Statistica Sinica **29** (2019), 769-787
doi:https://doi.org/10.5705/ss.202017.0112

ORTHOGONAL SERIES ESTIMATION OF THE PAIR CORRELATION FUNCTION OF A SPATIAL POINT PROCESS

Abdollah Jalilian, Yongtao Guan and Rasmus Waagepetersen

Razi University, University of Miami and Aalborg University

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 an orthogonal series non-parametric estimator. It is consistent and asymptotically normal according to our theoretical and simulation results. In our simulations the new estimator outperforms the kernel estimators, in particular for Poisson and clustered point processes.

Key words and phrases: 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 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)) and 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, 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, Guan and Waagepetersen (2013)).

In this paper we adapt orthogonal series density estimators (see e.g. the reviews in Hall (1987) and Efromovich (2010)) to the non-parametric 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 orthogonal series estimator and study its performance in a simulation study and in an application to a tropical rain forest data set.

2. Background

2.1. Spatial point processes

We denote by X a point process on $\mathbb{R}^d$, $d \geq 1$, 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 $A, B \subset \mathbb{R}^d$,

$$\mathbb{E}\{N(B)\} = \int_B \rho(u)\mathrm{d}u,$$

$$\mathbb{E}\{N(A)N(B)\} = \int_{A\cap B} \rho(u)\mathrm{d}u + \int_A \int_B \rho^{(2)}(u,v)\mathrm{d}u\mathrm{d}v. \tag{2.1}$$

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 (2.1),

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

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, Møller and Waagepetersen (2017)). In this paper we assume that g is isotropic, 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, recalling that g is assumed to be isotropic, a kernel density estimator (Stoyan and Stoyan (1994), Baddeley, Møller and Waagepetersen (2000)) of g is

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

where $\varsigma_d = 2\pi^{d/2}/\Gamma(d/2)$ is the surface area of the unit sphere in $\mathbb{R}^d$, $\sum^{\neq}$ denotes sum over all pairs of 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)),

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

and the bias corrected estimator

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

assuming k has bounded support $[-1, 1]$. Regarding the choice of kernel, Illian et al. (2008) recommend use of the uniform kernel $k(r) = \mathbb{I}(|r| \leq 1)/2$, where $\mathbb{I}(\cdot)$ denotes the indicator function; the Epanechnikov kernel $k(r) = (3/4)(1 - r^2)\mathbb{I}(|r| \leq 1)$ 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) 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, Rubak and Turner (2015)), following Stoyan and Stoyan (1994), is to use the Epanechnikov kernel with $b = 0.15/\sqrt{\hat{\rho}}$.

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

$$\mathrm{MISE}(\hat{g}_m, w) = \varsigma_d \int_I \mathbb{E}\{\hat{g}_m(r; b) - g(r)\}^2 w(r - r_{\min})\mathrm{d}r, \tag{2.2}$$

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 estimating the mean integrated squared error by

$$M(b) = \varsigma_d \int_0^R \{\hat{g}_m(r;b)\}^2 r^{d-1}\mathrm{d}r - 2 \sum_{\substack{u,v \in X_W \\ \|v-u\| \le R}}^{\neq} \frac{\hat{g}_m^{-\{u,v\}}(\|v-u\|;b)}{\rho(u)\rho(v)|W \cap W_{v-u}|}, \tag{2.3}$$

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 (2.2). We return to (2.3) in Sec. 5.

3. Orthogonal Series Estimation

3.1. The estimator

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

$$g(r) = \sum_{k=1}^{\infty} \theta_k \phi_k(r - r_{\min}), \tag{3.1}$$

where $\{\phi_k\}_{k \ge 1}$ is an orthonormal basis of functions on $(0, R)$ with respect to some weight function $w(r) \ge 0$, $r \in (0, R)$. Thus,

$$\int_0^R \phi_k(r)\phi_l(r)w(r)\mathrm{d}r = \mathbb{I}(k = l)$$

and the coefficients in the expansion are given by

$$\theta_k = \int_{r_{\min}}^{r_{\min}+R} g(r)\phi_k(r - r_{\min})w(r - r_{\min})\mathrm{d}r = \int_0^R g(r + r_{\min})\phi_k(r)w(r)\mathrm{d}r.$$

For the cosine basis, $w(r) = 1$, $\phi_1(r) = 1/\sqrt{R}$, and

$$\phi_k(r) = \left(\frac{2}{R}\right)^{1/2} \cos\left(\frac{(k-1)\pi r}{R}\right), \quad k \ge 2.$$

Another example is the Fourier-Bessel basis with $w(r) = r^{d-1}$ and

$$\phi_k(r) = \frac{2^{1/2}}{R J_{\nu+1}(\alpha_{\nu,k})} J_\nu\left(\frac{r\alpha_{\nu,k}}{R}\right) r^{-\nu}, \quad k \ge 1,$$

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 $J_\nu(r)$. Plots of cosine and Fourier-Bessel basis functions are shown in Figure S6 of the supplementary material.

An estimator of g is obtained by replacing the θ_k in (3.1) 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

$$\hat{\theta}_k = \frac{1}{\varsigma_d} \sum_{\substack{u,v \in X_W \\ r_{\min} < \|u-v\| < r_{\min}+R}}^{\neq} \frac{\phi_k(\|v-u\| - r_{\min}) w(\|v-u\| - r_{\min})}{\rho(u)\rho(v)\|v-u\|^{d-1}|W \cap W_{v-u}|}, \tag{3.2}$$

which is unbiased by the second order Campbell formula, see Lemma 1 below. 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

$$\hat{g}_o(r;b) = \sum_{k=1}^{\infty} b_k \hat{\theta}_k \phi_k(r - r_{\min}), \tag{3.3}$$

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

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

The factor $\|v-u\|^{d-1}$ in (3.2) can 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 the proof of Lemma 1 (see Sec. S2.1 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, two points in X cannot be closer than some $\delta > 0$, this is trivially satisfied. Another example is a determinantal point process (Lavancier, Møller and Rubak (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 (3.3) has the mean integrated squared error

$\text{MISE}(\hat{g}_o, w)$

$$= \varsigma_d \int_{r_{\min}}^{r_{\min}+R} \mathbb{E}\big\{\hat{g}_o(r;b) - g(r)\big\}^2 w(r - r_{\min}) \mathrm{d}r = \varsigma_d \sum_{k=1}^{\infty} \mathbb{E}(b_k \hat{\theta}_k - \theta_k)^2$$

$$= \varsigma_d \sum_{k=1}^{\infty} \big[b_k^2 \mathbb{E}\{(\hat{\theta}_k)^2\} - 2b_k\theta_k^2 + \theta_k^2\big]. \quad (3.4)$$

Each term in (3.4) is minimized with b_k equal to (cf. Hall (1987))

$$b_k^* = \frac{\theta_k^2}{\mathbb{E}\{(\hat{\theta}_k)^2\}} = \frac{\theta_k^2}{\theta_k^2 + \mathbb{V}\mathrm{ar}(\hat{\theta}_k)}, \quad k \geq 0, \quad (3.5)$$

leading to the minimal value $\varsigma_d \sum_{k=1}^{\infty} b_k^* \mathbb{V}\mathrm{ar}(\hat{\theta}_k)$ 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 (3.3) 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 $b_k(\psi) = \mathbb{I}(k \leq K)$. A more refined scheme is $b_k(\psi) = \mathbb{I}(k \leq K)\hat{b}_k^*$ where $\hat{b}_k^* = \widehat{\theta_k^2}/(\hat{\theta}_k)^2$ is an estimate of the optimal smoothing coefficient b_k^* given in (3.5). Here $\widehat{\theta_k^2}$ is the asymptotically unbiased estimator of θ_k^2 derived in Sec. 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 $b_k(\psi) = \mathbb{I}(k \leq K)/(1 + c_1 k^{c_2})$. In practice we choose the smoothing parameter ψ by minimizing an estimate of the mean integrated squared error, see Sec. 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 is

$$\tilde{g}_o(r;b) = 1 + \sum_{k=1}^{\infty} b_k \hat{\vartheta}_k \phi_k(r - r_{\min})$$

where $\hat{\vartheta}_k = \hat{\theta}_k - \int_0^{r_{\min}+R} \phi_k(r) w(r) \mathrm{d}r$ is an estimator of

$$\vartheta_k = \int_0^R \{g(r + r_{\min}) - 1\} \phi_k(r) w(r) \mathrm{d}r.$$

Here $\mathbb{V}\text{ar}(\hat{\vartheta}_k) = \mathbb{V}\text{ar}(\hat{\theta}_k)$ and $\tilde{g}_o(r;b) - \mathbb{E}\{\tilde{g}_o(r;b)\} = \hat{g}_o(r;b) - \mathbb{E}\{\hat{g}_o(r;b)\}$. These identities imply that the results regarding consistency and asymptotic normality established for $\hat{g}_o(r;b)$ in Sec. 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

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

where $B_{r_{\min}}^R = \{h \in \mathbb{R}^d \mid r_{\min} < \|h\| < r_{\min} + R\}$ and

$$e_n(h) = \frac{|W_n \cap (W_n)_h|}{|W_n|}. \tag{4.1}$$

Further,

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

where

$$\varphi_n(h, r) = \sum_{k=1}^{K_n} b_k(\psi_n) \phi_k(\|h\| - r_{\min}) \phi_k(r - r_{\min}). \tag{4.2}$$

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

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

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

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) (see Sec. S2.1 of the supplementary material for the proof).

Lemma 1. *For $k \geq 1$, $\mathbb{E}(\hat{\theta}_{k,n}) = \theta_{k,n}$. Moreover, $\mathbb{V}\mathrm{ar}(\hat{\theta}_{k,n}) \leq C_1/|W_n|$ for some $0 < C_1 < \infty$, if the following conditions hold.*

- $V1$ *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*

$$\sup_{h_1, h_2 \in B^R_{r_{\min}}} \int_{\mathbb{R}^d} \left| g^{(4)}(h_1, h_3, h_2 + h_3) - g(h_1)g(h_2) \right| \mathrm{d}h_3 \leq C_4.$$

The first part of V2 is needed to ensure finite variances of the $\hat{\theta}_{k,n}$ and is discussed in detail in Sec. 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 Sec. 4.3 and for estimation of θ_k^2 in Sec. 5.

Regarding the smoothing scheme, we assume

- S1 $B = \sup_{k,\psi} \left| b_k(\psi) \right| < \infty$ and for all ψ, $\sum_{k=1}^{\infty} \left| b_k(\psi) \right| < \infty$.
- S2 $\psi_n \to \psi^*$ for some ψ^*, and $\lim_{\psi \to \psi^*} \max_{1 \leq k \leq m} \left| b_k(\psi) - 1 \right| = 0$ for all $m \geq 1$.
- S3 $|W_n|^{-1} \sum_{k=1}^{\infty} \left| b_k(\psi_n) \right| \to 0$.

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

The following result is proved in Sec. S2.3 of the supplementary material.

Theorem 1. *Under conditions $V1$-$V3$ and $S1$-$S3$, $\hat{g}_{o,n}$ is a consistent estimator of $g_{o,n}$.*

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

$$S_n = \frac{1}{\varsigma_d |W_n|} \sum_{\substack{u,v \in X_{W_n} \\ v-u \in B^R_{r_{\min}}}}^{\neq} \frac{f_n(v-u)}{\rho(u)\rho(v)e_n(v-u)} \tag{4.3}$$

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

To establish asymptotic normality of estimators of the form (4.3) 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, Paparoditis and Romano (1998)),

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

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

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

To verify asymptotic normality we need assumptions beyond V1 (the conditions V2 and V3 are not needed here 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 $g^{(k)}(h_1, \ldots, h_{k-1}) \le L_1$ for $k = 2, \ldots, 2(2 + \lceil \eta \rceil)$ and all $h_1, \ldots, h_{k-1} \in \mathbb{R}^d$.

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). We will discuss N4 in further detail when applying the general framework to $\hat{\theta}_{k,n}$ and $\hat{g}_{o,n}$. The following result is proved in Sec. S2.4 of the supplementary material.

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

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 then holds 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 Sec. 3.2. For simplicity we here assume $r_{\min} > 0$.

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

$$\{\mathbb{V}\mathrm{ar}(\hat{\theta}_{k,n})\}^{-1/2}(\hat{\theta}_{k,n} - \theta_k) \xrightarrow{\mathscr{D}} N(0,1).$$

For $\hat{g}_{o,n}(r;b) = S_n$, $f_n(h) = \varphi_n(h,r)w(\|h\| - r_{\min})/\|h\|^{d-1}$, where φ_n is defined in (4.2). 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 (4.2) 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 take $\tilde{S}_n := K_n^{-\omega} S_n$ and $\tilde{\tau}_n^2 := K_n^{-2\omega}\tau_n^2$. Then, assuming $\liminf_{n\to\infty} \tilde{\tau}_n^2 > 0$, Thm. 2 gives the asymptotic normality of $\tilde{\tau}_n^{-1}|W_n|^{1/2}\{\tilde{S}_n - \mathbb{E}(\tilde{S}_n)\}$ which is $\tau_n^{-1}|W_n|^{1/2}\{S_n - \mathbb{E}(S_n)\}$.

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 that* $r_{\min} > 0$. *Then, for* $r \in (r_{\min}, r_{\min} + R)$,

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

In case of the simple smoothing scheme $b_k(\psi_n) = \mathbb{I}(k \le K_n)$, 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 Sec. S3 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

$$\varsigma_d I(\psi) = \mathrm{MISE}(\hat{g}_o, w) - \int_{r_{\min}}^{r_{\min}+R} \{g(r)-1\}^2 w(r)\mathrm{d}r$$
$$= \sum_{k=1}^{K} \left[b_k(\psi)^2 \mathbb{E}\{(\hat{\theta}_k)^2\} - 2b_k(\psi)\theta_k^2 \right]. \tag{5.1}$$

In practice we must replace (5.1) by an estimate. Let $\widehat{\theta_k^2}$ be

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

The estimator $\widehat{\theta_k^2}$ is obtained from $(\hat{\theta}_k)^2$ by retaining only terms in which 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. The proof of the following is given in Sec. S2.2 of the supplementary material.

Lemma 2. *Assuming condition* V3, $\widehat{\theta_{k,n}^2}$ *is an asymptotically unbiased estimator of* θ_k^2.

Thus

$$\hat{I}(\psi) = \sum_{k=1}^{K} \left\{ b_k(\psi)^2 (\hat{\theta}_k)^2 - 2b_k(\psi)\widehat{\theta_k^2} \right\} \tag{5.2}$$

is an asymptotically unbiased estimator of (5.1). Moreover, (5.2) is equivalent to a slight modification of Guan (2007a)'s criterion (2.3):

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

For the simple smoothing scheme $b_k(K) = \mathbb{I}(k \le K)$, (5.2) reduces to

$$\hat{I}(K) = \sum_{k=1}^{K} \left\{ (\hat{\theta}_k)^2 - 2\widehat{\theta_k^2} \right\} = \sum_{k=1}^{K} (\hat{\theta}_k)^2 (1 - 2\hat{b}_k^*), \tag{5.3}$$

where $\hat{b}_k^* = \widehat{\theta_k^2}/(\hat{\theta}_k)^2$ is an estimator of b_k^* in (3.5).

In practice, uncertainties of $\hat{\theta}_k$ and $\widehat{\theta_k^2}$ lead to numerical instabilities in the minimization of (5.2) with respect to ψ. To obtain a numerically stable procedure we first determine K as

$$\hat{K} = \inf\left(2 \le k \le K_{\max} : (\hat{\theta}_{k+1})^2 - 2\widehat{\theta^2_{k+1}} > 0\right) = \inf\left(2 \le k \le K_{\max} : \hat{b}^*_{k+1} < \frac{1}{2}\right). \quad (5.4)$$

Here $\hat{K}$ is the first local minimum of (5.3) larger than 1 and smaller than an upper limit $K_{\max}$ that we chose to be 49 in our applications. This choice of K is also used for the refined and the Wahba smoothing schemes. For the refined smoothing scheme we take $b_k = \mathbb{I}(k \le \hat{K})\hat{b}^*_k$. For the Wahba smoothing scheme $b_k = \mathbb{I}(k \le \hat{K})/(1 + \hat{c}_1 k^{\hat{c}_2})$, where $\hat{c}_1$ and $\hat{c}_2$ minimize $\sum_{k=1}^{\hat{K}} \{(\hat{\theta}_k)^2/(1 + c_1 k^{c_2})^2 - 2\widehat{\theta^2_k}/(1 + c_1 k^{c_2})\}$ over $c_1 > 0$ and $c_2 > 1$.

The expected value of $\hat{K}$ in (5.4) depends on how fast the coefficients θ_k tend to zero. The smoothness of $g(r)$ and its behavior near boundaries of the interval $(r_{\min}, r_{\min} + R)$ determine the decay rate of the coefficients θ_k. For example, if $g(r)$ is twice differentiable and $\int_0^R |g''(r_{\min} + r)|w(r)\mathrm{d}r < \infty$ then $\theta_k = O(k^{-2})$. To ensure that θ_k decreases faster than k^{-2} we need to assume that $g(r)$ has three derivatives and $g'(r_{\min}) = g'(r_{\min} + R) = 0$ (see Efromovich (2008, p. 32)).

Finally, as in the case of probability density estimation (Efromovich (2010)), the orthogonal series estimator may result in negative values of $\hat{g}_o(r; b)$ for some $r_{\min} < r < r_{\min} + R$. Since $g(r) \ge 0$ for any $r \ge 0$, a natural solution is to use $\hat{g}^+_o(r; b) = \max\{0, \hat{g}_o(r; b)\}$ in practice.

6. Simulation Study

We compared 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 $\rho = 100$; here the expected number of points (100 for $W = [0, 1]^2$ and 400 for $W = [0, 2]^2$) are the same for all point processes. More specifically, we took $n_{\mathrm{sim}} = 1{,}000$ 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 $\omega = 0.01845$, see Jalilian, Guan and Waagepetersen (2013)), and a determinantal point process with pair correlation function $g(r) = 1 - \exp(-2(r/\alpha)^2)$ and $\alpha = 0.056$. The parameters of the processes were chosen such that they had the same scale of pairwise correlation; $|g(r) - 1|/|g(0) - 1| \le 0.01$ for $r \ge 0.08$. The pair correlation functions of these point processes are shown in Figure 1.

For each realization, $g(r)$ was estimated for r in $(r_{\min}, r_{\min} + R)$, with $r_{\min} =$

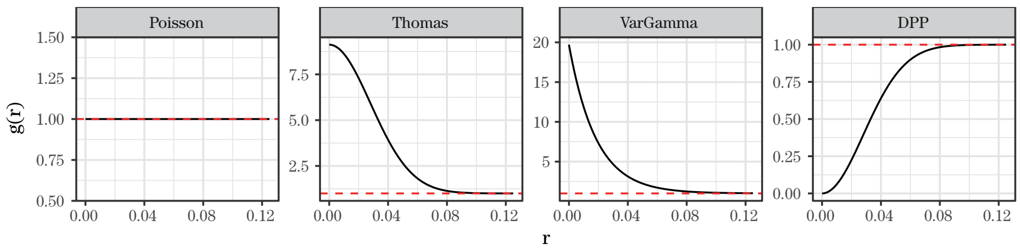

Figure 1. Pair correlation functions for the point processes considered in the simulation study.

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 $b = 0.15/\sqrt{\hat{\rho}}$ was used for $\hat{g}_k(r;b)$ and $\hat{g}_d(r;b)$ while the bandwidth of $\hat{g}_c(r;b)$ was chosen by minimizing Guan (2007a)'s estimate (2.3) of the mean integrated squared error. For the orthogonal series estimator, we considered both the cosine and the Fourier-Bessel bases with simple, refined or Wahba smoothing schemes. For the Fourier-Bessel basis we used the modified orthogonal series estimator described in Sec. 3.4. The parameters for the smoothing scheme were chosen according to Sec. 5.

From the simulations we estimated the mean integrated squared error (2.2) 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 considered the kernel estimator $\hat{g}_k$ as the baseline estimator and compared 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 outperformed or did as well as the kernel estimators both at small lags and over all lags. The detailed conclusions depended on whether the non-repulsive Poisson, Thomas and Var Gamma processes or the repulsive determinantal process was considered. Orthogonal-Bessel with refined or Wahba smoothing was 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 was between or better than the performance of the kernel estimators for Poisson, Thomas and Var Gamma and as good as the best kernel estimator for determinantal. The above conclusions were stable over the three R values considered. For $W = [0,2]^2$ (see Figure S1 in the supplementary material) the conclusions

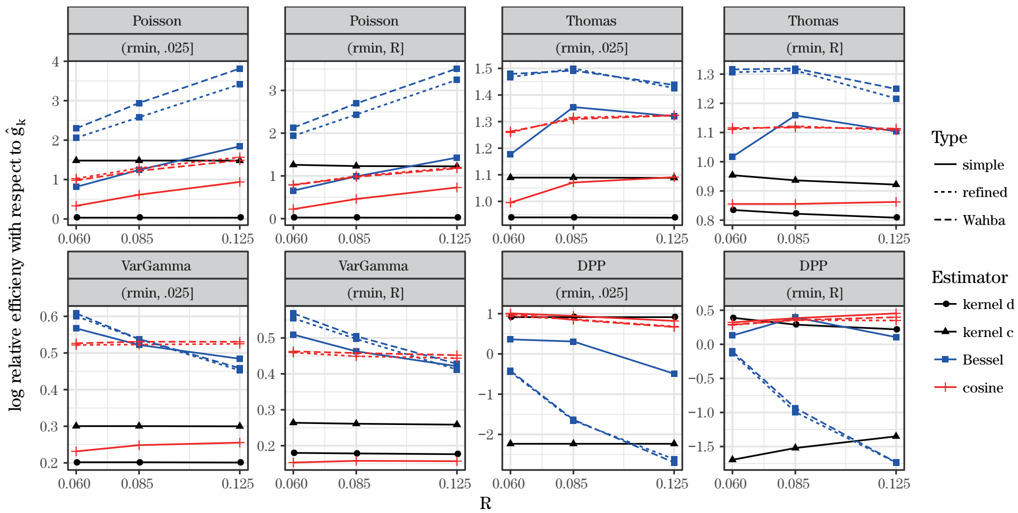

Figure 2. Plots of log relative efficiencies for small lags $(r_{\min}, 0.025]$ and all lags $(r_{\min}, R]$, $R = 0.06, 0.085, 0.125$, and $W = [0,1]^2$. Lines serve to ease visual interpretation.

were similar but with clearer superiority of the orthogonal series estimators for Poisson and Thomas. For Var Gamma the performance of $\hat{g}_c$ was similar to the orthogonal series estimators. For determinantal and $W = [0,2]^2$, $\hat{g}_c$ was better than orthogonal-Bessel-refined/Wahba but still inferior to orthogonal-Bessel-simple and orthogonal-cosine. Figure 3 and Figure S2 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 cutoff K in Table S1 tends to be larger for the Variance Gamma process than for the other processes because the boundary condition $g'(r_{\min}) = 0$ holds for all processes but the Variance Gamma. See also the difference between the Variance Gamma pair correlation function and the mean of the orthogonal-Bessel and orthogonal-cosine estimates in Figure 3. Table S1 also shows that a larger K tends to be needed for cosine than for Fourier-Bessel. 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 Sec. 4 we considered 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

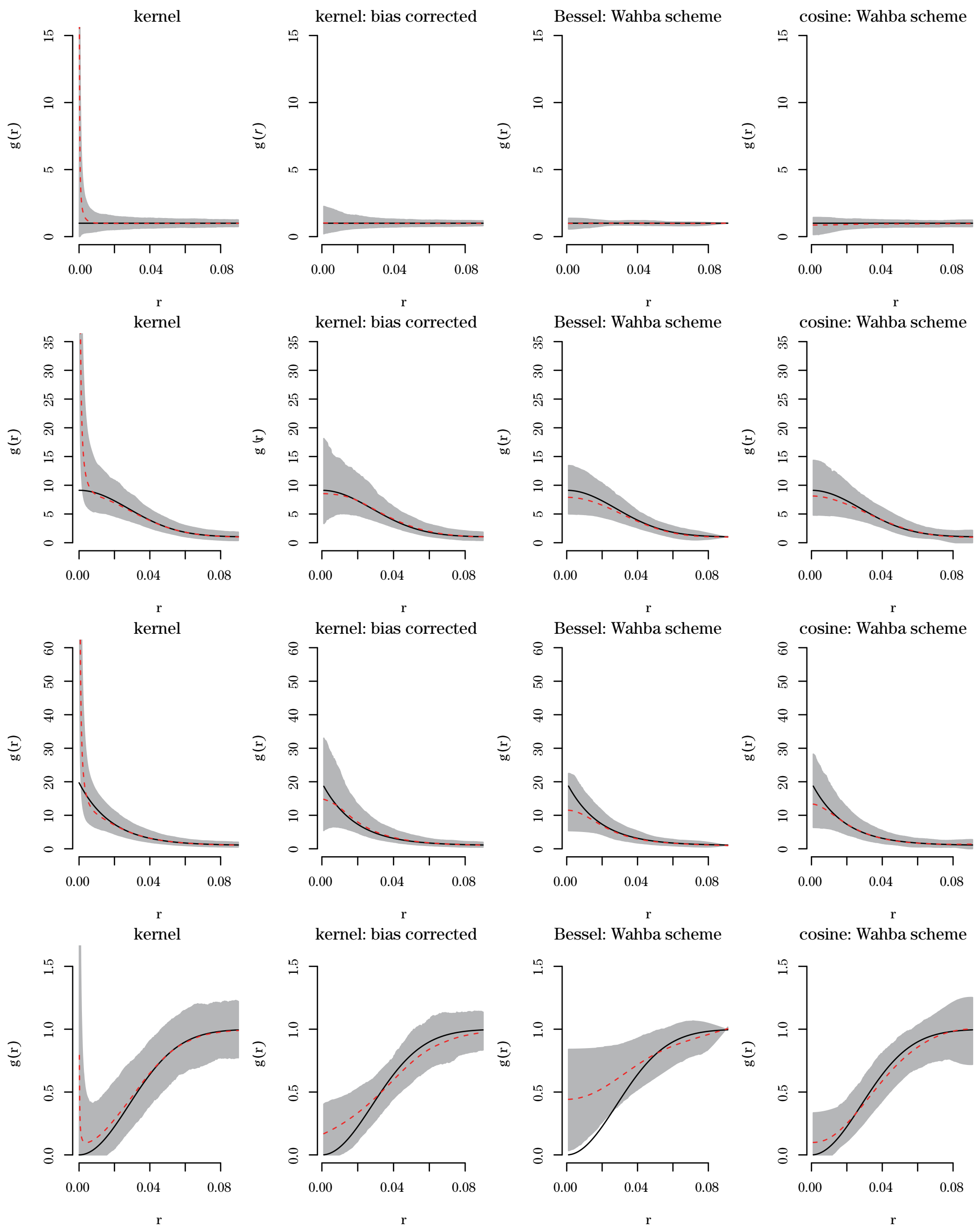

Figure 3. True pair correlation function (solid line), Monte Carlo mean (dashed lines) and 95% pointwise probability interval (grey area) of estimates based on 1,000 simulations from the Poisson (first row), Thomas (second row), Variance Gamma (third row) and determinantal (fourth row) point processes on $W = [0,1]^2$.

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	$g(r)$	$\hat{\mathbb{E}}\{\hat{g}_o(r)\}$	$[\hat{\mathbb{V}\text{ar}}\{\hat{g}_o(r)\}]^{1/2}$	$\hat{\text{S}}(\hat{g}_o(r))$	$\hat{\text{K}}(\hat{g}_o(r))$
W_1	0.025	3.972	3.961	0.923	1.145	5.240
	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
	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
	0.1	1.2187	1.2017	0.0951	0.2913	2.9573

to $W = [0,1]^2$ and $W = [0,2]^2$, $W = [0,3]^2$ was 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 in the supplementary material. 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* (3,299 trees) species in the 1995 census for the 1,000m × 500m Barro Colorado Island plot (Hubbell and Foster (1983), Condit (1998)). To estimate the intensity function of each species, we used a log-linear regression model depending on covariates related to soil conditions and topographical variables. The regression parameters were estimated using the quasi-likelihood approach in Guan, Jalilian and Waagepetersen (2015). The point patterns and fitted intensity functions are shown in Figure S4 in the supplementary material.

The pair correlation function of each species was then estimated using the bias corrected kernel estimator $\hat{g}_c(r;b)$ with b determined by minimizing (2.3) and the orthogonal 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 (5.4); see Figure 4.

For *Lonchocarpus* the three estimates were quite similar, while for *Acalypha* and *Capparis* the estimates deviated markedly for small lags, then similar for lags greater than respectively 2 and 8 meters. For *Capparis* and the cosine basis, the number of selected coefficients coincided with the chosen upper limit 49 for

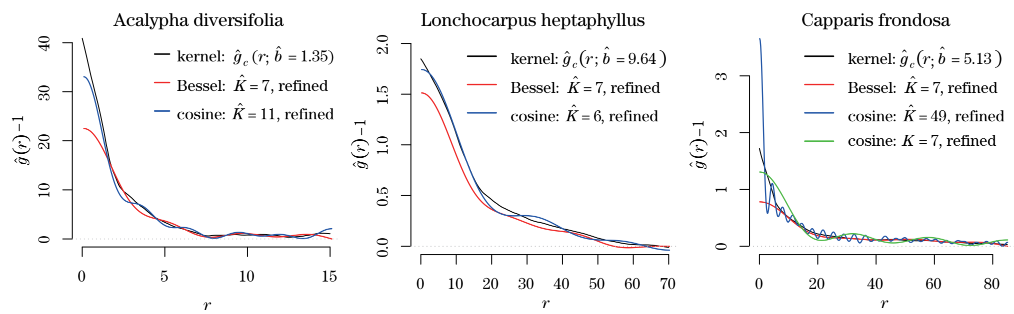

Figure 4. Estimated pair correlation functions for tropical rain forest trees.

the number of coefficients. The cosine estimate displays oscillations which appear to be artifacts of using high frequency components of the cosine basis. The function (5.3) decreases very slowly after $K = 7$ (Figure S5 in the supplementary material) so we also tried the cosine estimate with $K = 7$ which gives a more reasonable estimate.

8. Conclusion

The orthogonal series estimators introduced in this paper have a sound theoretical basis, being consistent and asymptotically normal under reasonable conditions. Our simulations suggest that they perform very well compared with the usual kernel estimators. The methodology is by no means confined to the considered Fourier-Bessel and cosine bases. A topic of future research is to consider such other bases as Laguerre polynomials and wavelet bases.

Acknowledgement

We thank the referees for valuable comments. Rasmus Waagepetersen was supported by The Danish Council for Independent Research — Natural Sciences, grant DFFV 7014-00074 “Statistics for point processes in space and beyond”, and by the ”Centre for Stochastic Geometry and Advanced Bioimaging”, funded by grant 8721 from the Villum Foundation.

Supplementary Material

Supplementary material available at *Statistica Sinica* online provides proofs and further material regarding 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, Boca Raton.

Baddeley, A. J., Møller, J. and Waagepetersen, R. (2000). Non- and semiparametric 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**, 1097–1125.

Coeurjolly, J.-F., Møller, J. and Waagepetersen, R. (2017). A tutorial on palm distributions for spatial point processes. *International Statistical Review* **85**, 404–420.

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

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

Efromovich, S. (2008). *Nonparametric Curve Estimation: Methods, Theory, and Applications*. Springer Science & Business Media.

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**, 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**, 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**, 677–697.

Hall, P. (1987). Cross-validation and the smoothing of orthogonal series density estimators. *Journal of Multivariate Analysis* **21**, 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 *Tropical Rain Forest: Ecology and Management* (Edited by S. L. Sutton, T. C. Whitmore, and A. C. Chadwick), 25–41. Blackwell Scientic Publications, Oxford.

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

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

Jalilian, A., Guan, Y. and Waagepetersen, R. (2013). Decomposition of variance for spatial cox processes. *Scandinavian Journal of Statistics* **40**, 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**, 761–779.

Møller, J. and Waagepetersen, R. P. (2003). *Statistical Inference and Simulation for Spatial Point Processes*. Boca Raton: Chapman and Hall/CRC.

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**, 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**, 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**, 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**, 463–474.

Department of Statistics, Razi University, Bagh-e-Abrisham, Kermanshah 67149-67346, Iran.

E-mail: jalilian@razi.ac.ir

Department of Management Science, University of Miami, Coral Gables, Florida 33124–6544, USA.

E-mail: yguan@bus.miami.edu

Department of Mathematical Sciences, Aalborg University, Fredrik Bajersvej 7G, DK-9220 Aalborg, Denmark.

E-mail: rw@math.aau.dk

(Received March 2017; accepted September 2017)