top}}{\lambda(x) \lambda(y)} \mathrm{d} u \mathrm{~d} v . \tag{14}
\end{gather*}
$$

Compared to (9) and (10), (13) and (14) utilize more information from the data. Intuitively, we would expect them to be more stable than their counterparts, (9) and (10), which are both based on nonoverlapping blocks. Indeed, overlapping blocks have been found to yield improved variance estimators for block bootstrap (e.g., Künsch, 1989). A similar result is anticipated to be true in the current setting.

To calculate (13) and (14), we need to approximate the integral terms involved. One obvious approach is to "tile" the region $D_{n}^{*}$ by a grid consisting of $k_{n}^{*}$ "small" cells, and then approximate the integrals by the corresponding Riemann sums (e.g., Politis and Sherman, 2001). This leads to the following estimates for $\hat{V}_{n, 1}$ and $\hat{V}_{n, 2}$ :

$$
\begin{gather*}
\hat{V}_{n, 1}=\frac{1}{k_{n}^{*}} \sum_{i=1}^{k_{n}} \sum_{j=1}^{k_{n}^{*}} \sum_{x \in D_{l_{n}}^{j} \cap N} \sum_{y \in D_{l_{n}}^{j} \cap N}^{\neq} \frac{\lambda^{(1)}\left(x-c_{j}+c_{i}\right)\left[\lambda^{(1)}\left(y-c_{j}+c_{i}\right)\right]^{\top}}{\lambda(x) \lambda(y)},  \tag{15}\\
\hat{V}_{n, 2}=\frac{1}{\left(k_{n}^{*}\right)^{2}} \sum_{i=1}^{k_{n}} \sum_{j_{1}=1}^{k_{n}^{*}} \sum_{j_{2}=1}^{k_{n}^{*}} \sum_{x \in D_{l_{n}}^{j_{1}} \cap N} \sum_{y \in D_{l_{n}}^{j_{2}} \cap N}^{\neq} \frac{\lambda^{(1)}\left(x-c_{j_{1}}+c_{i}\right)\left[\lambda^{(1)}\left(y-c_{j_{2}}+c_{i}\right)\right]^{\top}}{\lambda(x) \lambda(y)} . \tag{16}
\end{gather*}
$$

Note that (9) and (10) can be viewed as special cases of (15) and (16) with $k_{n}^{*}=k_{n}$, respectively.

## 3•2 Theoretical Justifications

Let $\hat{V}_{n}=\hat{V}_{n, 1}-\hat{V}_{n, 2}$, where $\hat{V}_{n, 1}$ and $\hat{V}_{n, 2}$ are given as in (9) and (10) in the nonoverlapping case or (13) and (14) in the overlapping case. We would like to show that $\hat{V}_{n}$ converges to

$$
V_{n}=B_{n}-A_{n}=\sum_{i=1}^{k_{n}} \sum_{j=1}^{k_{n}} \int_{D_{l_{n}}^{i}} \int_{D_{l_{n}}^{j}} \lambda^{(1)}\left(u ; \beta_{0}\right) \lambda^{(1)}\left(v ; \beta_{0}\right)[g(u-v)-1] d u d v .
$$

To show this, assume that for all $\beta$ in a small neighborhood of $\beta_{0}$,

$$
\begin{equation*}
\lambda(s ; \beta)>0, \lambda^{(i)}(s ; \beta)<\infty, i=1,2 . \tag{17}
\end{equation*}
$$

In addition, conditions on the cumulant function of the process are needed. Let $\mathrm{d} s$ be an infinitesimal region containing $s$ and $N(\mathrm{~d} s)$ be the number of events from $N$ contained in $\mathrm{d} s$. Define the $k$ th-order cumulant functions of $N$ as:

$$
Q_{k}\left(s_{1}, \cdots, s_{k}\right)=\lim _{\left|\mathrm{d} s_{i}\right| \rightarrow 0}\left\{\frac{\operatorname{Cum}\left[N\left(\mathrm{~d} s_{1}\right), \cdots, N\left(\mathrm{~d} s_{k}\right)\right]}{\left|\mathrm{d} s_{1}\right| \cdots\left|\mathrm{d} s_{k}\right|}\right\}, \quad i=1, \cdots, k,
$$

where $\operatorname{Cum}\left(Y_{1}, \cdots, Y_{k}\right)$ is the coefficient of $i^{k} t_{1} \cdots t_{k}$ in the Taylor series expansion of $\log \left\{\mathrm{E}\left[\exp \left(i \sum_{j=1}^{k} Y_{j} t_{j}\right)\right]\right\}$ about the origin (e.g. Brillinger, 1975) and $Y_{i}, i=1, \cdots, k$ are a set of random variables. The cumulant functions are a useful tool to describe the dependence between events of the process, where close-to-zero values of the cumulant functions often indicate near independence. In the extreme case of complete independence, i.e., when $N$ is Poisson, then $Q_{k}\left(s_{1}, \cdots, s_{k}\right)=0$ if at least two of $s_{1}, \cdots, s_{k}$ are different. In terms of the cumulant functions, assume

$$
\begin{equation*}
\sup _{s_{1}} \int \cdots \int\left|Q_{k}\left(s_{1}, \cdots, s_{k}\right)\right| \mathrm{d} s_{2} \cdots \mathrm{~d} s_{k}<\infty \text { for } k=2,3,4 \tag{18}
\end{equation*}
$$

Condition (18) is a fairly weak condition. It holds for broad class of inhomogeneous models including, but not limited to, the log Gaussian Cox Process (Møller et al., 1998), the inhomogeneous Neyman-Scott process (Waagepetersen, 2007), and any inhomogeneous process that is obtained by thinning a homogeneous process satisfying this condition. The following theorem establishes the consistency of $\hat{V}_{n}$.

Theorem 1. Assume Conditions (17) and (18). If $\left|D_{l_{n}}\right|=\mathrm{o}\left(\left|D_{n}\right|^{1 / 2}\right)$ and $\left|D_{n}\right|^{1 / 2}\left(\hat{\beta}_{n}-\right.$ $\left.\beta_{0}\right)=\mathrm{O}_{p}(1)$, then $\left(\hat{V}_{n}-V_{n}\right) /\left|D_{n}\right| \xrightarrow{L_{2}} 0$.

Proof. The proof of the theorem can be found in Appendix B.
The condition on the convergence rate of $\hat{\beta}_{n}$ to $\beta_{0}$ is a standard assumption and holds under conditions given Guan and Loh (2007). The condition on the subblock size is undesirable. However, a further relaxation of this condition appears to be difficult. It may not be reasonable after all to set the subblock size to be of order larger than or equal to $\left|D_{n}\right|^{1 / 2}$ given $\left|D_{n}\right|^{1 / 2}\left(\hat{\beta}_{n}-\beta_{0}\right)=\mathrm{O}_{p}(1)$. Finally, note that although we are focusing on estimating the variance of $\hat{\beta}_{n}$ in this paper, the proposed methods can be applied to estimate the variance for any statistic taking the form $\sum_{x \in N \cap D_{n}} Z\left(x ; \beta_{0}\right)$. This can be done simply by replacing $\lambda^{(1)}(\cdot)$ with $Z(\cdot) \lambda(\cdot)$ in the definitions of $\hat{V}_{n}$.

## 4. A Simulation Study

To illustrate the superior performance of the proposed methods over the TBB approach, we simulated realizations from an inhomogeneous Neyman-Scott process model on a unit square. The FOIF of the process was $\lambda(s)=\alpha+\beta X(s)$, where $\alpha=7.02$, $\beta=2$ and $X(s)$ was the same covariate process as being used in Guan and Loh (2007). For each simulation, we first simulated a homogeneous Poisson process as the parent process, where the intensity of the process $\kappa=50$. For each parent, we then generated a Poisson number of offspring. The position of each offspring relative to its parent was determined by a radially symmetric Gaussian random variable (e.g. Diggle, 2003) with a standard deviation $\omega=0.02,0.04$, which represent relatively strong and weak clustering, respectively.

One thousand realizations of the process were simulated for each $\omega$ value. For each realization, the TBB estimator as well as the proposed methods were all applied. For the TBB estimator, $B=499,999$ and $M=5,10,20$. The different $B$ and $M$ values allowed us to assess the sensitivity of the performance of the TBB estimator to these parameters. For all estimators, the subblock size was $0.25 \times 0.25$, which led to $k_{n}=16$ nonoverlapping blocks. For the estimator using overlapping blocks, $k_{n}^{*}=64,144$, where $k_{n}^{*}$ was the number of small cells used to calculate (15) and (16). For the convenience of discussion, let NOB and OB denote the nonoverlapping and overlapping block estimators, respectively.

Table 1 shows the bias, standard deviation (STD) and computational time (CT) for
each estimator. For all estimators, there is a negative bias. The bias for $\omega=0.04$ is larger, which is likely due to the fact that the range of dependence is larger in this case. The NOB and OB appear to be less biased than TBB. This is due to the removal of the term $-\lambda_{n} \hat{V}_{n, 3} / k_{n}$ in (12), which reduces the bias. For TBB, the STD generally decreases with $M$ (or $B$ ) for a fixed $B$ (or $M$ ). The effect of $B$ is not very significant in this example, likely because both $B$ values being considered here are much larger than the number of blocks to be resampled from. Note all the STDs for TBB are larger than their counterparts for NOB. This is as expected. For OB, the STDs are significantly smaller than those for NOB. This provides evidence for the importance of using overlapping blocks. The number of overlapping blocks, however, does not need to extremely big to achieve the most benefit. In terms of CT, NOB is the most computationally efficient. The CT for NOB for 1,000 simulations is only around $1 / 18$ of that for OB when $k_{n}^{*}=64$, and around $1 / 69$ of the smallest CT for TBB (when $M=5$ and $B=499$ ). For OB, the CT for $k_{n}^{*}=64$ is only $1 / 4$ of the smallest CT for TBB. The CT quickly increases when $k_{n}^{*}=144$. Even in this case, it is only slightly larger than the smallest CT for TBB. The CT for TBB increases roughly in proportion with $M$ and $B$. In summary, the proposed new methods greatly outperform the TBB approach in both accuracy and CT.

## Appendix A

## Computational Details

For $\hat{V}_{n, 1}$ given in (15), it can be seen that $\hat{V}_{n, 1}=\hat{V}_{n, 1}^{a}-\hat{V}_{n, 1}^{b}$, where

$$
\begin{gathered}
\hat{V}_{n, 1}^{a}=\frac{1}{k_{n}^{*}} \sum_{i=1}^{k_{n}} \sum_{j=1}^{k_{n}^{*}}\left[\sum_{x \in D_{l_{n}}^{j} \cap N} \frac{\lambda^{(1)}\left(x-c_{j}+c_{i}\right)}{\lambda(x)}\right]\left[\sum_{x \in D_{l_{n}}^{j} \cap N} \frac{\lambda^{(1)}\left(x-c_{j}+c_{i}\right)}{\lambda(x)}\right]^{\top}, \\
\hat{V}_{n, 1}^{b}=\frac{1}{k_{n}^{*}} \sum_{i=1}^{k_{n}} \sum_{j=1}^{k_{n}^{*}} \sum_{x \in D_{l_{n}}^{j} \cap N} \frac{\lambda^{(1)}\left(x-c_{j}+c_{i}\right)\left[\lambda^{(1)}\left(x-c_{j}+c_{i}\right)\right]^{\top}}{\lambda(x)^{2}} .
\end{gathered}
$$

For $\hat{V}_{n, 2}$ given in (16), it can be seen that $\hat{V}_{n, 2}=\hat{V}_{n, 2}^{a}-\hat{V}_{n, 2}^{b}$, where

$$
\hat{V}_{n, 2}^{a}=\frac{1}{\left(k_{n}^{*}\right)^{2}} \sum_{i=1}^{k_{n}}\left[\sum_{j=1}^{k_{n}^{*}} \sum_{x \in D_{l_{n}}^{j} \cap N} \frac{\lambda^{(1)}\left(x-c_{j_{1}}+c_{i}\right)}{\lambda(x)}\right]\left[\sum_{j=1}^{k_{n}^{*}} \sum_{x \in D_{l_{n}}^{j} \cap N} \frac{\lambda^{(1)}\left(x-c_{j_{1}}+c_{i}\right)}{\lambda(x)}\right]^{\top},
$$

$$
\hat{V}_{n, 2}^{b}=\frac{1}{\left(k_{n}^{*}\right)^{2}} \sum_{i=1}^{k_{n}} \sum_{j_{1}=1}^{k_{n}^{*}} \sum_{j_{2}=1}^{k_{n}^{*}} \sum_{x \in D_{l_{n}}^{j_{1}} \cap D_{l_{n}}^{j_{2} \cap N}} \frac{\lambda^{(1)}\left(x-c_{j_{1}}+c_{i}\right)\left[\lambda^{(1)}\left(x-c_{j_{2}}+c_{i}\right)\right]^{\top}}{\lambda(x)^{2}} .
$$

$\hat{V}_{n, 1}$ and $\hat{V}_{n, 2}$ in the nonoverlapping case can be obtained easily by noting that $k_{n}^{*}=k_{n}$.

## Appendix B

## Proof of Theorem 1

By a direct application of Taylor series expansion, it can be seen that it's sufficient to prove the theorem for $\hat{\beta}_{n}=\beta_{0}$. Here we only outline the proof for the overlapping block estimator. The proof in the nonoverlapping case follows trivially. Specifically, we consider the case when $\hat{V}_{n, 1}$ and $\hat{V}_{n, 2}$ are given by (15) and (16), respectively, since they, but not (13) and (14), are used in practice. To do so, first note that

$$
\begin{aligned}
\mathbb{E}\left(\hat{V}_{n}\right) & =\sum_{i=1}^{k_{n}} \iint_{D_{l_{n}}^{i}} \lambda^{(1)}(u) \lambda^{(1)}(v)[g(u-v)-1] \mathrm{d} u \mathrm{~d} v \\
& -\frac{1}{\left(k_{n}^{*}\right)^{2}} \sum_{i=1}^{k_{n}} \sum_{j_{1}=1}^{k_{n}^{*}=1} \sum_{j_{2}=1}^{k_{n}^{*}} \iint_{D_{l_{n}}^{*}} \lambda^{(1)}(u) \lambda^{(1)}(v)\left[g\left(u-v+c_{j_{1}}-c_{j_{2}}\right)-1\right] \mathrm{d} u \mathrm{~d} v .
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
\mathbb{E}\left(\hat{V}_{n}\right)-V_{n} & =-\sum_{i=1}^{k_{n}} \sum_{j \neq i} \int_{D_{l_{n}}^{i}} \int_{D_{l_{n}}^{j}} \lambda^{(1)}(u) \lambda^{(1)}(v)[g(u-v)-1] \mathrm{d} u \mathrm{~d} v \\
& -\frac{1}{\left(k_{n}^{*}\right)^{2}} \sum_{i=1}^{k_{n}} \sum_{j_{1}=1}^{k_{n}^{*}} \sum_{j_{2}=1}^{k_{n}^{*}} \iint_{D_{l_{n}}^{i}} \lambda^{(1)}(u) \lambda^{(1)}(v)\left[g\left(u-v+c_{j_{1}}-c_{j_{2}}\right)-1\right] \mathrm{d} u \mathrm{~d} v .
\end{aligned}
$$

The first term in the above is of order o $\left(\left|D_{n}\right|\right)$ from the proof of Theorem 2 of Guan and Loh (2007). The second term is also of order $\mathrm{o}\left(\left|D_{n}\right|\right)$ due to (17) and (18) and the fact that for each fixed $j_{1}$, the number of overlapping blocks that are within $l_{n}$ distance of $D_{l_{n}}^{j_{1}}$ is of order $\mathrm{o}\left(k_{n}^{*} / k_{n}\right)$.

For the variance of $\hat{V}_{n}$, some tedious yet elementary algebra suggests the variance is bounded by the following terms:

$$
\frac{C k_{n}^{2}}{\left(k_{n}^{*}\right)^{2}} \sum_{j_{1}=1}^{k_{n}^{*}} \sum_{j_{2}=1}^{k_{n}^{*}} \int_{D_{l_{n}}^{j_{1}} \cap D_{l_{n}}^{j_{2}}} \int_{D_{l_{n}}^{j_{1}} \cap D_{l_{n}}^{j_{2}}} g\left(x_{1}-x_{2}\right) \mathrm{d} x_{1} \mathrm{~d} x_{2}
$$

$$
\begin{gathered}
\frac{C k_{n}^{2}}{\left(k_{n}^{*}\right)^{2}} \sum_{j_{1}=1}^{k_{n}^{*}} \sum_{j_{2}=1}^{k_{n}^{*}} \int_{D_{l_{n}}^{j_{1}} \cap D_{l_{n}}^{j_{2}}} \int_{D_{l_{n}}^{j_{1}}} \int_{D_{l_{n}}^{j_{2}}}\left|Q_{3}\left(x_{1}, x_{2}, x_{3}\right)+2 Q_{2}\left(x_{1}, x_{2}\right)+Q_{2}\left(x_{2}, x_{3}\right)\right| \mathrm{d} x_{1} \mathrm{~d} x_{2} \mathrm{~d} x_{3}, \\
\frac{C k_{n}^{2}}{\left(k_{n}^{*}\right)^{2}} \sum_{j_{1}=1}^{k_{n}^{*}} \sum_{j_{2}=1}^{k_{n}^{*}} \int_{D_{l_{n}}^{j_{1}}} \int_{D_{l_{n}}^{j_{1}}} \int_{D_{l_{n}}^{j_{2}}} \int_{D_{l_{n}}^{j_{2}}}\left|Q_{3}\left(x_{1}, x_{2}, x_{3}\right)\right| \mathrm{d} x_{1} \mathrm{~d} x_{2} \mathrm{~d} x_{3} \mathrm{~d} x_{4}, \\
\frac{C k_{n}^{2}}{\left(k_{n}^{*}\right)^{2}} \sum_{j_{1}=1}^{k_{n}^{*}} \sum_{j_{2}=1}^{k_{n}^{*}}\left[\int_{D_{l_{n}}^{j_{1}}} \int_{D_{l_{n}}^{j_{2}}}\left|Q_{2}\left(x_{1}, x_{2}\right)\right| \mathrm{d} x_{1} \mathrm{~d} x_{2}\right]^{2} .
\end{gathered}
$$

All the above terms are of order $\left|D_{n}\right|^{2} / k_{n}$ due to (17) and (18) and the fact that for each fixed $j_{1}$, the number of overlapping blocks that are within $l_{n}$ distance of $D_{l_{n}}^{j_{1}}$ is of order $\mathrm{o}\left(k_{n}^{*} / k_{n}\right)$. Thus Theorem 1 is proved.

## REFERENCES

Baddeley, A. J., Møller, J. \& Waagepetersen, R. P. (2000). Non- and semiParametric estimation of interaction in inhomogeneous point patterns. Statistica Neerlandica, 54, 329-350.

Brillinger, D. R. (1975). Time series: data analysis and theory. Holt, Rinehart \& Winston: New York.

Diggle, P. J. (2003). Statistical analysis of spatial point patterns. Oxford University Press: New York.

Guan, Y. \& Loh, J. M. (2007). A thinned block bootstrap variance estimation procedure for inhomogeneous spatial Point Patterns. Journal of the American Statistical Association. To appear.

Künsch, H. R. (1993). The jackknife and the bootstrap for general stationary observations. Annals of Statistics, 17, 1217-1241.

Møller, J., Syversveen, A. R. and Waagepetersen, R. P. (1998). Log Gaussian Cox processes. Scandinavian Journal of Statistics, 25, 451-482.

Møller, J. \& Waagepetersen, R. P. (2003). Statistical inference and simulation for spatial point processes. Chapman \& Hall: New York.

Politis, D. N. \& Sherman, M. (2001). Moment estimation for statistics from marked point processes. Journal of the Royal Statistical Society, Series B, 63, 261-275.

Schoenberg, F. P. (2004). Consistent parametric estimation of the intensity of a spatial-temporal point process. Journal of Statistical Planning and Inference, 128, 79-93.

Wafgepetersen, R. P. (2007). An estimating function approach to inference for inhomogeneous Neyman-Scott processes. Biometrics. 62, 252-258.

Table 1: Bias, standard deviation (STD) and computational time for 1,000 simulations (in hours) for the various variance estimators using nonoverlapping blocks (NOB), overlapping blocks ( OB ) and the thinned block bootstrap (TBB). The computational time are in hours of the cpu time.

|  |  |  | OB |  | TBB $(B=499)$ |  |  | TBB $(B=999)$ |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | $\omega$ | NOB | $k_{n}^{*}=64$ | 144 | $M=5$ | 10 | 20 | $M=5$ | 10 | 20 |
| BIAS | 0.02 | -.1233 | -.1193 | -.1209 | -.1271 | -.1285 | -.1264 | -.1286 | -.1281 | -.1273 |
|  | 0.04 | -.1648 | -.1652 | -.1671 | -.1711 | -.1696 | -.1699 | -.1709 | -.1712 | -.1700 |
| STD | 0.02 | .1310 | .1197 | .1193 | .1340 | .1327 | .1330 | .1339 | .1327 | .1326 |
|  | 0.04 | .1216 | .1114 | .1124 | .1256 | .1240 | .1229 | .1241 | .1241 | .1228 |
| TIME | 0.02 | .0117 | .2076 | .8745 | .8041 | 1.5229 | 2.9513 | 1.4951 | 2.9204 | 5.7367 |
|  | 0.04 | .0118 | .2237 | .9487 | .8203 | 1.5418 | 2.9655 | 1.5096 | 2.9499 | 5.7571 |

