Orthogonal series estimation of the pair correlation function of a spatial point process

Abdollah Jalilian[1], Yongtao Guan[2], and Rasmus Waagepetersen[3]

[1]Department of Statistics, Razi University,Bagh-e-Abrisham, Kermanshah 67149-67346, Iran jalilian@razi.ac.ir
[2]Department of Management Science, University of Miami,Coral Gables, Florida 33124-6544, U.S.A. yguan@bus.miami.edu
[3]Department of Mathematical Sciences, Aalborg University,Fredrik Bajersvej 7G, DK-9220 Aalborg, Denmark rw@math.aau.dk

February 6, 2017

Abstract

The pair correlation function is a fundamental spatial point process characteristic that, given the intensity function, determines second order moments of the point process. Non-parametric estimation of the pair correlation function is a typical initial step of a statistical analysis of a spatial point pattern. Kernel estimators are popular but especially for clustered point patterns suffer from bias for small spatial lags. In this paper we introduce a new orthogonal series estimator. The new estimator is consistent and asymptotically normal according to our theoretical and simulation results. Our simulations further show that the new estimator can outperform the kernel estimators in particular for Poisson and clustered point processes.

Keywords: Asymptotic normality; Consistency; Kernel estimator; Orthogonal series estimator; Pair correlation function; Spatial point process.

1 Introduction

The pair correlation function is commonly considered the most informative second-order summary statistic of a spatial point process (Stoyan and Stoyan, 1994; Møller and Waagepetersen, 2003; Illian et al., 2008). Non-parametric estimates of the pair correlation function are useful for assessing regularity or clustering of a spatial point pattern and can moreover be used for inferring parametric models for spatial point processes via minimum contrast estimation (Stoyan and Stoyan, 1996; Illian et al., 2008). Although alternatives exist (Yue and Loh, 2013), kernel estimation is the by far most popular approach (Stoyan and Stoyan, 1994; Møller and Waagepetersen, 2003; Illian et al., 2008) which is closely related to kernel estimation of probability densities.

Kernel estimation is computationally fast and works well except at small spatial lags. For spatial lags close to zero, kernel estimators suffer from strong bias, see e.g. the discussion at page 186 in Stoyan and Stoyan (1994), Example 4.7 in Møller and

Waagepetersen (2003) and Section 7.6.2 in Baddeley et al. (2015). The bias is a major drawback if one attempts to infer a parametric model from the non-parametric estimate since the behavior near zero is important for determining the right parametric model (Jalilian et al., 2013).

In this paper we adapt orthogonal series density estimators (see e.g. the reviews in Hall, 1987; Efromovich, 2010) to the estimation of the pair correlation function. We derive unbiased estimators of the coefficients in an orthogonal series expansion of the pair correlation function and propose a criterion for choosing a certain optimal smoothing scheme. In the literature on orthogonal series estimation of probability densities, the data are usually assumed to consist of indendent observations from the unknown target density. In our case the situation is more complicated as the data used for estimation consist of spatial lags between observed pairs of points. These lags are neither independent nor identically distributed and the sample of lags is biased due to edge effects. We establish consistency and asymptotic normality of our new orthogonal series estimator and study its performance in a simulation study and an application to a tropical rain forest data set.

2 Background

2.1 Spatial point processes

We denote by X a point process on $\mathbb{R}^d$, $d \geq 1$, that is, X is a locally finite random subset of $\mathbb{R}^d$. For $B \subseteq \mathbb{R}^d$, we let $N(B)$ denote the random number of points in $X \cap B$. That X is locally finite means that $N(B)$ is finite almost surely whenever B is bounded. We assume that X has an intensity function ρ and a second-order joint intensity $\rho^{(2)}$ so that for bounded $A, B \subset \mathbb{R}^d$,

$$E\{N(B)\} = \int_B \rho(u)\mathrm{d}u, \quad E\{N(A)N(B)\} = \int_{A\cap B} \rho(u)\mathrm{d}u + \int_A \int_B \rho^{(2)}(u,v)\mathrm{d}u\mathrm{d}v. \tag{1}$$

The pair correlation function g is defined as $g(u,v) = \rho^{(2)}(u,v)/\{\rho(u)\rho(v)\}$ whenever $\rho(u)\rho(v) > 0$ (otherwise we define $g(u,v) = 0$). By (1),

$$\mathrm{cov}\{N(A), N(B)\} = \int_{A\cap B} \rho(u)\mathrm{d}u + \int_A \int_B \rho(u)\rho(v)\{g(v,u) - 1\}\mathrm{d}u\mathrm{d}v$$

for bounded $A, B \subset \mathbb{R}^d$. Hence, given the intensity function, g determines the covariances of count variables $N(A)$ and $N(B)$. Further, for locations $u, v \in \mathbb{R}^d$, $g(u,v) > 1$ (< 1) implies that the presence of a point at v yields an elevated (decreased) probability of observing yet another point in a small neighbourhood of u (e.g. Coeurjolly et al., 2016). In this paper we assume that g is isotropic, i.e. with an abuse of notation, $g(u,v) = g(\|v-u\|)$. Examples of pair correlation functions are shown in Figure 1.

2.2 Kernel estimation of the pair correlation function

Suppose X is observed within a bounded observation window $W \subset \mathbb{R}^d$ and let $X_W = X \cap W$. Let $k_b(\cdot)$ be a kernel of the form $k_b(r) = k(r/b)/b$, where k is a probability density and $b > 0$ is the bandwidth. Then a kernel density estimator (Stoyan and

Stoyan, 1994; Baddeley et al., 2000) of g is

$$\hat{g}_k(r;b) = \frac{1}{\mathrm{sa}_d r^{d-1}} \sum_{u,v \in X_W}^{\neq} \frac{k_b(r - \|v-u\|)}{\rho(u)\rho(v)|W \cap W_{v-u}|}, \quad r \geq 0,$$

where sa_d is the surface area of the unit sphere in $\mathbb{R}^d$, $\sum^{\neq}$ denotes sum over all distinct points, $1/|W \cap W_h|$, $h \in \mathbb{R}^d$, is the translation edge correction factor with $W_h = \{u - h : u \in W\}$, and $|A|$ is the volume (Lebesgue measure) of $A \subset \mathbb{R}^d$. Variations of this include (Guan, 2007a)

$$\hat{g}_d(r;b) = \frac{1}{\mathrm{sa}_d} \sum_{u,v \in X_W}^{\neq} \frac{k_b(r - \|v-u\|)}{\|v-u\|^{d-1}\rho(u)\rho(v)|W \cap W_{v-u}|}, \quad r \geq 0$$

and the bias corrected estimator (Guan, 2007a)

$$\hat{g}_c(r;b) = \hat{g}_d(r;b)/c(r;b), \quad c(r;b) = \int_{-b}^{\min\{r,b\}} k_b(t)\mathrm{d}t,$$

assuming k has bounded support $[-1,1]$. Regarding the choice of kernel, Illian et al. (2008), p. 230, recommend to use the uniform kernel $k(r) = \mathbb{1}(|r| \leq 1)/2$, where $\mathbb{1}(\,\cdot\,)$ denotes the indicator function, but the Epanechnikov kernel $k(r) = (3/4)(1 - r^2)\mathbb{1}(|r| \leq 1)$ is another common choice. The choice of the bandwidth b highly affects the bias and variance of the kernel estimator. In the planar ($d = 2$) stationary case, Illian et al. (2008), p. 236, recommend $b = 0.10/\sqrt{\hat{\rho}}$ based on practical experience where $\hat{\rho}$ is an estimate of the constant intensity. The default in `spatstat` (Baddeley et al., 2015), following Stoyan and Stoyan (1994), is to use the Epanechnikov kernel with $b = 0.15/\sqrt{\hat{\rho}}$.

Guan (2007b) and Guan (2007a) suggest to choose b by composite likelihood cross validation or by minimizing an estimate of the mean integrated squared error defined over some interval I as

$$\mathrm{MISE}(\hat{g}_m, w) = \mathrm{sa}_d \int_I E\big\{\hat{g}_m(r;b) - g(r)\big\}^2 w(r - r_{\min})\mathrm{d}r, \tag{2}$$

where $\hat{g}_m$, $m = k, d, c$, is one of the aforementioned kernel estimators, $w \geq 0$ is a weight function and $r_{\min} \geq 0$. With $I = (0, R)$, $w(r) = r^{d-1}$ and $r_{\min} = 0$, Guan (2007a) suggests to estimate the mean integrated squared error by

$$M(b) = \mathrm{sa}_d \int_0^R \big\{\hat{g}_m(r;b)\big\}^2 r^{d-1}\mathrm{d}r - 2 \sum_{\substack{u,v \in X_W \\ \|v-u\| \leq R}}^{\neq} \frac{\hat{g}_m^{-\{u,v\}}(\|v-u\|;b)}{\rho(u)\rho(v)|W \cap W_{v-u}|}, \tag{3}$$

where $\hat{g}_m^{-\{u,v\}}$, $m = k, d, c$, is defined as $\hat{g}_m$ but based on the reduced data $(X \setminus \{u,v\}) \cap W$. Loh and Jang (2010) instead use a spatial bootstrap for estimating (2). We return to (3) in Section 5.

3 Orthogonal series estimation

3.1 The new estimator

For an $R > 0$, the new orthogonal series estimator of $g(r)$, $0 \leq r_{\min} < r < r_{\min} + R$, is based on an orthogonal series expansion of $g(r)$ on $(r_{\min}, r_{\min} + R)$:

$$g(r) = \sum_{k=1}^{\infty} \theta_k \phi_k(r - r_{\min}), \tag{4}$$

where $\{\phi_k\}_{k \geq 1}$ is an orthonormal basis of functions on $(0, R)$ with respect to some weight function $w(r) \geq 0$, $r \in (0, R)$. That is, $\int_0^R \phi_k(r)\phi_l(r)w(r)\mathrm{d}r = \mathbb{1}(k = l)$ and the coefficients in the expansion are given by $\theta_k = \int_0^R g(r + r_{\min})\phi_k(r)w(r)\mathrm{d}r$.

For the cosine basis, $w(r) = 1$ and $\phi_1(r) = 1/\sqrt{R}$, $\phi_k(r) = (2/R)^{1/2} \cos\{(k-1)\pi r/R\}$, $k \geq 2$. Another example is the Fourier-Bessel basis with $w(r) = r^{d-1}$ and $\phi_k(r) = 2^{1/2} J_\nu(r\alpha_{\nu,k}/R)\, r^{-\nu}/\{RJ_{\nu+1}(\alpha_{\nu,k})\}$, $k \geq 1$, where $\nu = (d-2)/2$, J_ν is the Bessel function of the first kind of order ν, and $\{\alpha_{\nu,k}\}_{k=1}^{\infty}$ is the sequence of successive positive roots of $J_\nu(r)$.

An estimator of g is obtained by replacing the θ_k in (4) by unbiased estimators and truncating or smoothing the infinite sum. A similar approach has a long history in the context of non-parametric estimation of probability densities, see e.g. the review in Efromovich (2010). For θ_k we propose the estimator

$$\hat{\theta}_k = \frac{1}{\mathrm{sa}_d} \sum_{\substack{u,v \in X_W \\ r_{\min} < \|u-v\| < r_{\min}+R}}^{\neq} \frac{\phi_k(\|v-u\| - r_{\min})w(\|v-u\| - r_{\min})}{\rho(u)\rho(v)\|v-u\|^{d-1}|W \cap W_{v-u}|}, \tag{5}$$

which is unbiased by the second order Campbell formula, see Section S2 of the supplementary material. This type of estimator has some similarity to the coefficient estimators used for probability density estimation but is based on spatial lags $v - u$ which are not independent nor identically distributed. Moreover the estimator is adjusted for the possibly inhomogeneous intensity ρ and corrected for edge effects.

The orthogonal series estimator is finally of the form

$$\hat{g}_o(r; b) = \sum_{k=1}^{\infty} b_k \hat{\theta}_k \phi_k(r - r_{\min}), \tag{6}$$

where $b = \{b_k\}_{k=1}^{\infty}$ is a smoothing/truncation scheme. The simplest smoothing scheme is $b_k = \mathbb{1}[k \leq K]$ for some cut-off $K \geq 1$. Section 3.3 considers several other smoothing schemes.

3.2 Variance of $\hat{\theta}_k$

The factor $\|v-u\|^{d-1}$ in (5) may cause problems when $d > 1$ where the presence of two very close points in X_W could imply division by a quantity close to zero. The expression for the variance of $\hat{\theta}_k$ given in Section S2 of the supplementary material indeed shows that the variance is not finite unless $g(r)w(r - r_{\min})/r^{d-1}$ is bounded for $r_{\min} < r < r_{\min} + R$. If $r_{\min} > 0$ this is always satisfied for bounded g. If $r_{\min} = 0$ the condition is still satisfied in case of the Fourier-Bessel basis and bounded g.

For the cosine basis $w(r) = 1$ so if $r_{\min} = 0$ we need the boundedness of $g(r)/r^{d-1}$. If X satisfies a hard core condition (i.e. two points in X cannot be closer than some $\delta > 0$), this is trivially satisfied. Another example is a determinantal point process (Lavancier et al., 2015) for which $g(r) = 1 - c(r)^2$ for a correlation function c. The boundedness is then e.g. satisfied if $c(\cdot)$ is the Gaussian ($d \leq 3$) or exponential ($d \leq 2$) correlation function. In practice, when using the cosine basis, we take $r_{\min}$ to be a small positive number to avoid issues with infinite variances.

3.3 Mean integrated squared error and smoothing schemes

The orthogonal series estimator (6) has the mean integrated squared error

$$\begin{aligned}\text{MISE}\big(\hat{g}_o, w\big) &= \text{sa}_d \int_{r_{\min}}^{r_{\min}+R} E\big\{\hat{g}_o(r;b) - g(r)\big\}^2 w(r - r_{\min})\mathrm{d}r \\ &= \text{sa}_d \sum_{k=1}^{\infty} E(b_k\hat{\theta}_k - \theta_k)^2 = \text{sa}_d \sum_{k=1}^{\infty} \big[b_k^2 E\{(\hat{\theta}_k)^2\} - 2b_k\theta_k^2 + \theta_k^2\big]. \quad (7)\end{aligned}$$

Each term in (7) is minimized with b_k equal to (cf. Hall, 1987)

$$b_k^* = \frac{\theta_k^2}{E\{(\hat{\theta}_k)^2\}} = \frac{\theta_k^2}{\theta_k^2 + \text{var}(\hat{\theta}_k)}, \quad k \geq 0, \tag{8}$$

leading to the minimal value $\text{sa}_d \sum_{k=1}^{\infty} b_k^* \text{var}(\hat{\theta}_k)$ of the mean integrated square error. Unfortunately, the b_k^* are unknown.

In practice we consider a parametric class of smoothing schemes $b(\psi)$. For practical reasons we need a finite sum in (6) so one component in ψ will be a cut-off index K so that $b_k(\psi) = 0$ when $k > K$. The simplest smoothing scheme is $b_k(\psi) = \mathbb{1}(k \leq K)$. A more refined scheme is $b_k(\psi) = \mathbb{1}(k \leq K)\hat{b}_k^*$ where $\hat{b}_k^* = \widehat{\theta_k^2}/(\hat{\theta}_k)^2$ is an estimate of the optimal smoothing coefficient b_k^* given in (8). Here $\widehat{\theta_k^2}$ is an asymptotically unbiased estimator of θ_k^2 derived in Section 5. For these two smoothing schemes $\psi = K$. Adapting the scheme suggested by Wahba (1981), we also consider $\psi = (K, c_1, c_2)$, $c_1 > 0, c_2 > 1$, and $b_k(\psi) = \mathbb{1}(k \leq K)/(1 + c_1 k^{c_2})$. In practice we choose the smoothing parameter ψ by minimizing an estimate of the mean integrated squared error, see Section 5.

3.4 Expansion of $g(\cdot) - 1$

For large R, $g(r_{\min} + R)$ is typically close to one. However, for the Fourier-Bessel basis, $\phi_k(R) = 0$ for all $k \geq 1$ which implies $\hat{g}_o(r_{\min} + R) = 0$. Hence the estimator cannot be consistent for $r = r_{\min} + R$ and the convergence of the estimator for $r \in (r_{\min}, r_{\min} + R)$ can be quite slow as the number of terms K in the estimator increases. In practice we obtain quicker convergence by applying the Fourier-Bessel expansion to $g(r) - 1 = \sum_{k\geq 1} \vartheta_k \phi_k(r - r_{\min})$ so that the estimator becomes $\tilde{g}_o(r;b) = 1 + \sum_{k=1}^{\infty} b_k \hat{\vartheta}_k \phi_k(r - r_{\min})$ where $\hat{\vartheta}_k = \hat{\theta}_k - \int_0^{r_{\min}+R} \phi_k(r)w(r)\mathrm{d}r$ is an estimator of $\vartheta_k = \int_0^R \{g(r + r_{\min}) - 1\}\phi_k(r)w(r)\mathrm{d}r$. Note that $\text{var}(\hat{\vartheta}_k) = \text{var}(\hat{\theta}_k)$ and $\tilde{g}_o(r;b) - E\{\tilde{g}_o(r;b)\} = \hat{g}_o(r;b) - E\{\hat{g}_o(r;b)\}$. These identities imply that the results regarding consistency and asymptotic normality established for $\hat{g}_o(r;b)$ in Section 4 are also valid for $\tilde{g}_o(r;b)$.

4 Consistency and asymptotic normality

4.1 Setting

To obtain asymptotic results we assume that X is observed through an increasing sequence of observation windows W_n. For ease of presentation we assume square observation windows $W_n = \times_{i=1}^d [-na_i, na_i]$ for some $a_i > 0$, $i = 1, \ldots, d$. More general sequences of windows can be used at the expense of more notation and assumptions. We also consider an associated sequence ψ_n, $n \geq 1$, of smoothing parameters satisfying conditions to be detailed in the following. We let $\hat{\theta}_{k,n}$ and $\hat{g}_{o,n}$ denote the estimators of θ_k and g obtained from X observed on W_n. Thus

$$\hat{\theta}_{k,n} = \frac{1}{\mathrm{sa}_d |W_n|} \sum_{\substack{u,v \in X_{W_n} \\ v-u \in B_{r_{\min}}^R}}^{\neq} \frac{\phi_k(\|v-u\| - r_{\min}) w(\|v-u\| - r_{\min})}{\rho(u)\rho(v)\|v-u\|^{d-1} e_n(v-u)}),$$

where

$$B_{r_{\min}}^R = \{h \in \mathbb{R}^d \mid r_{\min} < \|h\| < r_{\min} + R\} \quad \text{and} \quad e_n(h) = |W_n \cap (W_n)_h| / |W_n|. \tag{9}$$

Further,

$$\hat{g}_{o,n}(r; b) = \sum_{k=1}^{K_n} b_k(\psi_n) \hat{\theta}_{k,n} \phi_k(r - r_{\min}) = \frac{1}{\mathrm{sa}_d |W_n|} \sum_{\substack{u,v \in X_{W_n} \\ v-u \in B_{r_{\min}}^R}}^{\neq} \frac{w(\|v-u\|) \varphi_n(v-u, r)}{\rho(u)\rho(v)\|v-u\|^{d-1} e_n(v-u)|},$$

where

$$\varphi_n(h, r) = \sum_{k=1}^{K_n} b_k(\psi_n) \phi_k(\|h\| - r_{\min}) \phi_k(r - r_{\min}). \tag{10}$$

In the results below we refer to higher order normalized joint intensities $g^{(k)}$ of X. Define the k'th order joint intensity of X by the identity

$$E\left\{ \sum_{u_1, \ldots, u_k \in X}^{\neq} \mathbb{1}(u_1 \in A_1, \ldots, u_k \in A_k) \right\} = \int_{A_1 \times \cdots \times A_k} \rho^{(k)}(v_1, \ldots, v_k) \mathrm{d}v_1 \cdots \mathrm{d}v_k$$

for bounded subsets $A_i \subset \mathbb{R}^d$, $i = 1, \ldots, k$, where the sum is over distinct $u_1, \ldots, u_k$. We then let $g^{(k)}(v_1, \ldots, v_k) = \rho^{(k)}(v_1, \ldots, v_k) / \{\rho(v_1) \cdots \rho(v_k)\}$ and assume with an abuse of notation that the $g^{(k)}$ are translation invariant for $k = 3, 4$, i.e. $g^{(k)}(v_1, \ldots, v_k) = g^{(k)}(v_2 - v_1, \ldots, v_k - v_1)$.

4.2 Consistency of orthogonal series estimator

Consistency of the orthogonal series estimator can be established under fairly mild conditions following the approach in Hall (1987). We first state some conditions that ensure (see Section S2 of the supplementary material) that $\mathrm{var}(\hat{\theta}_{k,n}) \leq C_1/|W_n|$ for some $0 < C_1 < \infty$:

V1 There exists $0 < \rho_{\min} < \rho_{\max} < \infty$ such that for all $u \in \mathbb{R}^d$, $\rho_{\min} \leq \rho(u) \leq \rho_{\max}$.

V2 For any $h, h_1, h_2 \in B^R_{r_{\min}}$, $g(h)w(\|h\| - r_{\min}) \leq C_2\|h\|^{d-1}$ and $g^{(3)}(h_1, h_2) \leq C_3$ for constants $C_2, C_3 < \infty$.

V3 A constant $C_4 < \infty$ can be found such that $\sup_{h_1,h_2 \in B^R_{r_{\min}}} \int_{\mathbb{R}^d} \Big| g^{(4)}(h_1, h_3, h_2 + h_3) - g(h_1)g(h_2) \Big| \mathrm{d}h_3 \leq C_4$.

The first part of V2 is needed to ensure finite variances of the $\hat\theta_{k,n}$ and is discussed in detail in Section 3.2. The second part simply requires that $g^{(3)}$ is bounded. The condition V3 is a weak dependence condition which is also used for asymptotic normality in Section 4.3 and for estimation of θ_k^2 in Section 5.

Regarding the smoothing scheme, we assume

S1 $B = \sup_{k,\psi} \big|b_k(\psi)\big| < \infty$ and for all ψ, $\sum_{k=1}^\infty \big|b_k(\psi)\big| < \infty$.

S2 $\psi_n \to \psi^*$ for some ψ^*, and $\lim_{\psi\to\psi^*} \max_{1\leq k\leq m} \big|b_k(\psi) - 1\big| = 0$ for all $m \geq 1$.

S3 $|W_n|^{-1} \sum_{k=1}^\infty \big|b_k(\psi_n)\big| \to 0$.

E.g. for the simplest smoothing scheme, $\psi_n = K_n$, $\psi^* = \infty$ and we assume $K_n/|W_n| \to 0$.

Assuming the above conditions we now verify that the mean integrated squared error of $\hat g_{o,n}$ tends to zero as $n \to \infty$. By (7), $\text{MISE}\big(\hat g_{o,n}, w\big)/\text{sa}_d = \sum_{k=1}^\infty \big[b_k(\psi_n)^2 \text{var}(\hat\theta_k) + \theta_k^2\{b_k(\psi_n) - 1\}^2\big]$. By V1-V3 and S1 the right hand side is bounded by

$$BC_1|W_n|^{-1} \sum_{k=1}^\infty \big|b_k(\psi_n)\big| + \max_{1\leq k\leq m} \theta_k^2 \sum_{k=1}^m (b_k(\psi_n) - 1)^2 + (B^2 + 1) \sum_{k=m+1}^\infty \theta_k^2.$$

By Parseval's identity, $\sum_{k=1}^\infty \theta_k^2 < \infty$. The last term can thus be made arbitrarily small by choosing m large enough. It also follows that θ_k^2 tends to zero as $k \to \infty$. Hence, by S2, the middle term can be made arbitrarily small by choosing n large enough for any choice of m. Finally, the first term can be made arbitrarily small by S3 and choosing n large enough.

4.3 Asymptotic normality

The estimators $\hat\theta_{k,n}$ as well as the estimator $\hat g_{o,n}(r; b)$ are of the form

$$S_n = \frac{1}{\text{sa}_d|W_n|} \sum_{\substack{u,v \in X_{W_n} \\ v-u \in B^R_{r_{\min}}}}^{\neq} \frac{f_n(v-u)}{\rho(u)\rho(v)e_n(v-u)} \tag{11}$$

for a sequence of even functions $f_n : \mathbb{R}^d \to \mathbb{R}$. We let $\tau_n^2 = |W_n|\text{var}(S_n)$.

To establish asymptotic normality of estimators of the form (11) we need certain mixing properties for X as in Waagepetersen and Guan (2009). The strong mixing coefficient for the point process X on $\mathbb{R}^d$ is given by (Ivanoff, 1982; Politis et al., 1998)

$$\alpha_{\mathbf{X}}(m; a_1, a_2) = \sup\big\{\big|\text{pr}(E_1 \cap E_2) - \text{pr}(E_1)\text{pr}(E_2)\big| : E_1 \in \mathcal{F}_X(B_1), E_2 \in \mathcal{F}_X(B_2),$$
$$|B_1| \leq a_1, |B_2| \leq a_2, \mathcal{D}(B_1, B_2) \geq m, B_1, B_2 \in \mathcal{B}(\mathbb{R}^d)\big\},$$

where $\mathcal{B}(\mathbb{R}^d)$ denotes the Borel σ-field on $\mathbb{R}^d$, $\mathcal{F}_X(B_i)$ is the σ-field generated by $X \cap B_i$ and

$$\mathcal{D}(B_1, B_2) = \inf\big\{\max_{1\le i\le d}|u_i - v_i| : u = (u_1,\dots,u_d) \in B_1, v = (v_1,\dots,v_d) \in B_2\big\}.$$

To verify asymptotic normality we need the following assumptions as well as V1 (the conditions V2 and V3 are not needed due to conditions N2 and N4 below):

- N1 The mixing coefficient satisfies $\alpha_X(m;(s+2R)^d,\infty) = O(m^{-d-\varepsilon})$ for some $s,\varepsilon > 0$.
- N2 There exists a $\eta > 0$ and $L_1 < \infty$ such that $g^{(k)}(h_1,\dots,h_{k-1}) \le L_1$ for $k = 2,\dots,2(2+\lceil\eta\rceil)$ and all $h_1,\dots,h_{k-1} \in \mathbb{R}^d$.
- N3 $\liminf_{n\to\infty}\tau_n^2 > 0$.
- N4 There exists $L_2 < \infty$ so that $|f_n(h)| \le L_2$ for all $n \ge 1$ and $h \in B^R_{r_{\min}}$.

The conditions N1-N3 are standard in the point process literature, see e.g. the discussions in Waagepetersen and Guan (2009) and Coeurjolly and Møller (2014). The condition N3 is difficult to verify and is usually left as an assumption, see Waagepetersen and Guan (2009), Coeurjolly and Møller (2014) and Dvořák and Prokešová (2016). However, at least in the stationary case, and in case of estimation of $\hat\theta_{k,n}$, the expression for $\mathrm{var}(\hat\theta_{k,n})$ in Section S2 of the supplementary material shows that $\tau_n^2 = |W_n|\mathrm{var}(\hat\theta_{k,n})$ converges to a constant which supports the plausibility of condition N3. We discuss N4 in further detail below when applying the general framework to $\hat\theta_{k,n}$ and $\hat g_{o,n}$. The following theorem is proved in Section S3 of the supplementary material.

Theorem 1 *Under conditions V1, N1-N4,* $\tau_n^{-1}|W_n|^{1/2}\{S_n - E(S_n)\} \xrightarrow{D} N(0,1)$.

4.4 Application to $\hat\theta_{k,n}$ and $\hat g_{o,n}$

In case of estimation of θ_k, $\hat\theta_{k,n} = S_n$ with $f_n(h) = \phi_k(\|h\| - r_{\min})w(\|h\| - r_{\min})/\|h\|^{d-1}$. The assumption N4 is then straightforwardly seen to hold in the case of the Fourier-Bessel basis where $|\phi_k(r)| \le |\phi_k(0)|$ and $w(r) = r^{d-1}$. For the cosine basis, N4 does not hold in general and further assumptions are needed, cf. the discussion in Section 3.2. For simplicity we here just assume $r_{\min} > 0$. Thus we state the following

Corollary 1 *Assume V1, N1-N4, and, in case of the cosine basis, that* $r_{\min} > 0$. *Then*

$$\{var(\hat\theta_{k,n})\}^{-1/2}(\hat\theta_{k,n} - \theta_k) \xrightarrow{D} N(0,1).$$

For $\hat g_{o,n}(r;b) = S_n$,

$$f_n(h) = \frac{\varphi_n(h,r)w(\|h\| - r_{\min})}{\|h\|^{d-1}} = \frac{w(\|h\| - r_{\min})}{\|h\|^{d-1}}\sum_{k=1}^{K_n} b_k(\psi_n)\phi_k(\|h\|-r_{\min})\phi_k(r-r_{\min}),$$

where φ_n is defined in (10). In this case, f_n is typically not uniformly bounded since the number of not necessarily decreasing terms in the sum defining φ_n in (10) grows with n. We therefore introduce one more condition:

N5 There exist an $\omega > 0$ and $M_\omega < \infty$ so that

$$K_n^{-\omega} \sum_{k=1}^{K_n} b_k(\psi_n) \big| \phi_k(r - r_{\min}) \phi_k(\|h\| - r_{\min}) \big| \le M_\omega$$

for all $h \in B_{r_{\min}}^R$.

Given N5, we can simply rescale: $\tilde{S}_n := K_n^{-\omega} S_n$ and $\tilde{\tau}_n^2 := K_n^{-2\omega} \tau_n^2$. Then, assuming $\liminf_{n\to\infty} \tilde{\tau}_n^2 > 0$, Theorem 1 gives the asymptotic normality of $\tilde{\tau}_n^{-1} |W_n|^{1/2} \{\tilde{S}_n - E(\tilde{S}_n)\}$ which is equal to $\tau_n^{-1} |W_n|^{1/2} \{S_n - E(S_n)\}$. Hence we obtain

Corollary 2 *Assume V1, N1-N2, N5 and $\liminf_{n\to\infty} K_n^{-2\omega} \tau_n^2 > 0$. In case of the cosine basis, assume further $r_{\min} > 0$. Then for $r \in (r_{\min}, r_{\min} + R)$,*

$$\tau_n^{-1} |W_n|^{1/2} \big[\hat{g}_{o,n}(r; b) - E\{\hat{g}_{o,n}(r; b)\}\big] \xrightarrow{D} N(0,1).$$

In case of the simple smoothing scheme $b_k(\psi_n) = \mathbb{1}(k \le K_n)$, we take $\omega = 1$ for the cosine basis. For the Fourier-Bessel basis we take $\omega = 4/3$ when $d = 1$ and $\omega = d/2 + 2/3$ when $d > 1$ (see the derivations in Section S6 of the supplementary material).

5 Tuning the smoothing scheme

In practice we choose K, and other parameters in the smoothing scheme $b(\psi)$, by minimizing an estimate of the mean integrated squared error. This is equivalent to minimizing

$$\mathrm{sa}_d I(\psi) = \mathrm{MISE}(\hat{g}_o, w) - \int_{r_{\min}}^{r_{\min}+R} \big\{g(r) - 1\big\}^2 w(r) \mathrm{d}r = \sum_{k=1}^{K} \big[b_k(\psi)^2 E\{(\hat{\theta}_k)^2\} - 2 b_k(\psi) \theta_k^2\big]. \tag{12}$$

In practice we must replace (12) by an estimate. Define $\widehat{\theta_k^2}$ as

$$\sum_{\substack{u,v,u',v' \in X_W \\ v-u, v'-u' \in B_{r_{\min}}^R}}^{\neq} \frac{\phi_k(\|v-u\| - r_{\min}) \phi_k(\|v'-u'\| - r_{\min}) w(\|v-u\| - r_{\min}) w(\|v'-u'\| - r_{\min})}{\mathrm{sa}_d{}^2 \rho(u)\rho(v)\rho(u')\rho(v') \|v-u\|^{d-1} \|v'-u'\|^{d-1} |W \cap W_{v-u}| |W \cap W_{v'-u'}|}.$$

Then, referring to the set-up in Section 4 and assuming V3,

$$\lim_{n\to\infty} E(\widehat{\theta_{k,n}^2}) \to \left\{ \int_0^R g(r + r_{\min}) \phi_k(r) w(r) \mathrm{d}r \right\}^2 = \theta_k^2$$

(see Section S4 of the supplementary material) and hence $\widehat{\theta_{k,n}^2}$ is an asymptotically unbiased estimator of θ_k^2. The estimator is obtained from $(\hat{\theta}_k)^2$ by retaining only terms where all four points u, v, u', v' involved are distinct. In simulation studies, $\widehat{\theta_k^2}$ had a smaller root mean squared error than $(\hat{\theta}_k)^2$ for estimation of θ_k^2.

Thus

$$\hat{I}(\psi) = \sum_{k=1}^{K} \{b_k(\psi)^2 (\hat{\theta}_k)^2 - 2 b_k(\psi) \widehat{\theta_k^2}\} \tag{13}$$

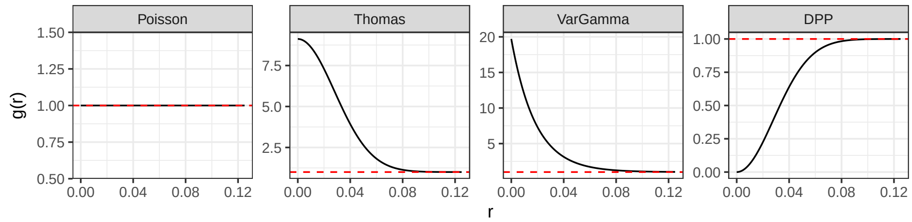

Figure 1: Pair correlation functions for the point processes considered in the simulation study.

is an asymptotically unbiased estimator of (12). Moreover, (13) is equivalent to the following slight modification of Guan (2007a)'s criterion (3):

$$\int_{r_{\min}}^{r_{\min}+R} \left\{\hat{g}_o(r;b)\right\}^2 w(r-r_{\min})\mathrm{d}r - \frac{2}{\mathrm{sa}_d} \sum_{\substack{u,v\in X_W \\ v-u\in B^R_{r_{\min}}}}^{\neq} \frac{\hat{g}_o^{-\{u,v\}}(\|v-u\|;b)w(\|v-u\|-r_{\min})}{\rho(u)\rho(v)|W\cap W_{v-u}|}.$$

For the simple smoothing scheme $b_k(K) = \mathbb{1}(k \le K)$, (13) reduces to

$$\hat{I}(K) = \sum_{k=1}^{K} \{(\hat{\theta}_k)^2 - 2\widehat{\theta^2_k}\} = \sum_{k=1}^{K} (\hat{\theta}_k)^2(1-2\hat{b}^*_k), \tag{14}$$

where $\hat{b}^*_k = \widehat{\theta^2_k}/(\hat{\theta}_k)^2$ is an estimator of b^*_k in (8).

In practice, uncertainties of $\hat{\theta}_k$ and $\widehat{\theta^2_k}$ lead to numerical instabilities in the minimization of (13) with respect to ψ. To obtain a numerically stable procedure we first determine K as

$$\hat{K} = \inf\{2 \le k \le K_{\max} : (\hat{\theta}_{k+1})^2 - 2\widehat{\theta^2_{k+1}} > 0\} = \inf\{2 \le k \le K_{\max} : \hat{b}^*_{k+1} < 1/2\}. \tag{15}$$

That is, $\hat{K}$ is the first local minimum of (14) larger than 1 and smaller than an upper limit $K_{\max}$ which we chose to be 49 in the applications. This choice of K is also used for the refined and the Wahba smoothing schemes. For the refined smoothing scheme we thus let $b_k = \mathbb{1}(k \le \hat{K})\hat{b}^*_k$. For the Wahba smoothing scheme $b_k = \mathbb{1}(k \le \hat{K})/(1+\hat{c}_1 k^{\hat{c}_2})$, where $\hat{c}_1$ and $\hat{c}_2$ minimize $\sum_{k=1}^{\hat{K}} \left\{(\hat{\theta}_k)^2/(1+c_1k^{c_2})^2 - 2\widehat{\theta^2_k}/(1+c_1k^{c_2})\right\}$ over $c_1 > 0$ and $c_2 > 1$.

6 Simulation study

We compare the performance of the orthogonal series estimators and the kernel estimators for data simulated on $W = [0,1]^2$ or $W = [0,2]^2$ from four point processes with constant intensity $\rho = 100$. More specifically, we consider $n_{\text{sim}} = 1000$ realizations from a Poisson process, a Thomas process (parent intensity $\kappa = 25$, dispersion standard deviation $\omega = 0.0198$), a Variance Gamma cluster process (parent intensity $\kappa = 25$, shape parameter $\nu = -1/4$, dispersion parameter $\omega = 0.01845$, Jalilian et al., 2013), and a determinantal point process with pair correlation function $g(r) = 1 - \exp\{-2(r/\alpha)^2\}$ and $\alpha = 0.056$. The pair correlation functions of these point processes are shown in Figure 1.

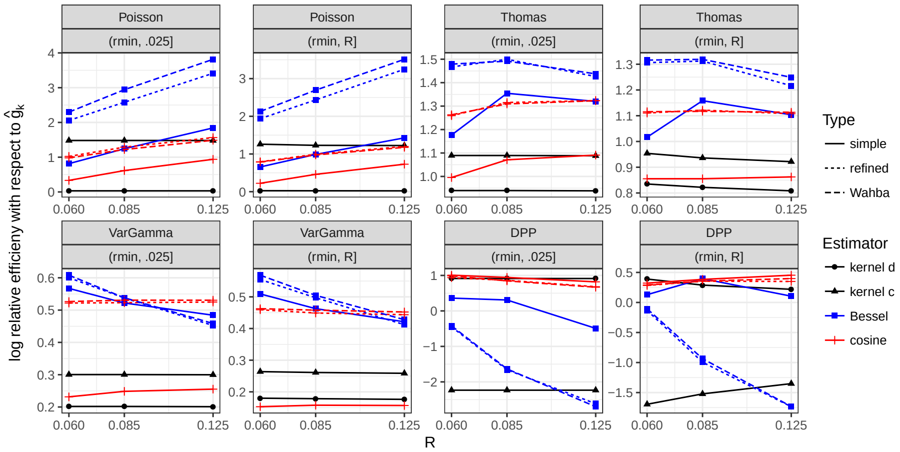

Figure 2: Plots of log relative efficiencies for small lags $(r_{\min}, 0.025]$ and all lags $(r_{\min}, R]$, $R = 0.06, 0.085, 0.125$, and $W = [0,1]^2$. Black: kernel estimators. Blue and red: orthogonal series estimators with Bessel respectively cosine basis. Lines serve to ease visual interpretation.

For each realization, $g(r)$ is estimated for r in $(r_{\min}, r_{\min} + R)$, with $r_{\min} = 10^{-3}$ and $R = 0.06, 0.085, 0.125$, using the kernel estimators $\hat{g}_k(r; b)$, $\hat{g}_d(r; b)$ and $\hat{g}_c(r; b)$ or the orthogonal series estimator $\hat{g}_o(r; b)$. The Epanechnikov kernel with bandwidth $b = 0.15/\sqrt{\hat{\rho}}$ is used for $\hat{g}_k(r; b)$ and $\hat{g}_d(r; b)$ while the bandwidth of $\hat{g}_c(r; b)$ is chosen by minimizing Guan (2007a)'s estimate (3) of the mean integrated squared error. For the orthogonal series estimator, we consider both the cosine and the Fourier-Bessel bases with simple, refined or Wahba smoothing schemes. For the Fourier-Bessel basis we use the modified orthogonal series estimator described in Section 3.4. The parameters for the smoothing scheme are chosen according to Section 5.

From the simulations we estimate the mean integrated squared error (2) with $w(r) = 1$ of each estimator $\hat{g}_m$, $m = k, d, c, o$, over the intervals $[r_{\min}, 0.025]$ (small spatial lags) and $[r_{\min}, r_{\min} + R]$ (all lags). We consider the kernel estimator $\hat{g}_k$ as the baseline estimator and compare any of the other estimators $\hat{g}$ with $\hat{g}_k$ using the log relative efficiency $e_I(\hat{g}) = \log\{\widehat{\mathrm{MISE}}_I(\hat{g}_k)/\widehat{\mathrm{MISE}}_I(\hat{g})\}$, where $\widehat{\mathrm{MISE}}_I(\hat{g})$ denotes the estimated mean squared integrated error over the interval I for the estimator $\hat{g}$. Thus $e_I(\hat{g}) > 0$ indicates that $\hat{g}$ outperforms $\hat{g}_k$ on the interval I. Results for W=$[0,1]^2$ are summarized in Figure 2.

For all types of point processes, the orthogonal series estimators outperform or does as well as the kernel estimators both at small lags and over all lags. The detailed conclusions depend on whether the non-repulsive Poisson, Thomas and Var Gamma processes or the repulsive determinantal process are considered. Orthogonal-Bessel with refined or Wahba smoothing is superior for Poisson, Thomas and Var Gamma but only better than $\hat{g}_c$ for the determinantal point process. The performance of the orthogonal-cosine estimator is between or better than the performance of the kernel estimators for Poisson, Thomas and Var Gamma and is as good as the best kernel estimator for determinantal. Regarding the kernel estimators, $\hat{g}_c$ is better than $\hat{g}_d$ for Poisson, Thomas and Var Gamma and worse than $\hat{g}_d$ for determinantal. The above conclusions are stable over the three R values considered. For $W = [0,2]^2$ (see Figure S1

Table 1: Monte Carlo mean, standard error, skewness (S) and kurtosis (K) of $\hat{g}_o(r)$ using the Bessel basis with the simple smoothing scheme in case of the Thomas process on observation windows $W_1 = [0,1]^2$, $W_2 = [0,2]^2$ and $W_3 = [0,3]^3$.

	r	$g(r)$	$\hat{E}\{\hat{g}_o(r)\}$	$[\hat{\mathrm{var}}\{\hat{g}_o(r)\}]^{1/2}$	$\hat{\mathrm{S}}\{\hat{g}_o(r)\}$	$\hat{\mathrm{K}}\{\hat{g}_o(r)\}$
W_1	0.025	3.972	3.961	0.923	1.145	5.240
W_1	0.1	1.219	1.152	0.306	0.526	3.516
W_2	0.025	3.972	3.959	0.467	0.719	4.220
W_2	0.1	1.219	1.187	0.150	0.691	4.582
W_3	0.025	3.972	3.949	0.306	0.432	3.225
W_3	0.1	1.2187	1.2017	0.0951	0.2913	2.9573

in the supplementary material) the conclusions are similar but with more clear superiority of the orthogonal series estimators for Poisson and Thomas. For Var Gamma the performance of $\hat{g}_c$ is similar to the orthogonal series estimators. For determinantal and $W = [0,2]^2$, $\hat{g}_c$ is better than orthogonal-Bessel-refined/Wahba but still inferior to orthogonal-Bessel-simple and orthogonal-cosine. Figures S2 and S3 in the supplementary material give a more detailed insight in the bias and variance properties for $\hat{g}_k$, $\hat{g}_c$, and the orthogonal series estimators with simple smoothing scheme. Table S1 in the supplementary material shows that the selected K in general increases when the observation window is enlargened, as required for the asymptotic results. The general conclusion, taking into account the simulation results for all four types of point processes, is that the best overall performance is obtained with orthogonal-Bessel-simple, orthogonal-cosine-refined or orthogonal-cosine-Wahba.

To supplement our theoretical results in Section 4 we consider the distribution of the simulated $\hat{g}_o(r;b)$ for $r = 0.025$ and $r = 0.1$ in case of the Thomas process and using the Fourier-Bessel basis with the simple smoothing scheme. In addition to $W = [0,1]^2$ and $W = [0,2]^2$, also $W = [0,3]^2$ is considered. The mean, standard error, skewness and kurtosis of $\hat{g}_o(r)$ are given in Table 1 while histograms of the estimates are shown in Figure S3. The standard error of $\hat{g}_o(r;b)$ scales as $|W|^{1/2}$ in accordance with our theoretical results. Also the bias decreases and the distributions of the estimates become increasingly normal as $|W|$ increases.

7 Application

We consider point patterns of locations of *Acalypha diversifolia* (528 trees), *Lonchocarpus heptaphyllus* (836 trees) and *Capparis frondosa* (3299 trees) species in the 1995 census for the 1000m $\times$ 500m Barro Colorado Island plot (Hubbell and Foster, 1983; Condit, 1998). To estimate the intensity function of each species, we use a log-linear regression model depending on soil condition (contents of copper, mineralized nitrogen, potassium and phosphorus and soil acidity) and topographical (elevation, slope gradient, multiresolution index of valley bottom flatness, ncoming mean solar radiation and the topographic wetness index) variables. The regression parameters are estimated using the quasi-likelihood approach in Guan et al. (2015). The point patterns and fitted intensity functions are shown in Figure S5 in the supplementary material.

The pair correlation function of each species is then estimated using the bias corrected kernel estimator $\hat{g}_c(r;b)$ with b determined by minimizing (3) and the orthogonal

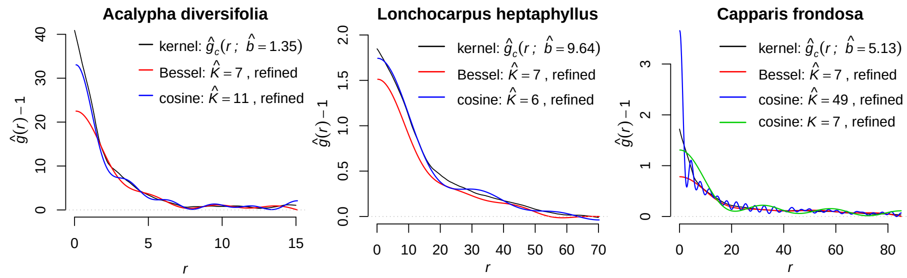

Figure 3: Estimated pair correlation functions for tropical rain forest trees.

series estimator $\hat{g}_o(r; b)$ with both Fourier-Bessel and cosine basis, refined smoothing scheme and the optimal cut-offs $\hat{K}$ obtained from (15); see Figure 3.

For *Lonchocarpus* the three estimates are quite similar while for *Acalypha* and *Capparis* the estimates deviate markedly for small lags and then become similar for lags greater than respectively 2 and 8 meters. For *Capparis* and the cosine basis, the number of selected coefficients coincides with the chosen upper limit 49 for the number of coefficients. The cosine estimate displays oscillations which appear to be artefacts of using high frequency components of the cosine basis. The function (14) decreases very slowly after $K = 7$ so we also tried the cosine estimate with $K = 7$ which gives a more reasonable estimate.

Acknowledgement

Rasmus Waagepetersen is supported by the Danish Council for Independent Research — Natural Sciences, grant "Mathematical and Statistical Analysis of Spatial Data", and by the "Centre for Stochastic Geometry and Advanced Bioimaging", funded by the Villum Foundation.

Supplementary material

Supplementary material includes proofs of consistency and asymptotic normality results and details of the simulation study and data analysis.

References

Baddeley, A., Rubak, E., and Turner, R. (2015). *Spatial Point Patterns: Methodology and Applications with R*. Chapman & Hall/CRC Interdisciplinary Statistics. Chapman & Hall/CRC, Boca Raton.

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.

Coeurjolly, J.-f. and Møller, J. (2014). Variational approach for spatial point process intensity estimation. *Bernoulli*, 20(3):1097–1125.

Coeurjolly, J.-F., Møller, J., and Waagepetersen, R. (2016). A tutorial on Palm distributions for spatial point processes. *International Statistical Review*. to appear.

Condit, R. (1998). *Tropical forest census plots*. Springer-Verlag and R. G. Landes Company, Berlin, Germany and Georgetown, Texas.

Dvořák, J. and Prokešová, M. (2016). Asymptotic properties of the minimum contrast estimators for projections of inhomogeneous space-time shot-noise cox processes. *Applications of Mathematics*, 61(4):387–411.

Efromovich, S. (2010). Orthogonal series density estimation. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2(4):467–476.

Guan, Y. (2007a). A least-squares cross-validation bandwidth selection approach in pair correlation function estimations. *Statistics and Probability Letters*, 77(18):1722–1729.

Guan, Y. (2007b). A composite likelihood cross-validation approach in selecting bandwidth for the estimation of the pair correlation function. *Scandinavian Journal of Statistics*, 34(2):336–346.

Guan, Y., Jalilian, A., and Waagepetersen, R. (2015). Quasi-likelihood for spatial point processes. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 77(3):677–697.

Hall, P. (1987). Cross-validation and the smoothing of orthogonal series density estimators. *Journal of Multivariate Analysis*, 21(2):189–206.

Hubbell, S. P. and Foster, R. B. (1983). Diversity of canopy trees in a Neotropical forest and implications for the conservation of tropical trees. In Sutton, S. L., Whitmore, T. C., and Chadwick, A. C., editors, *Tropical Rain Forest: Ecology and Management.*, pages 25–41. Blackwell Scientific Publications, Oxford.

Illian, J., Penttinen, A., Stoyan, H., and Stoyan, D. (2008). *Statistical Analysis and Modelling of Spatial Point Patterns*, volume 76. Wiley, London.

Ivanoff, G. (1982). Central limit theorems for point processes. *Stochastic Processes and their Applications*, 12(2):171–186.

Jalilian, A., Guan, Y., and Waagepetersen, R. (2013). Decomposition of variance for spatial Cox processes. *Scandinavian Journal of Statistics*, 40(1):119–137.

Lavancier, F., Møller, J., and Rubak, E. (2015). Determinantal point process models and statistical inference. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 77:853–877.

Loh, J. M. and Jang, W. (2010). Estimating a cosmological mass bias parameter with bootstrap bandwidth selection. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 59(5):761–779.

Møller, J. and Waagepetersen, R. P. (2003). *Statistical inference and simulation for spatial point processes*. Chapman and Hall/CRC, Boca Raton.

Politis, D. N., Paparoditis, E., and Romano, J. P. (1998). Large sample inference for irregularly space dependent opservations based on subsampling. *Sankhya: The Indian Journal of Statistics*, 60(2):274–292.

Stoyan, D. and Stoyan, H. (1994). *Fractals, Random Shapes and Point Fields: Methods of Geometrical Statistics*. Wiley.

Stoyan, D. and Stoyan, H. (1996). Estimating pair correlation functions of planar cluster processes. *Biometrical Journal*, 38(3):259–271.

Waagepetersen, R. and Guan, Y. (2009). Two-step estimation for inhomogeneous spatial point processes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71(3):685–702.

Wahba, G. (1981). Data-based optimal smoothing of orthogonal series density estimates. *Annals of Statistics*, 9:146–l56.

Yue, Y. R. and Loh, J. M. (2013). Bayesian nonparametric estimation of pair correlation function for inhomogeneous spatial point processes. *Journal of Nonparametric Statistics*, 25(2):463–474.