Second-order variational equations for spatial point processes with a view to pair correlation function estimation

Coeurjolly, Jean François; Cuevas-Pacheco, Francisco; Waagepetersen, Rasmus

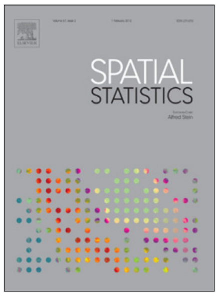
Published in:
Spatial Statistics

DOI (link to publication from Publisher):
10.1016/j.spasta.2019.03.001

Publication date:
2019

Document Version
Accepted author manuscript, peer reviewed version

Accepted Manuscript



Second-order variational equations for spatial point processes with a view to pair correlation function estimation

Jean-François Coeurjolly, Francisco Cuevas-Pacheco,
Rasmus Waagepetersen

PII: S2211-6753(19)30002-8
DOI: https://doi.org/10.1016/j.spasta.2019.03.001
Reference: SPASTA 350

To appear in: *Spatial Statistics*

Received date : 3 January 2019
Accepted date : 8 March 2019

Second-order variational equations for spatial point processes with a view to pair correlation function estimation

Jean-François Coeurjolly[a,*], Francisco Cuevas-Pacheco[b], Rasmus Waagepetersen[b]

[a] *Department of Mathematics, Universiteé du Québec à Montreal (UQAM), Canada*
[b] *Department of Mathematical Sciences, Aalborg University, Denmark*

Abstract

Second-order variational type equations for spatial point processes are established. In case of log linear parametric models for pair correlation functions, it is demonstrated that the variational equations can be applied to construct estimating equations with closed form solutions for the parameter estimates. This result is used to fit orthogonal series expansions of log pair correlation functions of general form.

Keywords: estimating equation, non-parametric estimation, orthogonal series expansion, pair correlation function, variational equation.

1. Introduction

Spatial point processes are models for sets of random locations of possibly interacting objects. Background on spatial point processes can be found in Møller and Waagepetersen (2004), Illian et al. (2008) or Baddeley et al. (2015) which gives both an accessible introduction as well as details on implementation in the R package `spatstat`. Moments of counts of objects for spatial point processes are typically expressed in terms of so-called joint intensity functions or Papangelou conditional intensity functions which are defined via the Campbell or Georgii-Nguyen-Zessin equations (see the aforementioned references or the concise review of intensity functions and Campbell formulae in Section 2). In this paper we consider a third type of equation called variational equations.

A key feature of variational equations compared to Campbell and Georgii-Nguyen-Zessin equations is that they are formulated in terms of the gradient of the log intensity or conditional intensity function rather than the (conditional)

*Corresponding author
Email addresses: `coeurjolly.jean-francois@uqam.ca` (Jean-François Coeurjolly), `francisco@math.aau.dk` (Francisco Cuevas-Pacheco), `rw@math.aau.dk` (Rasmus Waagepetersen)

intensity itself. Variational equations were introduced for parameter estimation in Markov random fields by Almeida et al. (1993). The authors suggested the terminology 'variational' due to the analogy between the derivation of their estimating equation and the variational Euler-Lagrange equations in partial differential equations. The resulting equation consisted in an equilibrium equation involving the gradient of the log conditional probability of the Markov random field. Later, Baddeley and Dereudre (2013) obtained variational equations for Gibbs point processes and exploited them to infer a log-linear parametric model of the conditional intensity function. Coeurjolly and Møller (2014) established a first-order variational equation for general spatial point processes and used it to estimate parameters in a log-linear parametric model for the intensity function.

The first contribution of this paper is to establish second-order variational equations. The second-order properties of a spatial point process are characterized by the so-called pair correlation function which is a normalized version of the second-order joint intensity function. We assume that the pair correlation function is translation invariant and also consider the case when it is isotropic. Since the new variational equations are based on the gradient of the log pair correlation function, they take a particularly simple form for pair correlation functions of log-linear form.

Our second contribution is to propose a new non-parametric estimator of the pair correlation function. The classical approach is to use a kernel estimator, see for example Møller and Waagepetersen (2004). More recently, Jalilian et al. (2019) investigated the estimation of the pair correlation function using an orthogonal series expansion. In the setting of their simulation studies, the orthogonal series estimator was shown to be more efficient than the standard kernel estimator. One drawback, however, is that the orthogonal series estimator is not guaranteed to be non-negative. We therefore propose to use our second-order variational equations to estimate coefficients in an orthogonal series expansion of the log pair correlation function. This ensures that the resulting pair correlation function estimator is non-negative. We compare our new estimator with the previous ones in a simulation study and also illustrate its use on real datasets.

2. Background and main results

2.1. Spatial point processes

Throughout this paper we let $\mathbf{X}$ be a spatial point process defined on $\mathbb{R}^d$. That is, $\mathbf{X}$ is a random subset of $\mathbb{R}^d$ with the property that the intersection of $\mathbf{X}$ with any bounded subset of $\mathbb{R}^d$ is of finite cardinality. The joint intensity functions $\rho^{(k)}$, $k \geq 1$, are characterized (when they exist) by the Campbell formulae (equations) (see for example Møller and Waagepetersen, 2004): for any $h : (\mathbb{R}^d)^k \to \mathbb{R}^+$ (with $\mathbb{R}^+$ the non-negative real numbers)

$$\mathbb{E} \sum_{u_1,\dots,u_k \in \mathbf{X}}^{\neq} h(u_1,\dots,u_k) = \int \cdots \int h(u_1,\dots,u_k)\rho^{(k)}(u_1,\dots,u_k)\mathrm{d}u_1 \cdots \mathrm{d}u_k. \tag{1}$$

More intuitively, for any pairwise distinct points $u_1,\dots,u_k \in \mathbb{R}^d$, $\rho^{(k)}(u_1,\dots,u_k)\mathrm{d}u_1 \cdots \mathrm{d}u_k$ is the probability that for each $i = 1,\dots,k$, $\mathbf{X}$ has a point in an infinitesimally small region around u_i with volume $\mathrm{d}u_i$. The intensity function ρ corresponds to the case $k = 1$, i.e. $\rho = \rho^{(1)}$. The pair correlation function is obtained by normalizing the second-order joint intensity $\rho^{(2)}$:

$$g(u,v) = \frac{\rho^{(2)}(u,v)}{\rho(u)\rho(v)} \tag{2}$$

for pairwise distinct u, v and where $g(u,v)$ is set to 0 if $\rho(u)$ or $\rho(v)$ is zero. Intuitively, $g(u,v) > 1$ [$g(u,v) < 1$] means that presence of a point at u increases [decreases] the probability of observing a further point at v and vice versa. We assume that $\mathbf{X}$ is observed on some bounded domain $W \subset \mathbb{R}^d$ with volume $|W| > 0$ and without loss of generality we assume that $\rho(u) > 0$ for all $u \in W$ (otherwise we just replace W by $\{u \in W \mid \rho(u) > 0\}$ provided the latter set has positive volume).

We will always assume that $\mathbf{X}$ is second-order intensity reweighted stationary (Baddeley et al., 2000), meaning that its pair correlation function g is invariant by translations. We then, with an abuse of notation, write $g(v-u)$ for $g(u,v)$ for any $u, v \in \mathbb{R}^d$. We will also consider the case of an isotropic pair correlation function in which case $g(v-u)$ depends only on the distance $\|v-u\|$.

For the presentation of the second-order variational type equation in the next section some additional notation is needed. For a function $h : \mathbb{R}^d \to \mathbb{R}$ which is differentiable on $\mathbb{R}^d$, we denote by

$$\nabla h(w) = \left\{ \frac{\partial h}{\partial w_1}(w), \dots, \frac{\partial h}{\partial w_d}(w) \right\}^\top, \quad w \in \mathbb{R}^d$$

the gradient vector with respect to the d coordinates. The inner product is denoted by $a \cdot b$ and for $h : \mathbb{R}^d \to \mathbb{R}^d$, a multivariate function such that each component is differentiable on $\mathbb{R}^d$, we define the divergence operator by

$$\operatorname{div} h(w) = \sum_{i=1}^{d} \frac{\partial h_i}{\partial w_i}(w).$$

2.2. Second-order variational equations

In this section, we present in Theorem 1 and Theorem 2 our new second-order variational equations. The prominent feature of the equations is that they are given in terms of expectations of random sums where the sums only depend on the pair correlation function through its gradient (Theorem 1) or, in the isotropic case, its derivative (Theorem 2). This allows us to construct in Section 3 closed form estimators of pair correlation functions of log linear form.

Theorem 1. *Assume* $\mathbf{X}$ *is second-order intensity reweighted stationary. Let* $h : \mathbb{R}^d \to \mathbb{R}^d$ *be a componentwise continuously differentiable function on* $\mathbb{R}^d$. *Assume that* g *is continuously differentiable on* $\mathbb{R}^d$, *that* $\|h\|\|\nabla g\| \in L^1(\mathbb{R}^d)$, *and that there exists a sequence of increasing bounded domains* $(B_n)_{n\geq 1}$ *such that* $B_n \to \mathbb{R}^d$ *as* $n \to \infty$, *with piecewise smooth boundary* ∂B_n *and such that*

$$\lim_{n\to\infty} \int_{\partial B_n} g(w)h(w) \cdot \nu(\mathrm{d}w) = 0 \tag{3}$$

where ν *stands for the outer normal measure to* ∂B_n. *Then*

$$\mathbb{E}\left\{ \sum_{u,v\in\mathbf{X}\cap W}^{\neq} e(u,v)\nabla \log g(v-u) \cdot h(v-u) \right\} = - \mathbb{E}\left\{ \sum_{u,v\in\mathbf{X}\cap W}^{\neq} e(u,v)\mathrm{div}\, \mathrm{h}(v-u) \right\}, \tag{4}$$

where $e : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^+$ *denotes the function* $e(u,v) = \{\rho(u)\rho(v)|W \cap W_{v-u}|\}^{-1}$ *for any* $u, v \in \mathbb{R}^d$ *and where* W_w *denotes the domain* W *translated by* $w \in \mathbb{R}^d$.

The proof of Theorem 1 is given in Appendix A. We note that condition (3) is in particular satisfied if the function h is compactly supported.

We next consider the case where the pair correlation function is isotropic, i.e. for any $u, v \in \mathbb{R}^d$ there exists $g_0 : \mathbb{R}^+ \to \mathbb{R}^+$ such that $g(u,v) = g(v-u) = g_0(\|v-u\|)$.

Theorem 2. *Assume* $\mathbf{X}$ *is second-order intensity reweighted stationary with isotropic pair correlation function* g_0. *Let* $h : \mathbb{R}^+ \to \mathbb{R}$ *be continuously differentiable on* $\mathbb{R}^+$. *Assume that* g_0 *is continuously differentiable on* $\mathbb{R}^+$ *and that either*

$$t \mapsto h(t)g_0'(t) \in L^1(\mathbb{R}^+) \quad \text{and} \quad \lim_{n\to\infty} \{g_0(n)h(n) - g_0(0)h(0)\} = 0 \tag{5}$$

or

$$t \mapsto t^{d-1}h(t)g_0(t) \in L^1(\mathbb{R}^+) \quad \text{and} \quad \lim_{n\to\infty} \{n^{d-1}g_0(n)h(n) - g_0(0)h(0)\mathbf{1}(d=1)\} = 0. \tag{6}$$

Then we have the two following cases. If (5) *is assumed,*

$$\mathbb{E}\left\{ \sum_{u,v\in\mathbf{X}\cap W}^{\neq} \frac{e(u,v)}{\|v-u\|^{d-1}} h(\|v-u\|)(\log g_0)'(\|v-u\|) \right\} = - \mathbb{E}\left\{ \sum_{u,v\in\mathbf{X}\cap W}^{\neq} \frac{e(u,v)}{\|v-u\|^{d-1}} h'(\|v-u\|) \right\}, \tag{7}$$

where $e(u,v) = \{\rho(u)\rho(v)|W \cap W_{v-u}|\}^{-1}$ *for any* $u, v \in \mathbb{R}^d$*. Instead, if* (6) *is assumed,*

$$\mathbb{E}\Bigg\{ \sum_{u,v \in \mathbf{X}\cap W}^{\neq} e(u,v) h(\|v-u\|)(\log g_0)'(\|v-u\|) \Bigg\} = - \mathbb{E}\left[\sum_{u,v \in \mathbf{X}\cap W}^{\neq} e(u,v) \left\{ (d-1)\frac{h(\|v-u\|)}{\|v-u\|} + h'(\|v-u\|) \right\} \right]. \tag{8}$$

The proof of Theorem 2 is given in Appendix B. We stress that the derivatives involved in Theorem 2 are derivatives with respect to $t \ge 0$. Like for Theorem 1, conditions (5) and (6) are in particular satisfied if h is compactly supported in $(0,\infty)$.

Remark 1. In Theorem 1 and Theorem 2, the factor $|W \cap W_{v-u}|^{-1}$ in $e(u,v)$ is a so-called edge correction factor that allows us to rewrite the expectations (4), (7) and (8) as integrals that do not depend on $|W|$, see the proofs in the appendices. Other edge corrections (p. 188-189 in Illian et al., 2008) like minus sampling or, in the case of Theorem 2, the isotropic edge correction, could be used as well.

2.3. Sensitivity matrix

In the next section we use empirical versions of (7) and (8) to construct estimating functions for a parametric model of an isotropic pair correlation function g_0 depending on a K-dimensional parameter $\boldsymbol{\beta}$, $K \ge 1$. We here investigate the expression for the associated sensitivity matrices.

Consider functions $h_1, \dots, h_K$ all fulfilling (5) and possibly depending on $\boldsymbol{\beta}$. By stacking the K equations obtained by applying these functions for $h_1, \dots, h_K$ in (7) we obtain the estimating function

$$\sum_{u,v \in \mathbf{X}\cap W}^{\neq} \frac{e(u,v)}{\|v-u\|^{d-1}} \mathbf{h}(\|v-u\|)(\log g_0)'(\|v-u\|) + \sum_{u,v \in \mathbf{X}\cap W}^{\neq} \frac{e(u,v)}{\|v-u\|^{d-1}} \mathbf{h}'(\|v-u\|) \tag{9}$$

where $\mathbf{h}$ and $\mathbf{h}'$ are vector functions with components h_i and h_i'. The sensitivity matrix is obtained as the expectation of the negated derivative (with respect to $\boldsymbol{\beta}$) of (9). After applying (7) once again after differentiation we obtain the sensitivity matrix

$$S(\boldsymbol{\beta}) = -\mathbb{E} \sum_{u,v \in \mathbf{X}\cap W}^{\neq} \frac{e(u,v)}{\|v-u\|^{d-1}} \mathbf{h}(\|v-u\|) \frac{\mathrm{d}}{\mathrm{d}\boldsymbol{\beta}^\top} (\log g_0)'(\|v-u\|).$$

Applying the Campbell theorem and converting to polar coordinates, we obtain

$$S(\boldsymbol{\beta}) = -\varsigma_d \int_0^\infty \mathbf{h}(t) \left[\frac{\mathrm{d}}{\mathrm{d}\boldsymbol{\beta}^\top} (\log g_0)'(t) \right] g_0(t) \mathrm{d}t,$$

where ς_d is the surface area of the d-dimensional unit ball. In case of (8) we obtain a similar expression,

$$S(\boldsymbol{\beta}) = -\varsigma_d \int_0^\infty \mathbf{h}(t) \left[\frac{\mathrm{d}}{\mathrm{d}\boldsymbol{\beta}^\top} (\log g_0)'(t) \right] g_0(t) t^{d-1} \mathrm{d}t.$$

By choosing $\mathbf{h}(t) = -\psi(t) \frac{\mathrm{d}}{\mathrm{d}\boldsymbol{\beta}} (\log g_0)'(t)$ for some real function ψ, $S(\boldsymbol{\beta})$ becomes at least positive semi-definite.

3. Estimation of log linear pair correlation function

We now consider the estimation of an isotropic pair correlation function of the form

$$\log g_0(t) = \boldsymbol{\beta}^\top \mathbf{r}(t) = \boldsymbol{\beta}^\top \{r_1(t), \dots, r_K(t)\}^\top \tag{10}$$

where the functions $r_k : \mathbb{R}^+ \to \mathbb{R}$, $k = 1, \dots, K$ are known. Following Section 2.3, the idea is to apply Theorem 2 K times to functions h_i, $i = 1, \dots, K$, of the form $h_i(t) = -\psi(t) \frac{\partial}{\partial \boldsymbol{\beta}_i} (\log g_o)'(t) = -\psi(t) r_i'(t)$ where the function $\psi : \mathbb{R}^+ \to \mathbb{R}$ will be justified and specified later. It is then remarkable that we obtain a simple estimating equation of the form $\mathbf{A}\boldsymbol{\beta} + \mathbf{b} = 0$. The sensitivity matrix discussed in Section 2.3 is $S(\boldsymbol{\beta}) = -\mathbb{E}\mathbf{A}$. Provided $\mathbf{A}$ is invertible we obtain the explicit solution

$$\hat{\boldsymbol{\beta}} = -\mathbf{A}^{-1}\mathbf{b}. \tag{11}$$

The matrix $\mathbf{A}$ and the vector $\mathbf{b}$ are specified in the following corollary.

Corollary 1. *Let $\psi : \mathbb{R}^+ \to \mathbb{R}$. Assume that ψ and r_k ($k = 1, \dots, K$) are respectively continuously differentiable and twice continuously differentiable on $\mathbb{R}^+$. Assume either that*

$$t \mapsto \|\mathbf{r}'(t)\|^2 \psi(t) \in L^1(\mathbb{R}^+) \text{ and } \lim_{n\to\infty} \psi(n)\mathbf{r}(n)^\top \mathbf{r}'(n) - \psi(0)\mathbf{r}(0)^\top \mathbf{r}'(0) = 0 \tag{12}$$

or

$$\begin{aligned} & t \mapsto t^{d-1} \|\mathbf{r}'(t)\|^2 \psi(t) \in L^1(\mathbb{R}^d) \\ & \text{and } \lim_{n\to\infty} n^{d-1}\psi(n)\mathbf{r}(n)^\top \mathbf{r}'(n) - \psi(0)\mathbf{r}(0)^\top \mathbf{r}'(0) \mathbf{1}(d=1) = 0. \end{aligned} \tag{13}$$

If (12) is assumed we define the (K, K) matrix $\mathbf{A}$ and the vector $\mathbf{b} \in \mathbb{R}^K$ by

$$\mathbf{A} = \sum_{u,v \in \mathbf{X} \cap W}^{\neq} \frac{e(u,v)}{\|v-u\|^{d-1}} \psi(\|v-u\|) \mathbf{r}'(\|v-u\|) \{\mathbf{r}'(\|v-u\|)\}^\top \tag{14}$$

$$\mathbf{b} = \sum_{u,v \in \mathbf{X} \cap W}^{\neq} \frac{e(u,v)}{\|v-u\|^{d-1}} \{\psi'(\|v-u\|)\mathbf{r}'(\|v-u\|) + \psi(\|v-u\|)\mathbf{r}''(\|v-u\|)\} \tag{15}$$

where again the edge effect factor is $e(u,v) = \{\rho(u)\rho(v)|W \cap W_{v-u}|\}^{-1}$ for any $u, v \in \mathbb{R}^d$. Instead, in case of (13), we define

$$\mathbf{A} = \sum_{u,v \in \mathbf{X}\cap W}^{\neq} e(u,v)\psi(\|v-u\|)\mathbf{r}'(\|v-u\|)\{\mathbf{r}'(\|v-u\|)\}^\top \tag{16}$$

$$\mathbf{b} = \sum_{u,v \in \mathbf{X}\cap W}^{\neq} e(u,v)\Big\{(d-1)\frac{\psi(\|v-u\|)\mathbf{r}'(\|v-u\|)}{\|v-u\|} + \psi'(\|v-u\|)\mathbf{r}'(\|v-u\|) + \psi(\|v-u\|)\mathbf{r}''(\|v-u\|)\Big\} \tag{17}$$

Then, the equation

$$\mathbf{A}\boldsymbol{\beta} + \mathbf{b} = 0 \tag{18}$$

is an unbiased estimating equation.

Proof. The proof consists in applying Theorem 2 with $h(t) = -\psi(t)r'_k(t)$ for $k = 1, \dots, K$ and in noticing that $(\log g_0)'(t) = \boldsymbol{\beta}^\top \mathbf{r}'(t) = \mathbf{r}'(t)^\top \boldsymbol{\beta}$. □

We note that if ψ is compactly supported in $[0, \infty)$, then (12) or (13) are always valid assumptions. Another special case is also interesting: let $d > 1$ and $\psi = 1$, then (13) is true if for any $k, l = 1, \dots, K$, $t \mapsto t^{d-1} r'_k(t)^2 \in L^1(\mathbb{R}^d)$ and $\lim_{n\to\infty} n^{d-1} r_k(n) r'_l(n) = 0$. This simple condition is for instance satisfied if the r_k's' are exponential covariance functions.

The results above are for instance applicable to the case of a pair correlation function for a log Gaussian Cox process with covariance function given by a sum of known correlation functions scaled by unknown variance parameters. Assuming a known correlation function is on the other hand quite restrictive. However, any log pair correlation function can be approximated well on a finite interval using a suitable basis function expansion so that we can effectively represent it as a log linear model. We exploit this in Section 4 where we consider the case where the functions r_k are basis functions on a bounded real interval.

Remark 2. In applications of (14)-(15) for $d = 2$ or (16)-(17) for $d \geq 1$ the division by $\|v-u\|^{d-1}$ or $\|v-u\|$ may lead to numerical instability for pairs of close points u and v. This can be mitigated by a proper choice of the function ψ. In the spatial case of $d = 2$ we propose to define $\psi(t) = (t/b)^2(1-(t/b))^2\mathbf{1}(t \in [0, b])$ for some $b > 0$. With this choice of ψ the divisors $\|v-u\|^{d-1} = \|v-u\|$ cancel out preventing very large or infinite variances of (14)-(17).

Remark 3. The quantities (14)-(17) depend on the unknown intensity function. If the intensity function is constant equal to $\rho > 0$ we can multiply (18) by ρ^2 whereby the resulting estimating equation no longer depends on ρ. Thus g_0 can be estimated without estimating ρ. Otherwise, the intensity function has to be estimated first, for instance in a parametric way, see Guan et al. (2015), and plugged into (14)-(17).

4. Variational orthogonal series estimation of the pair correlation function

In this section we consider the estimation of an isotropic pair correlation function g_0 on a bounded interval $[r_{\min}, r_{\min}+R]$, $0 \leq r_{\min} < \infty$ and $0 < R < \infty$, using a series expansion of $\log g_0$. Let $\{\phi_k\}_{k\geq 1}$ denote an orthonormal basis of functions on $[0,R]$ with respect to some weight function $w(\cdot) \geq 0$, i.e. $\int_0^R \phi_k(t)\phi_l(t)w(t)\mathrm{d}t = \delta_{kl}$. Provided $\log g_0$ is square integrable (with respect to $w(\cdot)$) on $[r_{\min}, r_{\min}+R]$, we have the expansion

$$\log g_0(t) = \sum_{k=1}^{\infty} \beta_k \phi_k(t - r_{\min}) \tag{19}$$

where the coefficients β_k are defined by $\beta_k = \int_0^R \log g_0(t + r_{\min})\phi_k(t)w(t)\mathrm{d}t$.

We propose to approximate $\log g_0$ by truncating the infinite sum up to some $K \geq 1$ and obtain estimates $\hat{\beta}_1, \ldots, \hat{\beta}_K$ using (18). The resulting estimate thus becomes

$$\widehat{\log g_{0,K}}(t) = \sum_{k=1}^{K} \hat{\beta}_k \phi_k(t - r_{\min}).$$

In the sequel this estimator is referred to as the variational (orthogonal series) estimator (VSE for short). The approach is related to Zhao (2018) who also considers an estimating equation approach to estimate a pair correlation function of the form (19) but for a number $n > 1$ of independent point processes on $\mathbb{R}$. The approach in Zhao (2018) further does not yield closed form expressions for the estimates of the coefficients.

Orthogonal series estimators have already been considered by Jalilian et al. (2019) who expand $g_0 - 1$ instead of $\log g_0$. They propose very simple unbiased estimators of the coefficients but the resulting estimator of g_0, referred to as the OSE in the sequel, is not guaranteed to be non-negative.

4.1. Implementation of the VSE

Examples of orthogonal bases include the cosine basis with $w(r) = 1$, $\phi_1(r) = 1/\sqrt{R}$ and $\phi_k(r) = (2/R)^{1/2}\cos\{(k-1)\pi r/R\}$, $k \geq 2$. Another example is the Fourier-Bessel basis with $w(r) = r^{d-1}$ and

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

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 $J_\nu(r)$. In the context of the variational equation (18) we need that the basis functions ϕ_k have non-zero derivatives in order to estimate β_k. This is not the case for ϕ_1 of the cosine basis. We therefore consider in the following the Fourier-Bessel basis.

Let $b_k = 1[k \leq K]$, $k \geq 1$. The mean integrated squared error (MISE) for $\log g_0$ of the VSE over the interval $[r_{\min}, R + r_{\min}]$ is

$$\text{MISE}\big(\widehat{\log g_{0,K}}\big) = \varsigma_d \int_{r_{\min}}^{r_{\min}+R} \mathbb{E}\big\{\widehat{\log g_{0,K}}(r) - \log g_{0,K}(r)\big\}^2 w(r - r_{\min})\mathrm{d}r \quad (20)$$
$$= \varsigma_d \sum_{k=1}^{\infty} \mathbb{E}(b_k\hat{\beta}_k - \beta_k)^2 = \varsigma_d \sum_{k=1}^{\infty} \big[b_k^2 \mathbb{E}\{\hat{\beta}_k^2\} - 2b_k\beta_k\mathbb{E}\hat{\beta}_k + \beta_k^2\big].$$

Jalilian et al. (2019) chose K by minimizing an estimate of the MISE for g_0. We have, however, not been able to construct a useful estimate of (20). Instead we choose K by maximizing a composite likelihood cross-validation criterion

$$\mathrm{CV}(K) = \sum_{\substack{u,v \in \mathbf{X}\cap W: \\ r_{\min} \leq \|u-v\| \leq r_{\min}+\mathbb{R}}}^{\neq} \log[\rho(u)\rho(v)\exp[\widehat{\log g_{0,K}}^{-\{u,v\}}(\|v-u\|)]$$
$$- \sum_{\substack{u,v \in \mathbf{X}\cap W: \\ 0 \leq \|u-v\|-r_{\min} \leq \mathbb{R}}}^{\neq} \log \int_{W^2} 1[0 \leq \|u-v\| - r_{\min} \leq R]\rho(u)\rho(v)\exp[\widehat{\log g_{0,K}}(\|v-u\|)]\mathrm{d}u\mathrm{d}v$$

where $\widehat{\log g_{0,K}}^{-\{u,v\}}$ is the estimate of $\log g_0$ obtained using all pairs of points in $\mathbf{X}$ except (u,v) and (v,u). This is a simplified version of the cross-validation criterion introduced by Guan (2007a) in the context of non-parametric kernel estimation of the pair correlation function.

For computational simplicity and to guard against overfitting we choose inspired by Jalilian et al. (2019) the first local maximum of $\mathrm{CV}(K)$ larger than or equal to two rather than looking for a global maximum. Note that when $\mathbf{A}$ and $\mathbf{b}$ in (18) have been obtained for one value of K, then we obtain the $\mathbf{A}$ and $\mathbf{b}$ for $K+1$ by just adding one new row/column to the previous $\mathbf{A}$ and one new entry to the previous $\mathbf{b}$.

4.2. Simulation study

We study the performance of our variational estimator using simulations of point processes with constant intensity 200 on $W = [0,1]^2$ or $W = [0,2]^2$. We consider the case of a Poisson process for which the pair correlation function is constant equal to one, a Thomas process (parent intensity $\kappa = 25$, dispersal standard deviation $\omega = 0.0198$ and offspring intensity $\mu = 8$), a variance Gamma cluster process (parent intensity $\kappa = 25$, shape parameter $\nu = -1/4$, dispersion parameter $\omega = 0.01845$ and offspring intensity $\mu = 8$), and a determinantal point process (DPP) with exponential kernel $K(r) = \exp(-r/\alpha)$ and $\alpha = 0.039$. The pair correlation functions for the four point process models are shown in Figures 2 and 3 in the usual scale as well as in the log scale. The Thomas and variance Gamma processes are clustered with pair correlation functions bigger than one while the DPP is repulsive with pair correlation function less than one. In all cases we consider $R = 0.125$ and we let $r_{\min} = 0$ for Poisson, Thomas,

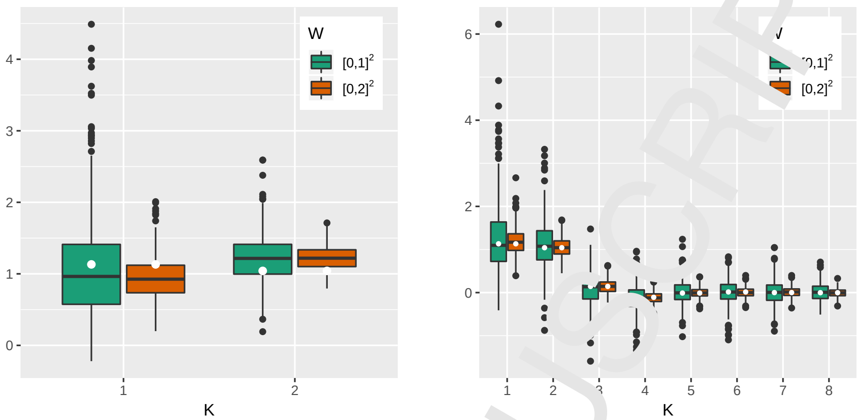


Figure 1: Estimates of the first K coefficients when (19) is truncated to $K = 2$ (left) or $K = 8$ (right) in case of the Thomas process. White points correspond to the true coefficient values. Observation window is either $W = [0, 1]^2$ or $W = [0, 2]^2$.

and variance Gamma. For the DPP the log pair correlation function is not well-defined for $r = 0$ and we therefore use $r_{\min} = 0.01$ in case of the DPP. We use (14) and (15) for computing $\mathbf{A}$ and $\mathbf{b}$ and referring to Remark 2 we let $b = r_{\min} + R$. For each point process we generate 500 simulations.

4.2.1. Estimates of coefficients

Equations (14) and (15) are derived from (7) in which g_0 is the true pair correlation function. In practice, when considering a truncated version of (19), the estimating equation (18) is not unbiased which results in bias of the coefficient estimates. This is exemplified in case of the Thomas process in the left plot of Figure 1 which shows boxplots of the first two coefficient estimates when (19) is truncated to $K = 2$. In the right plot, (19) is truncated to $K = 8$ which means that the truncated version of (19) is very close to the Thomas pair correlation function. Accordingly, the bias of the estimates is much reduced. However, the estimation variance increases when K is increased. This emphasizes the importance of selecting an appropriate trade-off between bias an variance. The plots in Figure 1 also show how the variance of the coefficient estimates decreases when the observation window W is increased from $[0, 1]^2$ to $[0, 2]^2$.

4.2.2. Comparison of estimators

In addition to our new VSE, we also for each simulation consider the OSE proposed by Jalilian et al. (2019) (using the Fourier-Bessel basis and their so-called simple smoothing scheme) and a standard non-parametric kernel density estimate (KDE) with bandwidth chosen by cross-validation (Guan, 2007b; Jalilian and Waagepetersen, 2018).

Figures 2 and 3 depict means of the simulated OSE and VSE estimates of g_0 and $\log g_0$ as well as 95% pointwise envelopes. Table 1 summarizes the root

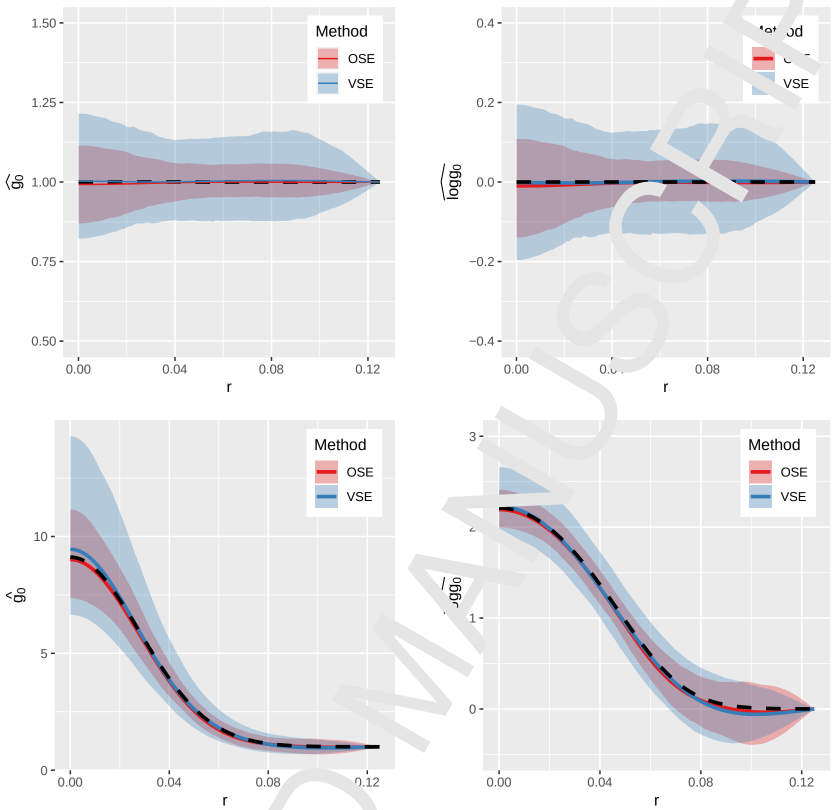


Figure 2: Mean VSE (red curves) and OSE (blue curves) of g_0 (first column) and $\log g_0$ (right column) for Poisson (first row) and Thomas (second row) point processes with $W = [0, 2]^2$. In each plot, the dashed black curve is the true pair correlation or log pair correlation function. The envelopes represent pointwise 95% probability intervals for the estimates.

MISE (square root of (20)) for the three estimators across the four models. Both the figures and the table show that the VSE has larger variance than the OSE. The root MISEs are also larger for VSE than for KDE except in the Poisson case.

We have also compared the computing time to evaluate the OSE and VSE. The OSE is generally cheaper except when the number of points and R are large, see also the case of *Capparis* in Section 4.3.

The numbers in parantheses in Table 1 report the averages of the selected K's for the variational estimator and the OSE. The averages of the selected K's are pretty similar for the Poisson and DPP models while the OSE tends to select higher K than the variational method for the Thomas and variance Gamma point processes.

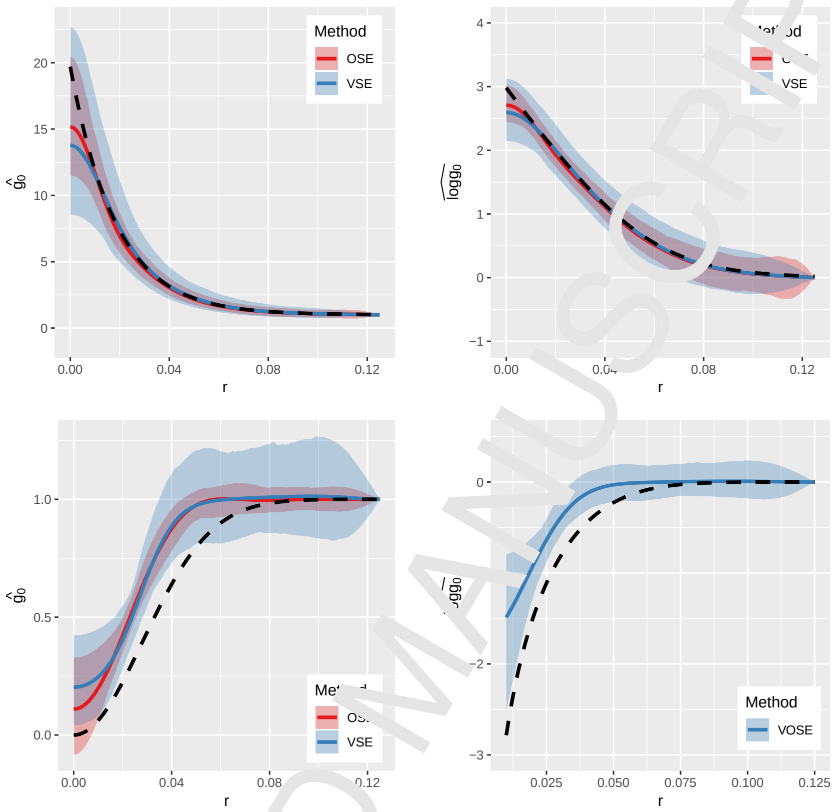


Figure 3: Mean VSE (red curves) and OSE (blue curves) of g_0 (first column) and $\log g_0$ (right column) for variance gamma (first row) and determinantal (second row, $r_{\min} = 0.01$) point processes with $W = [0,2]^2$. In each plot, the dashed black curve is the true pair correlation or log pair correlation function. The envelopes represent pointwise 95% probability intervals for the estimates.

4.3. Data example

To illustrate the use of the VSE in practice, we apply it (as well as the OSE and the KDE) to the data example considered in Jalilian et al. (2019). That is, 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 × 500m Barro Colorado Island plot (Hubbell and Foster, 1983; Condit et al., 1996; Condit, 1998). The intensity functions for the point patterns are estimated as in Jalilian et al. (2019) using log-linear regression models depending on various soil and topographical variables. The estimated pair correlation functions are shown in Figure 4. The selected number K for the VSE are 3, 9 and 5 for *Acalypha*, *Capparis*, and *Lonchocarpus*, while OSE selects $K = 7$ for all species.

In the case of *Capparis*, the computation time (4200 seconds) is higher for the

	Window	OSE	VSE	KDE
Poisson	$[0,1]^2$	0.027 (2.1)	0.051 (2.2)	0.09[illegible]
	$[0,2]^2$	0.012 (2.0)	0.024 (2.2)	0.[illegible]37
Thomas	$[0,1]^2$	0.0995 (3.7)	0.1418* (2.7)	0.111
	$[0,2]^2$	0.044 (4.2)	0.063 (2.9)	0.053
Variance Gamma	$[0,1]^2$	0.099 (6.5)	0.148 (3.8)	0.110
	$[0,2]^2$	0.050 (9.6)	0.072 ([illegible])	0.057
DPP	$[0,1]^2$	NA (3)	0.162[illegible] (3.[illegible])	NA
	$[0,2]^2$	NA (4.1)	0.1582 ([illegible].2)	NA

Table 1: Square-root of the MISE for different estimates of $\log g_0$, observation windows and models. The figures between brackets correspond to the average of the selected K's. The NA's are due to occurrence of non-positive estimates. ($\star$: in this setting one replication produced an outlier and is omitted in the root MISE estimation)

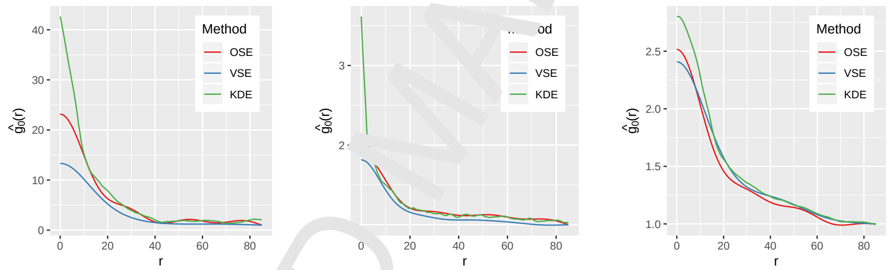


Figure 4: Estimates of g_0 for the three species *Acalypha* (left), *Capparis* (middle) and *Lonchocarpus* (right).

OSE than for the VSE (1244 seconds) due to the high number of points for this species. Comparing the values of the three estimators, the general observation is that they are very similar for large spatial lags but can differ substantially for small lags. This emphasizes the general difficulty of estimating the pair correlation function at small lags.

5. Discussion

In this paper we derive variational equations based on second order properties of a spatial point process. It is remarkable that in case of log-linear parametric models for the pair correlation function, it is possible to derive variational estimating equations which have closed form solutions for the unknown parameters. We exploit this to construct new variational orthogonal series type estimators for the pair correlation function. In contrast to previous kernel and orthogonal series estimators, our new estimate is guaranteed to be non-negative.

For large data sets, the new estimator is further computationally faster than the previous orthogonal series estimate. However, in terms of accuracy as measured by MISE, the new estimator does not outperform the previous estimators. In the data example, the new estimator and the OSE gave similar results.

We believe there is further scope for exploring variational equations. In Sections 3 and 4, we restricted attention to the case of an isotropic pair correlation function. However, by invoking Theorem 1 instead of Theorem 2 it is possible to extend the results to anisotropic translation invariant pair correlation functions. For the VSE we would then need basis function on a subset of $\mathbb{R}^d$ instead of an interval in $\mathbb{R}$. Similar, using basis functions on subsets of $\mathbb{R}^d \times \mathbb{R}$, the VSE could be extended to the space-time case. This is obviously at the expense of extra computations and an increased number of parameters.

Another option for future investigation is to consider non-orthogonal bases for expanding the log pair correlation function instead of the orthogonal Fourier-Bessel basis used in this work. One might for example consider so-called frames (Christensen, 2008) or spline bases.

Acknowledgments

We thank the referees for their helpful comments. Rasmus Waagepetersen's and Francisco Cuevas-Pachecho's research was supported by The Danish Council for Independent Research — Natural Sciences, grant DFF 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. The research of J.-F. Coeurjolly is funded by the Natural Sciences and Engineering Research Council of Canada.

The BCI forest dynamics research project was made possible by National Science Foundation grants to Stephen P. Hubbell: DEB-0640386, DEB-0425651, DEB-0346488, DEB-0129874, DEB-00753102, DEB-9909347, DEB-9615226, DEB-9615226, DEB-9405933, DEB-9221033, DEB-9100058, DEB-8906869, DEB-8605042, DEB-8206992, DEB-7922197, support from the Center for Tropical Forest Science, the Smithsonian Tropical Research Institute, the John D. and Catherine T. MacArthur Foundation, the Mellon Foundation, the Celera Foundation, and numerous private individuals, and through the hard work of over 100 people from 10 countries over the past two decades. The plot project is part of the Center for Tropical Forest Science, a global network of large-scale demographic tree plots.

The BCI soils data set were collected and analyzed by J. Dalling, R. John, K. Harms, R. Stallard and J. Yavitt with support from NSF DEB021104, 021115, 0212284, 0212818 and OISE 0314581, STRI and CTFS. Paolo Segre and Juan Di Trani provided assistance in the field. The covariates `dem`, `grad`, `mrvbf`, `solar` and `twi` were computed in SAGA GIS by Tomislav Hengl (`http://spatial-analyst.net/`).

M.P. Almeida, , and B. Gidas. A variational method for estimating the parameters of MRF from complete or incomplete data. *Annals of Applied Probability*, 3(1):103–136, 1993.

A. Baddeley and D. Dereudre. Variational estimators for the parameters of Gibbs point process models. *Bernoulli*, 19(3):905–930, 2013.

A. Baddeley, E. Rubak, and R. Turner. *Spatial point patterns: methodology and applications with R*. Chapman and Hall/CRC, 2015.

A. J. Baddeley, J. Møller, and R. Waagepetersen. Non- and semi-parametric estimation of interaction in inhomogeneous point patterns. *Statistica Neerlandica*, 54:329–350, 2000.

O. Christensen. *Frames and Bases - an introductory course*. Applied and numerical analysis. Birkhäuser, Basel, 2008.

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

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

R. Condit, S. P. Hubbell, and R. B. Foster. Changes in tree species abundance in a neotropical forest: impact of climate change. *Journal of Tropical Ecology*, 12:231–256, 1996.

L.C. Evans and R.F.. Gariepy. *Measure theory and fine properties of functions.* Studies in Advanced Mathematics. CRC Press, Boca Raton, FL, 1992.

Y. Guan. 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, 2007a.

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

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

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

J. Illian, A. Penttinen, H. Stoyan, and D. Stoyan. *Statistical analysis and modelling of spatial point patterns*, volume 70. John Wiley & Sons, 2008.

A. Jalilian and R. Waagepetersen. Fast bandwidth selection for estimation of the pair correlation function. *Journal of Statistical Computation and Simulation*, 88(10):2001–2011, 2018.

A. Jalilian, Y. Guan, and R. Waagepetersen. Orthogonal series estimation of the pair correlation function of a spatial point process. *Statistica Sinica*, 2019. To appear. Available at arXiv:1702.01736.

J. Møller and R. Waagepetersen. *Statistical inference and simulation for spatial point processes*. CRC Press, 2004.

C. Zhao. *Estimating equation estimators for the pair correlation function*. Open access dissertations 2166, University of Miami, 2018.

Appendix A. Proof of Theorem 1

Proof. Using the Campbell theorem (1) and since $\nabla \log g = (\nabla g)/g$, we start with

$$A := \mathbb{E}\left\{ \sum_{u,v \in \mathbf{X}\cap W}^{\neq} e(u,v)\nabla \log g(v-u)\cdot h(v-u) \right\}$$
$$= \int_W \int_W \frac{1}{|W \cap W_{v-u}|} \frac{\nabla g(v-u)\cdot h(v-u)}{g(v-u)\rho(u)\rho(v)} \rho^{(2)}(u,v)\mathrm{d}u\mathrm{d}v$$
$$= \int_W \int_W \frac{\nabla g(v-u)\cdot h(v-u)}{|W \cap W_{v-u}|}\mathrm{d}u\mathrm{d}v.$$

Using first the invariance by translation of h and ∇g, second Fubini's theorem, and third a change of variables, this reduces to

$$A = \int_{\mathbb{R}^d} \nabla g(w)\cdot h(w)\mathrm{d}w.$$

By assumption, we have using the dominated convergence theorem,

$$A = \lim_{n\to\infty} A_n \quad \text{where } A_n := \int_{B_n} \nabla g(w)\cdot h(w)\mathrm{d}w.$$

We can now use the standard trace theorem (see for instance Evans and Gariepy (1992)) and obtain

$$A_n = -\int_{B_n} g(w)(\operatorname{div} h)(w)\mathrm{d}w + \int_{\partial B_n} g(w)h(w)\cdot \nu(\mathrm{d}w).$$

From (3), we deduce from the dominated convergence theorem that

$$A = \lim_{n\to\infty} A_n = -\int_{\mathbb{R}^d} g(w)(\operatorname{div} h)(w)\mathrm{d}w.$$

Finally, using successively a change of variable and the Campbell theorem we get

$$A = -\int_W \int_W \frac{(\operatorname{div} h)(v-u)}{|W \cap W_{v-u}|} \frac{\rho^{(2)}(u,v)}{\rho(u)\rho(v)} \mathrm{d}u\mathrm{d}v$$
$$= -\mathbb{E}\left\{ \sum_{u,v \in \mathbf{X} \cap W}^{\neq} e(u,v) \, (\operatorname{div} h)(v-u) \right\}$$

which proves (4). □

Appendix B. Proof of Theorem 2

Proof. Both (7) and (8) are proved similarly. We focus only on (8) and follow the proof of Theorem 1. Using the Campbell theorem (1), the fact $(\log g_0)' = g_0'/g_0$ and finally a change to polar coordinates, we have

$$A := \mathbb{E}\left\{ \sum_{u,v \in \mathbf{X} \cap W}^{\neq} e(u,v)(\log g_0)'(\|v-u\|)h(\|v-u\|) \right\}$$
$$= \int_W \int_W \frac{1}{|W \cap W_{v-u}|} \frac{g_0'(\|v-u\|)h(\|v-u\|)}{g_0(\|v-u\|)\rho(u)\rho(v)} \rho^{(2)}(u,v)\mathrm{d}u\mathrm{d}v$$
$$= \int_W \int_W \frac{g_0'(\|v-u\|)h(\|v-u\|)}{|W \cap W_{v-u}|} \mathrm{d}u\mathrm{d}v$$
$$= \int_{\mathbb{R}^d} g_0'(\|w\|)h(\|w\|)\mathrm{d}w$$
$$= \varsigma_d \int_0^\infty t^{d-1} g_0'(t)h(t)\mathrm{d}t.$$

Using the dominated convergence theorem, partial integration and (6) we have

$$\int_0^\infty t^{d-1} g_0'(t)h(t)\mathrm{d}t = \lim_{n\to\infty} \int_0^n t^{d-1} g_0'(t)h(t)\mathrm{d}t$$
$$= -\lim_{n\to\infty} \int_0^n t^{d-1} g_0(t) \left\{ \frac{(d-1)h(t)}{t} + h'(t) \right\} \mathrm{d}t$$
$$= -\int_0^\infty t^{d-1} g_0(t) \left\{ \frac{(d-1)h(t)}{t} + h'(t) \right\} \mathrm{d}t.$$

A change to polar coordinates and the Campbell theorem again lead to

$$A = -\int_{\mathbb{R}^d} g_0(\|w\|) \left\{ \frac{(d-1)h(\|w\|)}{\|w\|} + h'(\|w\|) \right\} \mathrm{d}w$$
$$= -\int_W \int_W \left\{ \frac{(d-1)h(\|w\|)}{\|w\|} + h'(\|w\|) \right\} \frac{\rho^{(2)}(u,v)}{\rho(u)\rho(v)|W \cap W_{v-u}|} \mathrm{d}u\mathrm{d}v$$
$$= -\mathbb{E}\left[\sum_{u,v \in \mathbf{X} \cap W}^{\neq} e(u,v) \left\{ (d-1)\frac{h(\|v-u\|)}{\|v-u\|} + h'(\|v-u\|) \right\} \right].$$