Abdollah Jalilian, Yongtao Guan and Rasmus Waagepetersen

# Decomposition of Variance for Spatial Cox Processes

# Decomposition of Variance for Spatial Cox Processes

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

[1]Department of Statistics, Shahid Beheshti University, Evin, Tehran, 19839-63113, Iran
[2]Division of Biostatistics, Yale University, New Haven, Connecticut 06520-8034, U.S.A
[3]Department of Mathematical Sciences, Aalborg University, Fredrik Bajersvej 7G, DK-9220 Aalborg, Denmark, rw@math.aau.dk

### Abstract

Spatial Cox point processes is a natural framework for quantifying the various sources of variation governing the spatial distribution of rain forest trees. We introduce a general criterion for variance decomposition for spatial Cox processes and apply it to specific Cox process models with additive or log linear random intensity functions. We moreover consider a new and flexible class of pair correlation function models given in terms of Matérn covariance functions. The proposed methodology is applied to point pattern data sets of locations of tropical rain forest trees.

*Keywords:* additive random intensity, composite likelihood, Cox process, Matérn covariance function, pair correlation function, variance component.

## 1   Introduction

The spatial distributions of tropical rain forest trees are influenced by many factors including e.g. spatially varying environmental conditions, seed dispersal, infectious diseases, and gap creation by hurricanes. One natural question is, loosely speaking, 'how much of the variation' in the spatial distribution of a tropical rain forest tree species can be attributed to each of these factors? In this paper we study methods for addressing this question with a particular focus on the contribution of the environment.

The most fundamental summaries of variation are the variances of counts of trees in bounded regions. A generalized linear mixed model (GLMM) for such counts, with a variance component for each of the sources of variation, might be a starting point for decomposing the variance. However, with this approach one loses the fine-scale information contained in extensive tropical rain forest data sets which include locations of individual trees and not just numbers of trees in certain regions (e.g. Condit, 1998). Moreover, conclusions obtained from a fitted GLMM may in general be strongly dependent on the sizes and shapes of the regions used to create the count data.

A more natural and flexible approach is to model the individual tree locations as a spatial point process (e.g. Møller and Waagepetersen, 2004) and then derive summaries of variation from a fitted spatial point process model. With this approach one can compute variances and covariances of counts in arbitrary regions and therefore recover all summaries that would be obtained from a GLMM approach. In particular, Cox processes (e.g. Møller and Waagepetersen, 2004) provide a useful framework for integrating different sources of variation.

This paper introduces a general criterion for decomposition of variance for Cox processes. The criterion is applied to specific Cox process models with either additive or log linear random intensity functions. The additive model is appealing in the context of variance decomposition but has not been well studied in the point process literature. For either model, the resulting variance decomposition depends on the assumed pair correlation function which characterizes the spatial correlation in the Cox process. It is therefore important to have a wide class of pair correlation functions to choose from. To this end we consider a new class of shot-noise Cox processes with flexible Matérn pair correlation functions. These functions are closely related to their widely used counterparts in geostatistics and hence serve as a bridge to further connect the theories of geostatistics and point processes.

## 1.1 A tropical rain forest data example

Waagepetersen and Guan (2009) fitted so-called inhomogeneous Thomas processes to locations of three tree species, namely *Acalypha diversifolia* (528 trees), *Lonchocarpus heptaphyllus* (836 trees), and *Capparis frondosa* (3299 trees) that were alive in 1995 in the 1000 m by 500 m Barro Colorado Island plot (Condit et al., 1996; Condit, 1998; Hubbell and Foster, 1983). The significant covariates in the fitted Thomas models were elevation and potassium for Acalypha and Capparis and nitrogen and phosphorous for Lonchocarpus. The point patterns of tree locations and the covariate potassium are shown in Figure 1. In Section 6 we extend the analysis in Waagepetersen and Guan (2009) by studying decomposition of variance for these data sets. We moreover employ a much broader class of Cox processes than the inhomogeneous Thomas processes used in Waagepetersen and Guan (2009).

# 2 Background on spatial point processes

Let $X$ be a point process observed on $S \subseteq \mathbb{R}^2$ and let $N(B)$, for any bounded $B \subseteq S$, denote the number of points in $X \cap B$. The first- and second-order moments of the counts $N(B)$ are determined by the intensity function $\rho$ and the second-order product density $\rho^{(2)}$ of $X$ which are functions defined on $S$ and $S \times S$, respectively (see Møller and Waagepetersen, 2004). More precisely,

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

and

$$\mathbb{E}N(A)N(B) = \int_{A \cap B} \rho(u)\mathrm{d}u + \int_A \int_B \rho^{(2)}(u, v)\mathrm{d}u\mathrm{d}v$$
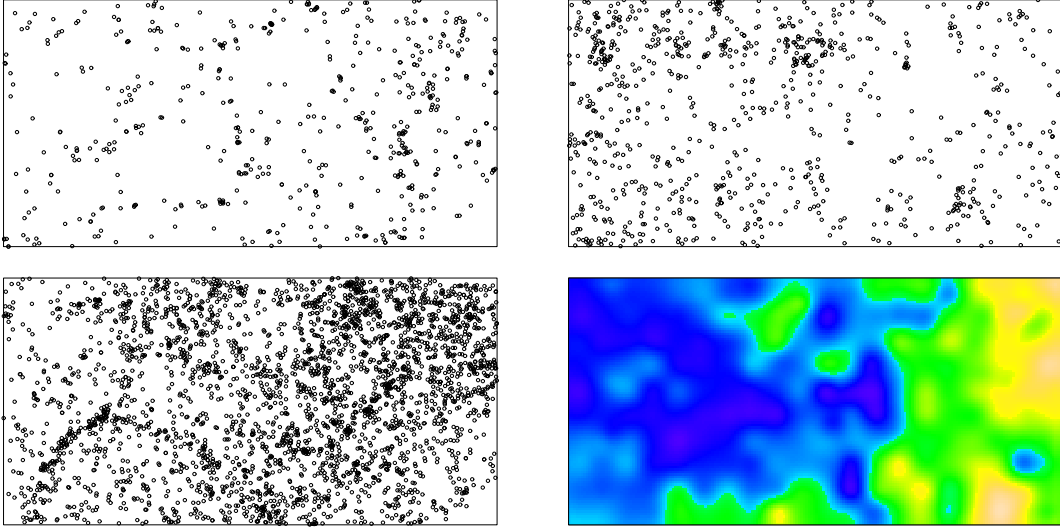
**Figure 1:** Locations of Acalypha (top left), Lonchocarpus (top right), and Capparis trees (bottom left) and image of interpolated potassium content in the surface soil (bottom right).

for bounded $A, B \subseteq S$. The pair correlation function $g$ is given by $g(u, v) = \rho^{(2)}(u, v)/[\rho(u)\rho(v)]$.

For a Poisson process, the counts in disjoint regions are independent and Poisson distributed. A Cox process is driven by a random intensity function $\Lambda = \{\Lambda(u) : u \in S\}$. Conditional on $\Lambda = \lambda$, the Cox process becomes a Poisson process with intensity function $\lambda$. For a Cox process, $\rho(u) = \mathbb{E}\Lambda(u)$ and $\rho^{(2)}(u, v) = \mathbb{E}[\Lambda(u)\Lambda(v)]$.

# 3 Decomposition of variance

To quantify the various sources of variation in a spatial point pattern we consider the contribution of each source to the total variation of a count $N(B)$ for a region $B$. According to the previous section,

$$\mathbb{V}\mathrm{ar}\, N(B) = \int_B \rho(u)\mathrm{d}u + \int_B \int_B \rho(u)\rho(v)[g(u, v) - 1]\mathrm{d}u\mathrm{d}v. \qquad (3.1)$$

The first term on the right-hand side of (3.1) is the variance of $N(B)$ for a Poisson process with intensity function $\rho(\cdot)$. The second term is the increase (or decrease) in variance due to possible attraction (or repulsion) between points corresponding to $g > 1$ (or $g < 1$).

For a Cox process $X$ driven by a random intensity function $\Lambda$, we can decompose a count $N(B)$ as $N(B) = I_{|\Lambda}(B) + \hat{N}_{|\Lambda}(B)$ where

$$\hat{N}_{|\Lambda}(B) = \mathbb{E}[N(B)|\Lambda] = \int_B \Lambda(u)\mathrm{d}u$$

is the minimum mean square error predictor of $N(B)$ given $\Lambda$ and $I_{|\Lambda}(B) = N(B) - \hat{N}_{|\Lambda}(B)$ is an innovation process in the terminology of Baddeley et al. (2005). Then

$I_{|\Lambda}(B)$ and $\hat{N}_{|\Lambda}(B)$ are uncorrelated processes with

$$\mathbb{E}I_{|\Lambda}(B) = 0, \quad \mathbb{C}\text{ov}[I_{|\Lambda}(A), I_{|\Lambda}(B)] = \int_{A \cap B} \rho(u)\mathrm{d}u$$

and

$$\mathbb{C}\text{ov}[\hat{N}_{|\Lambda}(A), \hat{N}_{|\Lambda}(B)] = \int_A \int_B \rho(u)\rho(v)[g(u,v) - 1]\mathrm{d}u\mathrm{d}v$$

since $\mathbb{C}\text{ov}[\Lambda(u), \Lambda(v)] = \rho(u)\rho(v)[g(u,v) - 1]$. The process $I_{|\Lambda}$ may be viewed as a 'nugget' process in geostatistical terminology since $I_{|\Lambda}(A)$ and $I_{|\Lambda}(B)$ are uncorrelated when $A$ and $B$ are disjoint. The spatially correlated sources of variation causing the extra-Poisson variance in (3.1) enters via $\hat{N}_{|\Lambda}$.

In Section 4 we model $\Lambda$ in terms of a random covariate process $Z$. To quantify how much of the spatially structured variation is due to $Z$ we further decompose $\hat{N}_{|\Lambda}(B) = I_{|\Lambda,Z}(B) + \hat{N}_{|\Lambda,Z}$ into uncorrelated components

$$\hat{N}_{|\Lambda,Z} = \mathbb{E}[\hat{N}_{|\Lambda}(B)|Z] = \int_B \mathbb{E}[\Lambda(u)|Z]\mathrm{d}u$$

and

$$I_{|\Lambda,Z}(B) = \hat{N}_{|\Lambda}(B) - \hat{N}_{|\Lambda,Z}(B) = \int_B \left\{\Lambda(u) - \mathbb{E}[\Lambda(u)|Z]\right\}\mathrm{d}u.$$

If all random variation in $\Lambda$ is due to $Z$ then $\mathbb{V}\text{ar}\, I_{|\Lambda,Z}(B) = 0$. Furthermore, $\mathbb{V}\text{ar}\, \hat{N}_{|\Lambda,Z}(B) = 0$ if $\Lambda$ is independent of $Z$. The quantity

$$R^2(B) = \frac{\mathbb{V}\text{ar}\, \hat{N}_{|\Lambda,Z}(B)}{\mathbb{V}\text{ar}\, \hat{N}_{|\Lambda}(B)} = \frac{\mathbb{V}\text{ar} \int_B \mathbb{E}[\Lambda(u)|Z]\mathrm{d}u}{\mathbb{V}\text{ar}[\int_B \Lambda(u)\mathrm{d}u]} \tag{3.2}$$

thus describes the proportion of the variance of $\int_B \Lambda(u)\mathrm{d}u$ explained by $Z$. For sufficiently small $B$, $R^2(B)$ may be approximated by

$$R^2 = \frac{\mathbb{V}\text{ar}\, \mathbb{E}[\Lambda(u_B)|Z]}{\mathbb{V}\text{ar}\, \Lambda(u_B)}, \quad u_B \in B \tag{3.3}$$

which in the stationary case does not depend on $B$. The quantity $R^2$ can also be viewed as an analogue of the $R^2$ statistic for linear regression. To see this connection, we define the expected 'sums of squares'

$$\text{SSR} = \mathbb{E}\int_S I_R^2(u)\mathrm{d}u, \quad \text{SSE} = \mathbb{E}\int_S I_E^2(u)\mathrm{d}u$$

and

$$\text{SST} = \text{SSR} + \text{SSE}$$

where $I_R(u) = \hat{\Lambda}_{|Z}(u) - \rho(u)$, $I_E(u) = \Lambda(u) - \hat{\Lambda}_{|Z}(u)$, and $\hat{\Lambda}_{|Z}(u) = \mathbb{E}[\Lambda(u)|Z]$. Then, in the case of a stationary random intensity function $\Lambda$, the ratio SSR/SST (known as $R^2$ for linear regression) coincides with (3.3). We discuss our $R^2$ criterion in relation to specific models in Section 4.

# 4 Models for rain forest data

Let $X$ be the process that generates the point pattern of locations for a particular tree species. We assume that $X$ is a Cox process driven by a random intensity function $\Lambda$ where $\Lambda$ depends on a stationary process $Z = \{Z(u) : u \in S\}$, $Z(u) = [Z_1(u), \ldots, Z_p(u)]$, of $p$ observed covariates and a non-negative process $\Lambda_0 = \{\Lambda_0(u) : u \in S\}$ representing unobserved sources of variation. We assume that $\Lambda$ depends on $Z$ through $\widetilde{Z} = \{\beta_0 + \beta_{1:p}Z(u)^\mathsf{T}\}_{u \in S}$ for some parameter $\beta = (\beta_0, \beta_{1:p}) = (\beta_0, \beta_1, \ldots, \beta_p) \in \mathbb{R}^{p+1}$ and that $\widetilde{Z}$ and $\Lambda_0$ are independent second-order stationary processes.

We further let $\rho_0 = \mathbb{E}\Lambda_0(u)$ and $g_0(u - v) = \mathbb{E}[\Lambda_0(u)\Lambda_0(v)]/\rho_0^2$. Then $\rho_0$ and $g_0$ become respectively the intensity and the pair correlation function of a Cox process driven by the random intensity function $\Lambda_0$. Note that

$$g_0(u - v) = c_0(u - v)/\rho_0^2 + 1 \tag{4.1}$$

where $c_0$ is the covariance function for $\Lambda_0$. Moreover $\sigma_0^2 = \mathbb{Var}[\Lambda_0(u)]$ is equal to $\rho_0^2[g_0(0) - 1]$. In Section 4.3 we discuss parametric models for $g_0$ or equivalently $c_0$.

To quantify how much variation is due to $Z$, it is appropriate to model $Z$ as a random field. However, since $Z$ is observed, it is often convenient to base parameter estimation on $X|Z$ where $Z$ is then treated as a fixed quantity, see Section 5.

## 4.1 An additive model

Variance decomposition is straightforward for the following additive model:

$$\Lambda(u) = \widetilde{Z}(u) + \Lambda_0(u) = \beta_0 + \beta_{1:p}Z(u)^\mathsf{T} + \Lambda_0(u). \tag{4.2}$$

Thus, $X$ can be considered as a superposition of two independent Cox processes with random intensity functions $\widetilde{Z}(u)$ and $\Lambda_0$. A drawback of this model is that $\Lambda$ is not positive for all values of $\beta$ and $Z$. This is probably why (4.2) has not attracted much interest in the point process literature; Best et al. (2000) is one notable exception. In Section 4.2 we discuss an alternative log linear model for which positivity of $\Lambda$ is guaranteed.

Conditional on $Z$, the intensity function of $X$ becomes

$$\rho(u|Z;\beta) = \rho_0 + \beta_0 + \beta_{1:p}Z(u)^\mathsf{T}.$$

In practice, a negative estimate of the intensity function may be obtained. This did, however, not happen in our data examples (Section 6).

Further, given $Z$, the second-order product density and the pair correlation function become

$$\rho^{(2)}(u, v|Z;\beta) = \rho(u|Z;\beta)\rho(v|Z;\beta) + c_0(u - v)$$

and $g(u, v|Z;\beta) = 1 + c_0(u - v)/[\rho(u|Z;\beta)\rho(v|Z;\beta)]$.

For (4.2) we obtain a straightforward decomposition of the variance of $\Lambda(u)$ into the sum of $\sigma_{\widetilde{Z}}^2 = \mathbb{Var}\,\widetilde{Z}(u)$ and $\sigma_0^2 = \mathbb{Var}\,\Lambda_0(u)$. Hence (3.3) becomes

$$R^2 = \frac{\sigma_{\widetilde{Z}}^2}{\sigma_{\widetilde{Z}}^2 + \sigma_0^2}.$$

## 4.2 A log linear model

The multiplicative log linear random intensity function

$$\Lambda(u) = \exp[\widetilde{Z}(u) + \log \Lambda_0(u)] = \Lambda_0(u) \exp[\beta_0 + \beta_{1:p} Z(u)^{\mathsf{T}}] \qquad (4.3)$$

is always non-negative. This model has an appealing interpretation in terms of location dependent thinning (Waagepetersen, 2007) where $X$ is obtained by independent thinning of a Cox process driven by $\Lambda_0$ and the probability of retaining a point at $u$ is proportional to $\exp[\widetilde{Z}(u)]$. Hence in the tropical rain forest context, the covariates may be regarded as influencing the survival of plants in a stationary process of seedlings.

### 4.2.1 Intensity and pair correlation

Without loss of generality we assume in case of (4.3) that $\rho_0 = \mathbb{E}\Lambda_0(u) = 1$. Conditional on $Z$, the intensity function of $X$ is

$$\rho(u|Z;\beta) = \mathbb{E}[\Lambda(u)|Z] = \exp[\widetilde{Z}(u)],$$

the pair correlation function $g(u - v|Z) = g_0(u - v)$ coincides with $g_0$, and

$$\rho^{(2)}(u, v|Z) = \rho(u|Z;\beta)\rho(v|Z;\beta) + \rho(u|Z;\beta)\rho(v|Z;\beta)c_0(u - v).$$

Unconditionally, the intensity and the pair correlation function become

$$\rho(u) = \rho_{\exp \widetilde{Z}} = \mathbb{E}\exp[\widetilde{Z}(u)] \text{ and } g(u - v) = g_0(u - v)g_{\exp \widetilde{Z}}(u - v)$$

where $\rho_{\exp \widetilde{Z}}$ and $g_{\exp \widetilde{Z}}$ are the intensity and the pair correlation function of a Cox process with random intensity function $\exp[\widetilde{Z}(\cdot)]$.

### 4.2.2 Decomposition of variance

For a Cox process with random intensity function (4.3),

$$\mathbb{V}\mathrm{ar}[\Lambda(u)] = \sigma^2_{\exp \widetilde{Z}} + \sigma_0^2\big[\sigma^2_{\exp \widetilde{Z}} + \rho^2_{\exp \widetilde{Z}}\big]$$

where

$$\sigma^2_{\exp \widetilde{Z}} = \mathbb{V}\mathrm{ar}\,\mathbb{E}[\Lambda(u)|Z] = \mathbb{V}\mathrm{ar}\exp[\widetilde{Z}(u)].$$

According to (3.3), $R^2$ becomes

$$R^2 = \frac{\sigma^2_{\exp \widetilde{Z}}}{\sigma^2_{\exp \widetilde{Z}} + \sigma_0^2\big[\sigma^2_{\exp \widetilde{Z}} + \rho^2_{\exp \widetilde{Z}}\big]} = \frac{g_{\exp \widetilde{Z}}(0) - 1}{g_{\exp \widetilde{Z}}(0)g_0(0) - 1}.$$

A related approach is to consider the proportion of variance of $\log \Lambda$ explained by $\widetilde{Z}$. If both $\widetilde{Z}$ and $\log \Lambda_0$ are Gaussian then (Møller et al., 1998)

$$\mathbb{V}\mathrm{ar}\log \Lambda(u) = \mathbb{V}\mathrm{ar}\,\widetilde{Z}(u) + \mathbb{V}\mathrm{ar}\log \Lambda_0(u) = \log g_0(0) + \log g_{\exp \widetilde{Z}}(0).$$

The proportion of variance for $\log \Lambda$ explained by $Z$ then becomes

$$\frac{\log g_{\exp \widetilde{Z}}(0)}{\log g_0(0) + \log g_{\exp \widetilde{Z}}(0)}$$

which is related to $R^2$ by the approximation $\log(x) \approx x - 1$.

## 4.3 Models for $c_0$

Our estimation procedure in Section 5 requires the specification of parametric models for the intensity function and the second-order product density of $X|Z$. Hence it remains to specify a parametric model for the covariance function $c_0$ of $\Lambda_0$ or equivalently (cf. (4.1)) the pair correlation function $g_0$. Below we consider covariance functions $c_0$ obtained from explicit constructions of non-negative random fields.

A log Gaussian random field (Møller et al., 1998) obtained by exponentiating a Gaussian random field $Y$ is one type of non-negative random field with covariance function of a known form

$$c_0(u - v) = \rho_0^2 \exp[c(u - v)] - \rho_0^2 \tag{4.4}$$

where $c(u - v) = \mathbb{C}\text{ov}[Y(u), Y(v)]$.

Another popular construction of non-negative random fields are shot-noise fields given by sums of positive kernel functions scaled by a parameter $\alpha > 0$ and centered around points of a homogeneous Poisson point process with intensity $\kappa > 0$. The resulting covariance function becomes

$$c_0(h) = \kappa\alpha^2 \int_{\mathbb{R}^2} k(u)k(u + h)\mathrm{d}u. \tag{4.5}$$

Any absolutely integrable covariance function in fact has a representation of the form (4.5) (see e.g. page 65 and 489 in Chilès and Delfiner, 1999) but it is not always easy to determine the function $k$. The random intensity function of a modified Thomas process is obtained when $k$ is a Gaussian density with standard deviation $\omega$ in which case $c_0$ is a Gaussian covariance function

$$c_0(h) = \sigma_0^2 \exp\left[-(\|h\|/\eta)^2\right] \tag{4.6}$$

where $\sigma_0^2 = \kappa\alpha^2/(\pi\eta^2)$ and $\eta = 2\omega$.

In the next Section 4.3.1 we consider a more flexible class of Bessel shot-noise fields where the kernel $k$ is given in terms of a Bessel function and the corresponding class of covariance functions is the so-called Matérn class. In addition, we in Section 4.3.2 consider covariance functions obtained when $k$ is a Cauchy density.

### 4.3.1 Bessel shot-noise fields and Matérn pair correlation functions

Consider the Bessel probability density function

$$k(u) = \frac{1}{\pi 2^{\nu'+1}\eta^2\Gamma(\nu'+1)}(\|u\|/\eta)^{\nu'}K_{\nu'}(\|u\|/\eta), \quad \eta > 0, \quad \nu' > -1/2 \tag{4.7}$$

where $K_{\nu'}$ is a modified Bessel function of the second kind (McKay, 1932; Nadarajah and Gupta, 2006). According to (4.5) and Table 1 on page 30 in Matérn (1986), the covariance function for the corresponding shot-noise field becomes

$$c_0(h) = \sigma_0^2\frac{(\|h\|/\eta)^\nu K_\nu(\|h\|/\eta)}{2^{\nu-1}\Gamma(\nu)}, \quad \nu = 2\nu' + 1, \quad \sigma_0^2 > 0. \tag{4.8}$$

A covariance function of this form is commonly known as a Matérn covariance function (e.g. Stein, 1999). Since (4.7) is symmetric, the integral term in (4.5) represents the density function of the sum of two independent random vectors with common density function $k$. Thus (4.8) divided by $\kappa\alpha^2$ is a probability density of the form (4.7) and hence $\sigma_0^2 = \kappa\alpha^2/(4\pi\eta^2\nu)$.

The resulting Matérn pair correlation function is

$$g_0(u) = 1 + \sigma_0^2 \frac{(\|h\|/\eta)^\nu K_\nu(\|h\|/\eta)}{2^{\nu-1}\Gamma(\nu)\rho_0^2}$$

where $\rho_0 = \kappa\alpha$. The smoothness parameter $\nu$ controls the shape of the pair correlation function and gives additional flexibility in the modeling. For instance, $\nu = 1/2$ yields the exponential model

$$g_0(u) = 1 + \sigma_0^2 \exp(-\|h\|/\eta)/\rho_0^2 \tag{4.9}$$

which offers more slowly decaying correlations than the Thomas pair correlation function obtained with (4.6). Note that the Gaussian covariance function (4.6) may be viewed as a limiting case of a Matérn covariance function when $\nu \to \infty$ (Stein, 1999, page 50).

### 4.3.2 Cauchy covariance function

Another model for a slowly decaying pair correlation function is obtained with

$$c_0(u) = \sigma_0^2 \big[1 + (\|u\|/\eta)^2\big]^{-3/2} \tag{4.10}$$

where the kernel $k$ is of Cauchy type (see Table 1, page 30 in Matérn, 1986)

$$k(u) = \frac{1}{2\pi\omega^2}\big[1 + (\|u\|/\omega)^2\big]^{-3/2},$$

$\sigma_0^2 = \kappa\alpha^2/(2\pi\eta^2)$, and $\eta = 2\omega$. This covariance function is a special case of the Cauchy covariance function $\sigma^2\big[1 + (\|u\|/\eta)^2\big]^{-\nu}$ which is absolutely integrable for $\nu > 1$. However, we do not know the kernel function $k$ for which the general Cauchy covariance function has a representation of the form (4.5).

# 5 Parameter estimation

In practice we consider a parametric model $c_0(\cdot; \psi)$ for the covariance function of $\Lambda_0$ where $c_0$ could for instance be the Matérn covariance function (4.8) with $\psi = (\sigma_0^2, \eta, \nu)$. Given $(X, Z)$ observed within $W$, we then obtain a plug-in estimate for $R^2$ in (3.3) after estimating the three parameters $\beta$, $\psi$, and $\sigma_{\tilde{Z}}^2$ (or $\sigma_{\exp \tilde{Z}}^2$).

## 5.1 Composite likelihood estimation

The inference about $\beta$ and $\psi$ is based on $X|Z$. The regression parameter $\beta$ can be estimated by the first-order composite log likelihood function ($\mathrm{CL}_1$) (Schoenberg, 2005; Waagepetersen, 2007)

$$\mathrm{CL}_1(\beta) = \sum_{u \in X} \log \rho(u|Z; \beta) - \int_W \rho(u|Z; \beta) \mathrm{d}u \tag{5.1}$$

which is formally equivalent to the log likelihood function of a Poisson process with intensity function $\rho(\cdot|Z; \beta)$. The Berman-Turner quadrature scheme can be used to approximate the integral in (5.1) both for the log linear and the additive model, see e.g. Baddeley and Turner (2000).

For a Cox process specified by the log linear random intensity function (4.3) one may subsequently estimate $\psi$ using minimum contrast estimation based on the $K$-function, see Waagepetersen (2007) and Waagepetersen and Guan (2009). However, for the additive model the $K$-function is not well-defined since the pair correlation function is not translation invariant (Baddeley et al., 2000). As suggested by Guan (2006) and Waagepetersen (2007), $\psi$ can instead be estimated using a second-order composite likelihood function

$$\mathrm{CL}_2(\psi|\beta) = \sum_{u,v \in X}^{\neq} h(u, v) \log \rho^{(2)}(u, v|Z; \beta, \psi)$$
$$- \iint_{W^2} h(u, v) \rho^{(2)}(u, v|Z; \beta, \psi) \mathrm{d}u \mathrm{d}v \tag{5.2}$$

based on the second-order product density where $h$ is a weight function, see e.g. (5.4) and (5.5) below. The integral term in (5.2) can also be approximated by a variant of the Berman-Turner scheme. The second-order composite likelihood can be evaluated for both the log linear model and the additive model and the maximized $\mathrm{CL}_2$ can be used as a criterion for model selection, see Section 6.

Instead of maximizing the right-hand side of (5.2) with respect to both $\beta$ and $\psi$, a computationally simpler approach is to obtain $\hat{\beta}$ by maximizing (5.1) and then $\hat{\psi}$ by maximizing $\mathrm{CL}_2(\cdot|\hat{\beta})$. Moreover, an equivalent version of $\mathrm{CL}_2(\psi|\hat{\beta})$ is given by

$$\mathrm{CL}_2^*(\psi|\hat{\beta}) = \sum_{u,v \in X}^{\neq} h(u, v) \log g(u, v|Z; \hat{\beta}, \psi)$$
$$- \iint_{W^2} h(u, v) \, \mathbb{C}\mathrm{ov}[\Lambda(u), \Lambda(v)|Z; \hat{\beta}, \psi] \mathrm{d}u \mathrm{d}v \tag{5.3}$$

which is obtained from $\mathrm{CL}_2(\psi|\hat{\beta})$ by subtracting the second-order composite likelihood of a Poisson process with intensity function $\rho(\cdot|Z;\hat{\beta})$. More stable convergence results were obtained using (5.3) instead of (5.2) when using a standard implementation of the Nelder-Mead algorithm for maximizing the second-order composite likelihood.

In simulation studies and applications to real data we discovered that the second-order composite likelihood estimates can be quite sensitive to the accuracy of the Berman-Turner quadrature scheme used to approximate the double integral in (5.3). This is especially the case when $c_0$ is steep at zero like for (4.8) with a small $\nu$ or small $\eta$. Regarding the weight function $h$,

$$h(u,v) = 1[\|u-v\| \leq t]/[\pi t^2] \tag{5.4}$$

is the standard choice which ensures that only $t$-close pairs of points are used in the composite likelihood. In the data example in Section 6 we also considered

$$h(u,v) = 1[\|u-v\| \leq t]/[\pi t^2 \rho(u|Z,\hat{\beta})\rho(u|Z,\hat{\beta})]. \tag{5.5}$$

For log linear models, this $h$ implies a simplification of the double integral since the intensity function is eliminated from the integrand.

## 5.2 Estimation of environmental variances

In practice $Z$ is observed on a grid $G = \{u_i\}_{i=1,\dots,M}$ covering $W$. Since we assume that $Z$ is a stationary process, simple estimators of $\sigma_{\tilde{Z}}^2$ and $\sigma_{\exp \tilde{Z}}^2$ are given by

$$\hat{\sigma}_{\tilde{Z}}^2 = \frac{1}{M} \sum_{u \in G} \left[\widehat{\tilde{Z}}(u) - \hat{\rho}_{\tilde{Z}}\right]^2 \tag{5.6}$$

and

$$\hat{\sigma}_{\exp \tilde{Z}}^2 = \frac{1}{M} \sum_{u \in G} \left\{ \exp \left[\widehat{\tilde{Z}}(u)\right] - \hat{\rho}_{\exp \tilde{Z}} \right\}^2 \tag{5.7}$$

where

$$\hat{\rho}_{\tilde{Z}} = \frac{1}{M} \sum_{u \in G} \widehat{\tilde{Z}}(u), \quad \hat{\rho}_{\exp \tilde{Z}} = \frac{1}{M} \sum_{u \in G} \exp \left[\widehat{\tilde{Z}}(u)\right], \tag{5.8}$$

and $\widehat{\tilde{Z}}(u) = \hat{\beta}Z(u)^\mathsf{T}$.

As a theoretical sanity check we discuss the joint asymptotic properties of the composite likelihood estimates and (5.6)-(5.8) in the Appendix. The asymptotic results seem less useful from a practical point of view due to the non-linear dependence of $R^2$ on the parameters. Also we expect slow convergence to normality of (5.6)–(5.8).

## 5.3 Further approaches to non-parametric estimation

Suppose we divide the observation window into quadrats $B_i$, $i = 1, \dots, M$, of identical size. Then we may estimate the denominator of (3.2) with $B = B_1$ simply as

$\widehat{\mathbb{V}\mathrm{ar}\, N(B)} - |B|\hat{\rho}$ where $\widehat{\mathbb{V}\mathrm{ar}\, N(B)} = \sum_{i=1}^{M}[N(B_i) - |B|\hat{\rho}]^2/M$ and $\hat{\rho} = N(W)/|W|$. A problem with this approach is that $\widehat{\mathbb{V}\mathrm{ar}\, N(B)} - |B|\hat{\rho}$ may be negative. If we choose the $B_i$ so small that $N_i(B)$ becomes binary, then indeed $\widehat{\mathbb{V}\mathrm{ar}\, N(B)} - |B|\hat{\rho} = -(|B|\hat{\rho})^2 < 0$.

In the case of the log linear model it is possible to estimate the pair correlation function $g_0(\cdot) = g(\cdot|Z)$ and hence $\sigma_0^2 = g_0(0) - 1$ non-parametrically (Baddeley et al., 2000) but the non-parametric estimate is not reliable for small distances due to bias and large variability, see e.g. Chapter 4 in Møller and Waagepetersen (2004). It is, however, still useful for model assessment, see Section 6.

# 6    Decomposition of variance for tropical rain forest data example

In this section we return to the data example in Section 1.1. Waagepetersen and Guan (2009) fitted inhomogeneous Thomas models with random intensity functions of the form (4.3) and a Gaussian covariance function (4.6) for $\Lambda_0$. Using the same covariates as in Waagepetersen and Guan (2009), we fit for each of the species the additive model (4.2) and the log linear model (4.3) with $c_0$ being the Gaussian covariance function (4.6), the Matérn covariance function (4.8), and the Cauchy covariance function (4.10). Note that these covariance functions have very distinct behaviors both at the origin and in the tails, see also Figure 3. We also consider a covariance function of the form (4.4) where $Y$ is a Gaussian random field with a Matérn covariance function. We denote this covariance function LG-Matérn since it corresponds to the case of a log Gaussian random intensity function $\Lambda_0$. In this case $\sigma_0^2 = \exp[\mathbb{V}\mathrm{ar}\, Y(u)] - 1$. As measures of fit for these models, we use the maximal values of the composite likelihoods $\mathrm{CL}_1$ and $\mathrm{CL}_2$.

Regarding the weight function $h$ we tried both (5.4) and (5.5) with $t = 125$ in (5.2). The integrals in the composite likelihoods $\mathrm{CL}_1$ and $\mathrm{CL}_2$ were approximated using a Berman-Turner quadrature scheme consisting of data points and $200 \times 100$ dummy points over the observation window $W = [0, 1000] \times [0, 500]$. The ranking of the models according to $\mathrm{CL}_2$ did not depend on the choice of $h$-function (except for a single swap of ranks between Cauchy and Matérn in case of the additive model for Lonchocarpus). However, for the log linear models we in general obtained somewhat smaller estimates of $\sigma_0^2$ with (5.4) than with (5.5) and vice versa for the additive models. According to a model check for the best fitting log linear models (see below) the estimates obtained with (5.5) gave the best fit. In the following we restrict attention to the results obtained with (5.5). Table 1 shows the parameter estimates, the maximal composite likelihood values, and the estimated $R^2$ for each species and model.

Considering first $\beta_{1:2}$ for each species, the regression parameters have similar signs and relative magnitudes for the log linear model and the additive model. However, the maximal first-order composite likelihood $\mathrm{CL}_1$ is always largest for the log linear model. The fitted $\widetilde{Z}$ are shown in Figure 2 for each species.

Considering $\mathrm{CL}_2$ and comparing log linear and additive models for each species, the log linear models always yield the largest $\mathrm{CL}_2$ values. Regarding the choice of

**Table 1:** Estimates of $\beta$, $\psi$ and $R^2$ and maximal values of the composite likelihoods CL$_1$ and CL$_2$. For each species, the first four rows in the last three columns correspond to the log linear model with Gaussian, Cauchy, Matérn and LG-Matérn covariance functions. The next four rows are for the additive model with the same covariance models. For LG-Matérn, abusing notation, $\hat{\psi}$ denotes the parameter estimate for the covariance function of the Gaussian field $Y$. In the first column, the estimate $\hat{\rho}$ is given by $N(W)/|W|$. The second column also shows estimates of $\hat{\sigma}^2_{\exp\tilde{Z}}$ and $\hat{\sigma}^2_{\tilde{Z}}$.

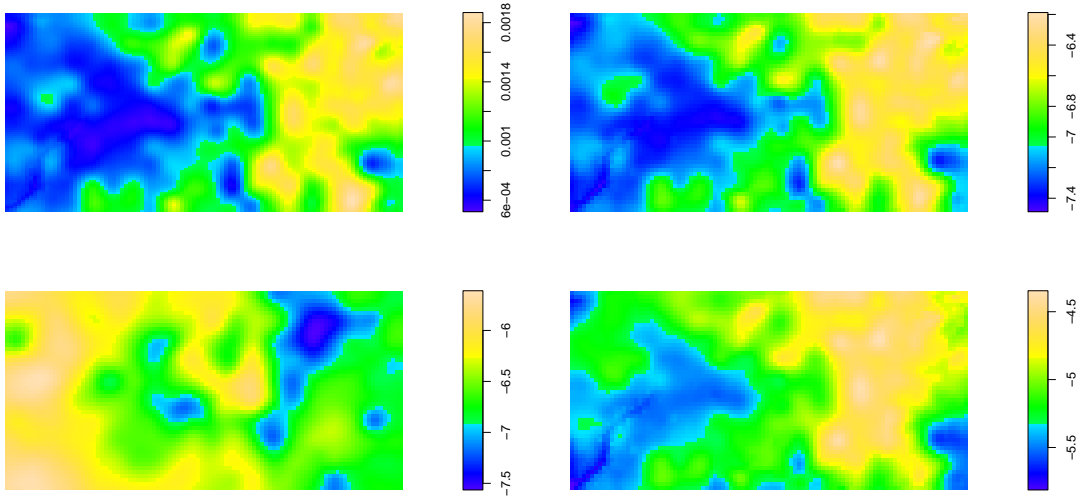| Species | $\hat{\beta}_{1:2}$ / CL$_1(\hat{\beta})$ | $\hat{\psi} = (\hat{\sigma}_0^2, \hat{\eta}, \hat{\nu})$ | CL$_2(\hat{\psi}\|\hat{\beta}) - C$ | $R^2$ |
|---|---|---|---|---|
| a | (0.02, 0.005) | $(13.7, 4.4, \infty)$ | 27 438.39 | 0.01 |
| | | $(11.9, 4.9, -)$ | 28 744.97 | 0.01 |
| | $-4117.7$ | $(8.5, 4.7, 0.69)$ | 28 507.05 | 0.01 |
| **Acalypha** | $\hat{\sigma}^2_{\exp\tilde{Z}} = 0.13 \times 10^{-6}$ | $(2.4, 5.6, 1.02)$ | 28 675.13 | 0.01 |
| | $(15.5, 4.9) \times 10^{-6}$ | $(11.6 \times 10^{-6}, 4.1, \infty)$ | 0 | 0.01 |
| | | $(8.1 \times 10^{-6}, 5.2, -)$ | 1724.32 | 0.01 |
| $\hat{\rho} = 1056 \times 10^{-6}$ | $-4119.7$ | $(6.1 \times 10^{-6}, 5.8, 0.56)$ | 1128.50 | 0.02 |
| $C = -6291053.0$ | $\hat{\sigma}^2_{\tilde{Z}} = 0.11 \times 10^{-6}$ | $(6.1 \times 10^{-6}, 5.8, 0.56)$ | 1128.50 | 0.02 |
| | $(-0.03, -0.16)$ | $(1.1, 28.4, \infty)$ | 82 006.98 | 0.11 |
| | | $(1.8, 18.4, -)$ | 82 174.76 | 0.07 |
| | $-6117.6$ | $(2.0, 14.0, 0.65)$ | 82 326.85 | 0.06 |
| **Lonchocarpus** | $\hat{\sigma}^2_{\exp\tilde{Z}} = 0.45 \times 10^{-6}$ | $(1.1, 14.8, 0.86)$ | 82 344.08 | 0.06 |
| | $(-38.2, -193.3) \times 10^{-6}$ | $(1.5 \times 10^{-6}, 36.6, \infty)$ | 0 | 0.17 |
| | | $(2.1 \times 10^{-6}, 27.4, -)$ | 934.78 | 0.12 |
| $\hat{\rho} = 1672 \times 10^{-6}$ | $-6121.9$ | $(2.8 \times 10^{-6}, 23.1, 0.41)$ | 702.26 | 0.09 |
| $C = -6168628.5$ | $\hat{\sigma}^2_{\tilde{Z}} = 0.29 \times 10^{-6}$ | $(2.8 \times 10^{-6}, 23.1, 0.41)$ | 702.26 | 0.09 |
| | $(0.03, 0.004)$ | $(0.25, 69.8, \infty)$ | 5012.70 | 0.28 |
| | | $(0.43, 43.1, -)$ | 5223.48 | 0.18 |
| | $-19693.0$ | $(0.76, 48.2, 0.22)$ | 5342.78 | 0.11 |
| **Capparis** | $\hat{\sigma}^2_{\exp\tilde{Z}} = 4.84 \times 10^{-6}$ | $(0.59, 49.7, 0.26)$ | 5361.99 | 0.11 |
| | $(193.2, 24.8) \times 10^{-6}$ | $(10.0 \times 10^{-6}, 70.2, \infty)$ | 0 | 0.29 |
| | | $(15.0 \times 10^{-6}, 48.7, -)$ | 285.51 | 0.21 |
| $\hat{\rho} = 6598 \times 10^{-6}$ | $-19700.1$ | $(28.8 \times 10^{-6}, 51.6, 0.21)$ | 466.02 | 0.12 |
| $C = -5089810.54$ | $\hat{\sigma}^2_{\tilde{Z}} = 4.06 \times 10^{-6}$ | $(28.8 \times 10^{-6}, 51.6, 0.21)$ | 466.02 | 0.12 |

**Figure 2:** The fitted regression term, $\widehat{\widetilde{Z}}$, for Acalypha under the additive model and the log linear model (top left and right), and for Lonchocarpus and Capparis under the log linear model (bottom left and right).

covariance function, the best fit for Acalypha is obtained with the Cauchy covariance function while for Lonchocarpus and Capparis, the LG-Matérn covariance function performs better, followed by the Matérn covariance function. Both for the log linear and the additive models, the smallest $CL_2$ is obtained with the Gaussian covariance function. This suggests that the fast decaying pair correlation function obtained with the Gaussian covariance function is not suitable for the tropical rain forest data. For the additive model, the LG-Matérn and the Matérn covariance functions are almost identical because in this case $\hat{\sigma}_0^2$ is small and $\exp(x) - 1 \approx x$ when $x \approx 0$.

Regarding $R^2$, the estimates vary considerably across models. However, the overall qualititative conclusion is stable: the contribution of the environment is smallest for Acalypha and largest for Capparis. This may be linked to the different modes of seed dispersal of the species, see Waagepetersen and Guan (2009). The $R^2$ obtained with the best fitting models are 0.01, 0.06, and 0.11 for Acalypha, Lonchocarpus, and Capparis.

The fitted covariance functions for the log linear models and for all species are shown in Figure 3. For all species, the Gaussian covariance function differ much from the other three covariance functions both a the origin and in the tail. For Lonchocarpus and Capparis, the fitted Matérn and LG-Matérn covariance functions appear rather similar.

We used the non-parametric estimate of the $g_0$-function (mentioned in Section 5.3) as a summary statistic for model assessment and computed pointwise 90 % confidence bands for this statistic using simulations from the best fitting log linear models according to $CL_2$. The non-parametric estimates of $c_0(\cdot) = g_0(\cdot) - 1$ and 90 % pointwise confidence bands are shown in Figure 4 and do not provide evidence against the fitted models.
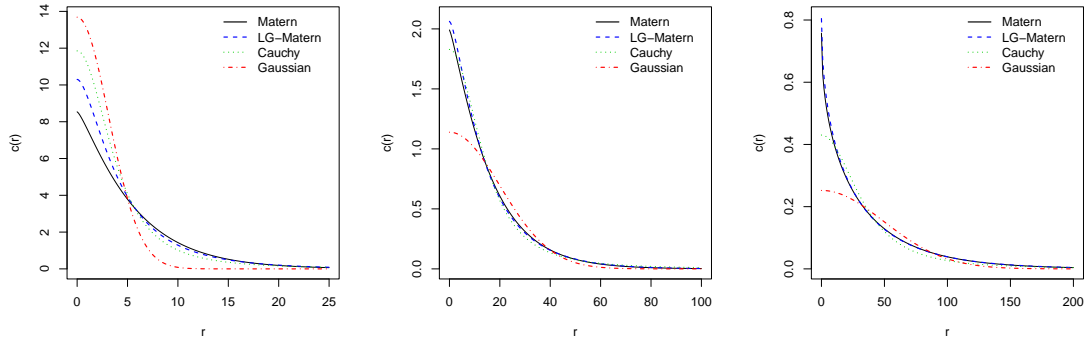
**Figure 3:** The various fitted covariance functions $c_0$ in case of the log linear model for Acalypha (left), Lonchocarpus (middle), and Capparis (right).
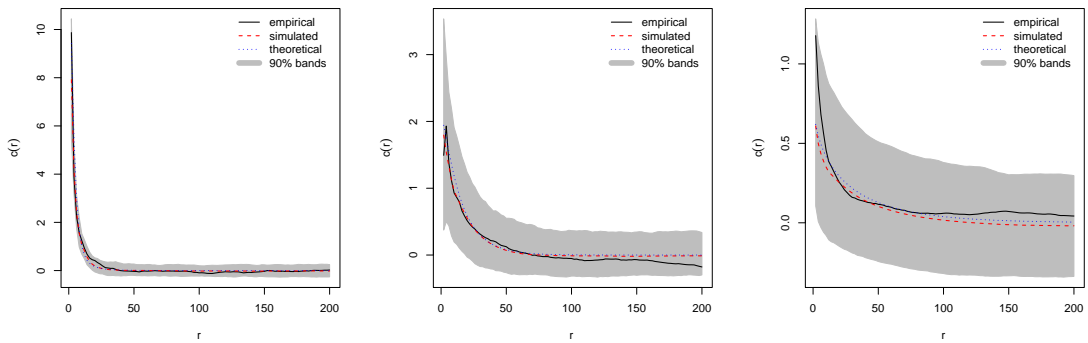


**Figure 4:** Non-parametric estimates of $c_0$ (solid line) and $90\%$ pointwise confidence bands (gray area) obtained from 100 simulations under best fitting models. The red dashed line shows mean of simulated non-parametric estimates and the blue dotted line shows the parametric estimate of $c_0$. Left: Acalypha, middle: Lonchocarpus, and right: Capparis.

# 7 Discussion

In this paper we introduced a method for decomposing variance for spatial Cox processes. As can be seen from our data analysis, the results can be sensitive to the choice of the pair correlation function. Fortunately, the flexible class of Matérn and Cauchy pair correlation functions allow us to compare results from a wide range of pair correlation functions. We can then report results obtained with the best fitting model according to the second-order composite likelihood criterion.

Concerning the second-order composite likelihood estimation, further studies regarding numerical implementation and choice of $h$-function seem needed. For the Matérn model, the joint estimation of $(\sigma_0^2, \eta, \nu)$ is computationally demanding and simulation studies indicate that the statistical properties of the estimates can be poor. For routine use, a more feasible approach is to maximize only with respect to $\sigma_0^2$ and $\eta$ for each $\nu$ in a moderate collection of $\nu$ values.

The Thomas process has enjoyed much popularity in the point process literature. However, at least for the tropical rain forest data considered in this paper, it seems inferior to models with a more slowly decaying pair correlation function. For all the data examples considered, the log linear model provided a better fit than the additive model. This is perhaps not so surprising since the log linear model has

an appealing interpretation in terms of survival of seedlings. On the other hand, variance decomposition is more straightforward for the additive model than for the log linear model.

Our proposed variance decomposition procedure assumes a stationary random environment. This may e.g. not be tenable for Acalypha because the fitted $\widetilde{Z}$ in $W$ shows a trend from right to left. Although this does not necessarily invalidate the stationarity assumption on $\widetilde{Z}$, it at least implies that $W$ may be too small to allow a precise estimate of the variance for $\widetilde{Z}$. Further research is required to handle the case where $\widetilde{Z}$ is not stationary but perhaps satisfies a weaker assumption of intrinsic stationarity.

# Acknowledgements

# References

Baddeley, A. and Turner, R. (2000). Practical maximum pseudolikelihood for spatial point patterns. *Australian and New Zealand Journal of Statistics* **42,** 283–322.

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

Baddeley, A. J., Turner, R., Møller, J., and Hazelton, M. (2005). Residual analysis for spatial point processes (with discussion). *Journal of the Royal Statistical Society, Series B* **67,** 617–666.

Best, N. G., Ickstadt, K., and Wolpert, R. L. (2000). Spatial Poisson regression for health and exposure data measured at disparate resolutions. *Journal of the American Statistical Society* **95,** 1076–1088.

Bolthausen, E. (1982). On the central limit theorem for stationary mixing random fields. *The Annals of Probability* **10,** 1047–1050.

Chilès, J.-P. and Delfiner, P. (1999). *Geostatistics - modeling spatial uncertainty*. Probability and Statistics. Wiley.

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

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

Guan, Y. (2006). A composite likelihood approach in fitting spatial point process models. *Journal of the American Statistical Association* **101,** 1502–1512.

Guyon, X. (1991). *Random fields on a network*. Springer-Verlag, New York.

Hubbell, S. P. and Foster, R. B. (1983). Diversity of canopy trees in a neotropical forest and implications for conservation. 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.

Matérn, B. (1986). *Spatial Variation*. Lecture Notes in Statistics 36, Springer-Verlag, Berlin.

McKay, A. T. (1932). A Bessel function distribution. *Biometrika* **24,** 39–44.

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

Møller, J. and Waagepetersen, R. P. (2004). *Statistical inference and simulation for spatial point processes*. Chapman and Hall/CRC, Boca Raton.

Nadarajah, S. and Gupta, A. K. (2006). Elliptically symmetric Bessel distribution: the distribution of $XY$. *Tamkang Journal of Mathematics* **37,** 185–192.

Schoenberg, F. P. (2005). Consistent parametric estimation of the intensity of a spatial-temporal point process. *Journal of Statistical Planning and Inference* **128,** 79–93.

Stein, M. L. (1999). *Spatial interpolation, some theory for kriging*. Springer Verlag, New York.

Waagepetersen, R. (2007). An estimating function approach to inference for inhomogeneous Neyman-Scott processes. *Biometrics* **63,** 252–258.

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

# A    Asymptotic properties of parameter estimates

In this appendix we discuss asymptotic properties of the parameter estimates given in Section 5. To do this we consider an increasing sequence of observation windows $W_n \subset \mathbb{R}^2$ and let $G_n = W_n \cap \mathbb{Z}^2$ be an increasing sequence of sampling grids for $Z$. We moreover let $\hat{\theta}_n = (\hat{\beta}_n, \hat{\psi}_n)$ denote the vector of composite likelihood estimates and $(\hat{\sigma}^2_{\widetilde{Z},n}, \hat{\sigma}^2_{\exp \widetilde{Z},n}, \hat{\rho}_{\widetilde{Z},n}, \hat{\rho}_{\exp \widetilde{Z},n})$ the parameter estimates (5.6)–(5.8) obtained from $X_n = X \cap W_n$ and $Z^n = \{Z(u)\}_{u \in G_n}$. We may view $\hat{\theta}_n$ as a solution of an unbiased estimating equation $u_n(\theta) = 0$ where $u_n$ is obtained by concatenating the gradients of the log composite likelihoods evaluated over $W_n$.

Under appropriate conditions, $|W_n|^{1/2}(\hat{\theta}_n - \theta^*)J_n(\hat{\theta}_n)\Sigma_n^{-1/2}$ has the same asymptotic $N(0, I)$ distribution as $|W_n|^{-1/2}u_n(\theta^*)\Sigma_n^{-1/2}$ where $J_n(\theta) = -|W_n|^{-1}\mathrm{d}u_n(\theta)/\mathrm{d}\theta^\mathsf{T}$, $\theta^*$ denotes the 'true' value of $\theta$, and $\Sigma_n$ is the covariance matrix of $|W_n|^{-1/2}u_n(\theta^*)$. For details we refer to Waagepetersen and Guan (2009) which contains a thorough asymptotic study of a two-step procedure based on the first-order composite likelihood and minimum contrast estimation using the $K$-function. If we assume further that $Z$ is a suitably mixing random field (e.g. Guyon, 1991), then $J_n(\hat{\theta}_n)$ and $\Sigma_n$ (which can be viewed as spatial averages) converge to fixed matrices $J$ and $\Sigma$.

From mixing of $Z$ we moreover obtain the central limit theorem (Bolthausen, 1982; Guyon, 1991) $M_n^{1/2}(\bar{f}_n^* - \mu_f) \to N(0, \Sigma_f)$ for estimators $\bar{f}_n^*$ of the form $\bar{f}_n^* = M_n^{-1} \sum_{u \in G_n} f[\widetilde{Z}^*(u)]$ where $f = (f_1, \ldots, f_q)$ is a vector function, $\widetilde{Z}^*(u) = \beta^* Z(u)^\mathsf{T}$, $M_n$ is the cardinality of $G_n$, $\mu_f = \mathbb{E}f[\widetilde{Z}^*(0)]$ and $\Sigma_f$ is the asymptotic variance matrix of $M_n^{1/2}\bar{f}_n^*$. By the unbiasedness of $u_n$ given $Z$ it follows that

$$\mathbb{C}\mathrm{ov}[\bar{f}_n^*, u_n(\theta^*)] = \mathbb{E}\,\mathbb{C}\mathrm{ov}[\bar{f}_n^*, u_n(\theta^*)|Z] + \mathbb{C}\mathrm{ov}\{\mathbb{E}[\bar{f}_n^*|Z], \mathbb{E}[u_n(\theta^*)|Z]\} = 0.$$

Thus, $|W_n|^{1/2}(\hat{\theta}_n - \theta^*)$ and $M_n^{1/2}(\bar{f}_n^* - \mu_f)$ are asymptotically zero-mean normal and independent with covariance matrices $(J^\mathsf{T})^{-1}\Sigma J^{-1}$ and $\Sigma_f$.

Finally $(\hat{\sigma}^2_{\widetilde{Z},n}, \hat{\sigma}^2_{\exp \widetilde{Z},n}, \hat{\rho}_{\widetilde{Z},n}, \hat{\rho}_{\exp \widetilde{Z},n})$ can be expressed in terms of

$$\hat{\bar{f}}_n = M_n^{-1} \sum_{u \in G_n} f[\widehat{\widetilde{Z}}_n(u)]$$

for a vector function $f$ involving squares and exponentials. Since $\widehat{\widetilde{Z}}_n(u) = \hat{\beta}_n Z(u)^\mathsf{T}$ and using a Taylor expansion of $g_u(\beta) = f[\beta Z(u)^\mathsf{T}]$ around $\beta^*$ we obtain that $M_n^{1/2}(\hat{\bar{f}}_n - \mu_f)$ has the same asymptotic distribution as

$$M_n^{1/2}(\bar{f}_n^* - \mu_f) + |W_n|^{1/2}(\hat{\beta}_n - \beta^*)\mathbb{E}\{Z(0)^\mathsf{T}f'[\widetilde{Z}^*(0)]\}$$

where $f'(z)$ is the vector of derivatives $\mathrm{d}f_i(z)/\mathrm{d}z$. Thus we also obtain asymptotic normality for $(\hat{\sigma}^2_{\widetilde{Z},n}, \hat{\sigma}^2_{\exp \widetilde{Z},n}, \hat{\rho}_{\widetilde{Z},n}, \hat{\rho}_{\exp \widetilde{Z},n})$.