Composite likelihood inference for space-time point processes

Abdollah Jalilian

Razi University, Iran

Lancaster Ecology and Epidemiology Group, Lancaster University, United Kingdom

email: stat4aj@gmail.com

and

Francisco Cuevas-Pacheco

Universidad Técnica Federico Santa María, Chile

email: francisco.cuevas@usm.cl

and

Ganggang Xu

University of Miami, United States of America

email: gangxu@bus.miami.edu

and

Rasmus Waagepetersen

Skjernvej 4A, DK-9220 Aalborg

Aalborg University

Denmark

email: rw@math.aau.dk

Summary: The dynamics of a rain forest is extremely complex involving births, deaths and growth of trees with complex interactions between trees, animals, climate, and environment. We consider the patterns of recruits (new trees) and dead trees between rain forest censuses. For a current census we specify regression models for the conditional

intensity of recruits and the conditional probabilities of death given the current trees and spatial covariates. We estimate regression parameters using conditional composite likelihood functions that only involve the conditional first order properties of the data. When constructing assumption lean estimators of covariance matrices of parameter estimates we only need mild assumptions of decaying conditional correlations in space while assumptions regarding correlations over time are avoided by exploiting conditional centering of composite likelihood score functions. Time series of point patterns from rain forest censuses are quite short while each point pattern covers a fairly big spatial region. To obtain asymptotic results we therefore use a central limit theorem for the fixed timespan - increasing spatial domain asymptotic setting. This also allows us to handle the challenge of using stochastic covariates constructed from past point patterns. Conveniently, it suffices to impose weak dependence assumptions on the innovations of the space-time process. We investigate the proposed methodology by simulation studies and an application to rain forest data.

KEY WORDS: Central limit theorem; composite likelihood; conditional centering; estimating function; point process; spatio-temporal.

1. Introduction

This paper develops composite likelihood methodology for analysing a discrete-time continuous-space time series of spatial point patterns. Our primary motivation is the need to understand the complex spatio-temporal development of a rain forest ecosystem. Essentially, this process can be characterized in terms of growth, recruitment and mortality (Wolf, 2005; Wiegand et al., 2009; Shen et al., 2013; Kohyama et al., 2018). Each of these processes depend on species specific factors (e.g. genetics, light requirements, seed dispersal), inter- and intraspecies interactions (e.g. competition), interactions with animals, as well as exogeneous factors such as climate or weather, soil properties, and topography (Rüger et al., 2009; Häbel et al., 2019; Hiura et al., 2019). We leave aside the aspect of tree growth and confine ourselves to considering recruitment and mortality.

Extensive rain forest tree census data have been collected within a global network of forest research sites including the Barro Colorado Island plot with eight censuses collected with 5 year intervals (Condit et al., 2019). In the ecological literature (e.g. Hubbell et al., 2001; Rüger et al., 2009; Johnson et al., 2017; Zhu et al., 2018; Zuleta et al., 2022) there is much interest in the spatial patterns of trees that were recruited or died between consecutive censuses (see Figure 3). In particular, biologists wants to assess how the intensity of recruits and how the death probabilities depend on various covariates including soil properties, topography, and covariates representing the influence of trees from previous censuses. The latter type of covariates can be regarded as stochastic or auto-regressive since they depend on the process of recruitment and mortality up to the current time interval. We model the intensity of recruits and the death probabilities using parametric log-linear and logistic regression models. This enables biologists to study the impact of covariates through estimates of regression parameters in these models.

We estimate regression parameters using conditional composite likelihood functions. This

is computionally very efficient since the composite likelihoods can be maximized using standard efficient logistic regression software. We obtain asymptotic distributions of regression parameter estimates using asymptotic results for sequences of conditionally centered random fields with increasing spatial domain but fixed time horizon (Jalilian et al., 2024). The aforementioned stochastic covariates do not violate conditional centering and are easily accommodated by the asymptotic framework.

To enhance robustness to model misspecification, we, inspired by approaches in spatial econometrics (Conley, 2010), avoid parametric modeling of second-order properties or further distributional characteristics of the spatial patterns of recruitment and mortality. Instead we estimate variance matrices of parameter estimates by exploiting conditional centering of the conditional composite likelihood score functions and model free estimators (Conley, 2010; Coeurjolly and Guan, 2014) of conditional variance matrices for the score functions. For the space-time correlation structure we only need mild assumptions of spatial decay of correlations for recruits and death events in each census interval conditional on the previous state of the forest. The estimation of variance matrices is computationally efficient. This is because it only involves pairs of recruits or of dead trees that are close in space within each of the time intervals rather than all pairs of points across all spatial distances and all time intervals.

Logistic regression models for mortality are used extensively in the ecological literature (Comita and Hubbell, 2009; Johnson et al., 2017; Zhu et al., 2018; Zuleta et al., 2022) with various approaches to handling spatial correlation including block bootstrap or random effects associated with grid cells partitioning the study region. Rathbun and Cressie (1994) and Hubbell et al. (2001) instead model spatial correlation between deaths using auto-logistic models and implement approximate maximum likelihood estimation using computationally heavy Markov chain Monte Carlo (MCMC) methods.

Regarding recruits, Rathbun and Cressie (1994) consider a single generation and explicitly model dependence between recruits using a spatial Cox point process and implement parameter estimation using MCMC. Wiegand et al. (2009) and Getzin et al. (2014) use spatial point process summary statistics to investigate associations between recruits and adults and test independence by randomization of the recruits. Regression modeling is not used in these papers. Rüger et al. (2009) use a negative binomial regression to model effects of light availability on grid cell counts of recruits implementing inference using MCMC. Brix and Møller (2001) consider a multi-type spatio-temporal log Gaussian Cox process for modeling weed recruits on a barley field and use minimum contrast parameter estimation. However, their approach does not accommodate covariates. Within rain forest ecology, May et al. (2015) use approximate Bayesian computation for a discrete-time continuous-space model based on ecological neutral theory (Hubbell, 2001).

In contrast to previous approaches to analyzing recruitment and mortality we avoid potentially restrictive assumptions regarding the space-time correlation structure, we avoid dependence on choice of grids for random effects or counts, and we avoid computationally intensive bootstrap, approximate Bayesian, or MCMC methods. We demonstrate the usefulness and validity of our approach by application to simulated and rain forest data.

2. Space-time model for rain forest census data

We consider marked spatial point pattern datasets originating from censuses that record location $u = (u_1, u_2) \in \mathbb{R}^2$, species $s \in \{1, \ldots, p\}$ and possibly further marks $m \in \mathcal{M}$ for all trees in a research plot. The mark m for a tree could represent size in terms of diameter at breast height and we assume for specificity that $\mathcal{M} = \mathbb{R}_+$. Such data can be viewed as a time series of multivariate marked space-time point processes, $X = \{X_t\}_{t \in T}$, where $X_t = (X_t^{(1)}, \ldots, X_t^{(p)})$ and $X_t^{(s)}$ is the marked point process consisting of marked points $x = (u, m)$ at time $t \in T$ for species s (Diggle, 2013; González et al., 2016). The

distribution of the point process X may depend on a space-time process $Z = \{Z_t\}_{t \in \mathbb{R}}$, where $Z_t = \{\mathbf{Z}_t(u)\}_{u \in \mathbb{R}^2}$ and $\mathbf{Z}_t(u) = (Z_t^{(1)}(u), \ldots, Z_t^{(q)}(u))^{\mathsf{T}}$ is a vector of $q \geqslant 1$ environmental covariates at location $u \in \mathbb{R}^2$ at time t.

We assume that the time index set T consists of equidistant time points $t_k = \Delta k$, $k = 0, 1, \ldots$, for some $\Delta > 0$, where we henceforth take $\Delta = 1$. For any $k \geqslant 0$, we let the 'observation history' $\mathcal{H}_k$ denote the information given by $X_0, \ldots, X_k$ and $Z_0, \ldots, Z_k$. In practice we only observe X and Z within $W = \widetilde{W} \times \mathcal{M}$ for a bounded $\widetilde{W} \subset \mathbb{R}^2$ (typically a rectangle) and for a finite number $K + 1$ of observation times $0, 1, \ldots, K$. More precisely, for an observation time $0 \leqslant k \leqslant K$, we observe those marked points $x = (u, m)$ in $X_k^{(s)}$ and covariate vectors $\mathbf{Z}_k(v)$ where $u, v \in \widetilde{W}$.

We focus on statistical modeling of recruits and deaths of trees for a single species but our models for recruit intensities and death probabilities may in general depend on existing trees of other species. Without loss of generality we consider the first species $s = 1$ and for two consecutive observation times $k-1$ and k we let $B_k = X_k^{(1)} \backslash X_{k-1}^{(1)}$ and $D_k = X_{k-1}^{(1)} \backslash X_k^{(1)}$ denote the recruitment and mortality processes of species $s = 1$ over the interval $]k-1, k]$.

The spatial pattern of recruits B_k often exhibit clustering around parent plants due to seed dispersal and favorable soil conditions (Wiegand et al., 2009). Moreover, when regeneration occurs in canopy gaps, recruits tend to aggregate and be positively correlated with dead trees (Wolf, 2005). The spatial pattern of deaths D_k is influenced by biotic and abiotic factors such as intra- and inter-specific interactions as well as environmental factors (Shen et al., 2013). The negative density-dependent mortality hypothesis for example implies clustering of dead individuals and repulsion between surviving and dead individuals (Wiegand and Moloney, 2013, p. 210).

In the next section we specify models for the intensity functions of recruits B_k and probabilities of death for trees $x \in X_{k-1}^{(1)}$ over intervals $]k-1, k]$, conditional on $\mathcal{H}_{k-1}$. We

do not impose assumptions regarding the dependence structure of the recruit and mortality processes until Sections 4 and 5 where spatially decaying correlations and spatial mixing are needed for the conditional distributions of recruits and deaths at time k conditional on $\mathcal{H}_{k-1}$, $k = 1, \ldots, K$.

2.1 *Models for recruit intensity and death probabilities*

Given $\mathcal{H}_{k-1}$, we model the recruits B_k as an inhomogeneous point process on $\mathbb{R}^2 \times \mathcal{M}$ with log linear intensity function of the form

$$\zeta_k(x|\mathcal{H}_{k-1}) = f(m)\exp\left[\beta_{\mathrm{b}0,k} + \mathbf{Z}_{k-1}^\mathsf{T}(u)\boldsymbol{\beta}_{\mathrm{b}} + \mathbf{c}_{k-1}(x)^\mathsf{T}\boldsymbol{\gamma}_{\mathrm{b}}\right], \quad x = (u, m) \in \mathbb{R}^2 \times \mathcal{M}, \quad (1)$$

where f is a probability density for the marks, $\beta_{\mathrm{b}0,k}$ is a time-dependent intercept, $\boldsymbol{\beta}_{\mathrm{b}} \in \mathbb{R}^q$ and $\boldsymbol{\gamma}_{\mathrm{b}} \in \mathbb{R}^p$ are vectors of regression parameters, and $\mathbf{c}_{k-1}(x) = (c_{k-1}^l(x))_{l=1}^p$ where $c_{k-1}^l(\cdot)$ represents the 'influence' of trees of species l on the intensity of recruits. In general, the intensity $\zeta_k(\cdot|\mathcal{H}_{k-1})$ may not capture all sources of variation of the recruits. It is therefore important to recognize the possibility of stochastic dependence between recruits conditional on $\mathcal{H}_{k-1}$. In forest ecology, the marks of recruits are often quite similar and in the following we focus on estimation of the log linear part treating $f(m)$ as a known density.

For $x = (u, m) \in X_{k-1}^{(1)}$ we define $I_k(x) = 1[x \in D_k]$ to be an indicator of death. The $I_k(x)$, $x \in X_{k-1}^{(1)}$, are Bernoulli random variables with death probabilities $p_k(x|\mathcal{H}_{k-1}) = P(I_k(x) = 1|\mathcal{H}_{k-1})$. We do not assume that the death indicators are independent. We model the death probabilities by logistic regressions $p_k(x|\mathcal{H}_{k-1}) = \exp\left[\eta_k(x|\mathcal{H}_{k-1})\right]/\{1 + \exp\left[\eta(x|\mathcal{H}_{k-1})\right]\}$ with

$$\eta_k(x|\mathcal{H}_{k-1}) = \beta_{\mathrm{d}0,k} + \alpha m + \mathbf{Z}_{k-1}^\mathsf{T}(u)\boldsymbol{\beta}_{\mathrm{d}} + \mathbf{d}_{k-1}(x)^\mathsf{T}\boldsymbol{\gamma}_{\mathrm{d}}, \quad (2)$$

where $\beta_{\mathrm{d}0,k}$ is a time-dependent intercept, α is a regression parameter for the mark m, $\boldsymbol{\beta}_{\mathrm{d}} \in \mathbb{R}^q$ and $\boldsymbol{\gamma}_{\mathrm{d}} \in \mathbb{R}^p$ are vectors of regression parameters, and $\mathbf{d}_{k-1}(x) = (d_{k-1}^l(x))_{l=1}^p$ where $d_{k-1}^l(x)$ represents the 'influence' of trees of species l on death of a tree x.

Unlike the spatial covariates $\mathbf{Z}_{k-1}(u)$ which are commonly assumed to be deterministic, we permit the influence covariates $\mathbf{c}_{k-1}$ and $\mathbf{d}_{k-1}$ to be stochastic being constructed (next subsection) from past marked point patterns X_{k-1}.

2.2 *Stochastic covariates for influence of existing trees*

Recruits may be positively dependent on previous conspecific trees since recruits arise from parent tree seed dispersal. Conversely, the impact of the remaining species could be negative due to competition for light and other resources. There is a rich ecological literature on models for seed dispersal kernels (Nathan et al., 2012; Bullock et al., 2019) and competition indices (e.g. Burkhart and Tomé, 2012; Britton et al., 2023).

We assume that the influence of an existing conspecific forest stand on a recruit $x = (u, m)$ is a function of the distance from x to the nearest neighbour in the forest stand. That is,

$$c_{k-1}^1(x) = \exp\{-[d(x, X_{k-1}^{(1)})]/\psi_1)^2\}, \tag{3}$$

with $d(x, X_{k-1}^{(1)}) = \min_{(u',m') \in X_{k-1}^{(1)}} \|u-u'\|/m'$ being a mark-weighted spatial distance between the marked point $x = (u, m)$ and the point pattern $X_{k-1}^{(1)}$, while $\psi_1 > 0$ controls the range of effect of the existing conspecific trees. Thus, the influence of the existing trees is determined by a Gaussian dispersal kernel placed at the location of the nearest existing tree.

For the impact of competition from existing trees we follow Burkhart and Tomé (2012) and use for $c_{k-1}^l(x)$, $l \neq 1$, and $d_{k-1}^l(x)$, $l = 1, \ldots, p$, $x = (u, m)$, indices of the form

$$\sum_{(u',m') \in X_{k-1}^{(l)} \setminus \{(u,m)\}} \frac{m'}{m} \exp[-(\|u - u'\|/\kappa_l)^2], \quad \kappa_l > 0. \tag{4}$$

For the data example in Section 7, m represents diameter at breast height which does not vary much between recruits. Further, diameter at breast height is missing for a large proportion of the dead trees. We therefore in Section 7 modified (4) by omitting division by m.

A huge variety of plausible models for influence of conspecific trees and competition could be proposed and compared e.g. in terms of resulting maximized composite likelihoods.

However, the exact choice of model for the influence of existing trees is not our primary focus and we leave it to future users to investigate further models.

Using stochastic covariates such as $\mathbf{c}_k(\cdot)$ and $\mathbf{d}_k(\cdot)$ in the models (1) and (2) is conceptually straightforward but challenging from a theoretical point of view. For example, the resulting intensity in (1) becomes stochastic, bearing a resemblance to the conditional intensity of the Hawkes process. However, the conditional intensity (1) allows for a more flexible dependence structure on the past $\mathcal{H}_{k-1}$. To address the theoretical challenges, we propose in Sections 4 and 5 a framework based on conditional centering for estimating functions.

2.3 *Campbell formulas and pair correlation function*

In Sections 3 and 4, for recruits B_k conditional on $\mathcal{H}_{k-1}$, we use the so-called first and second order Campbell formulas

$$\mathbb{E}[\sum_{x \in B_k} h_1(x)|\mathcal{H}_{k-1}] = \int h_1(x)\zeta_k(x|\mathcal{H}_{k-1})\mathrm{d}x$$

$$\mathbb{E}[\sum_{\substack{x,x' \in B_k: \\ x \neq x'}} h_2(x,x')|\mathcal{H}_{k-1}] = \int \int h_2(x,x')\zeta_k(x|\mathcal{H}_{k-1})\zeta_k(x'|\mathcal{H}_{k-1})g_{B_k}(x,x')\mathrm{d}x\mathrm{d}x'$$

for any non-negative functions h_1 and h_2 where $g_{B_k}(\cdot,\cdot)$ is the so-called pair correlation function of B_k conditional on $\mathcal{H}_{k-1}$ (the second equation is actually the defining equation for the pair correlation function). We also use the Campbell formulas for a Poisson process Y_k of intensity $\rho_0(\cdot)$ and independent of $\mathcal{H}_{k-1}$. The Campbell formulas then become

$$\mathbb{E}\sum_{x \in Y_k} h_1(x) = \int h_1(x)\rho_0(x)\mathrm{d}x \text{ and } \mathbb{E}\sum_{\substack{x,x' \in Y_k: \\ x \neq x'}} h_2(x,x') = \int \int h_2(x,x')\rho_0(x)\rho_0(x')\mathrm{d}x\mathrm{d}x'$$

since the pair correlation function of a Poisson process is one. Møller and Waagepetersen (2003) provide more details regarding Campbell formulas and pair correlation functions.

3. Composite likelihood estimation

Given observations X_k and Z_k, $k = 0, \ldots, K$, we infer regression parameters using estimating functions derived from composite likelihoods for the recruit and death patterns B_k and D_k, $k = 1, \ldots, K$. Let $\boldsymbol{\theta}_{\mathrm{b}} = (\beta_{\mathrm{b}0,1}, \ldots, \beta_{\mathrm{b}0,K}, \boldsymbol{\beta}_{\mathrm{b}}^{\mathsf{T}}, \boldsymbol{\gamma}_{\mathrm{b}}^{\mathsf{T}})^{\mathsf{T}}$ and $\boldsymbol{\theta}_{\mathrm{d}} = (\beta_{\mathrm{d}0,1}, \ldots, \beta_{\mathrm{d}0,K}, \alpha, \boldsymbol{\beta}_{\mathrm{d}}^{\mathsf{T}}, \boldsymbol{\gamma}_{\mathrm{d}}^{\mathsf{T}})^{\mathsf{T}}$ denote the parameter vectors for the recruit and mortality models. Since the models for recruits and deaths do not share parameters, we construct separate estimating functions for $\boldsymbol{\theta}_{\mathrm{b}}$ and $\boldsymbol{\theta}_{\mathrm{d}}$. The proposed estimating functions are unbiased, leading to consistent estimators of $\boldsymbol{\theta}_{\mathrm{b}}$ and $\boldsymbol{\theta}_{\mathrm{d}}$. Background on composite likelihood for intensity function estimation can be found in Møller and Waagepetersen (2017).

3.1 *Composite likelihoods for recruits at time k*

For the recruits B_k, we consider the following conditional composite log likelihood

$$\sum_{x \in B_k \cap W} \log \zeta_k(x | \mathcal{H}_{k-1}) - \int_W \zeta_k(x | \mathcal{H}_{k-1}) \mathrm{d}x$$

which would be the log likelihood if B_k was a Poisson process given $\mathcal{H}_{k-1}$. For a Poisson process, points occur independently of each other. Hence for the estimation of $\boldsymbol{\theta}_{\mathrm{b}}$ we ignore possible dependencies between recruits that are not explained by $\mathcal{H}_{k-1}$. Since $\int_{\mathcal{M}} f(m) \mathrm{d}m = 1$, the integral over W reduces to an integral over $\widetilde{W}$ involving just the log-linear part of $\zeta_k(\cdot | \mathcal{H}_{k-1})$. The score function (gradient) is

$$\sum_{x \in B_k \cap W} \frac{\nabla \zeta_k(x | \mathcal{H}_{k-1})}{\zeta_k(x | \mathcal{H}_{k-1})} - \int_W \nabla \zeta_k(x | \mathcal{H}_{k-1}) \mathrm{d}x, \tag{5}$$

where $\nabla \zeta_k(x | \mathcal{H}_{k-1})$ denotes the gradient of $\zeta_k(x | \mathcal{H}_{k-1})$ with respect to $\boldsymbol{\theta}_{\mathrm{b}}$. By the first order Campbell formula, the score function is conditionally centered meaning that it has expectation zero given $\mathcal{H}_{k-1}$.

In practice we need to estimate the integral in the score function (5). Following Waagepetersen (2008) and Baddeley et al. (2014), consider for each $k = 1, \ldots, K$, a dummy Poisson point process Y_k on W independent of B_k and $\mathcal{H}_{k-1}$ and with known intensity function

$\rho_0(x) = f(m)\rho(u)$ for $x = (u, m)$. By the first order Campbell formula for the union $B_k \cup Y_k$ with intensity $\zeta_k(\cdot|\mathcal{H}_{k-1}) + \rho_0(\cdot)$, the integral $\int_W \nabla\zeta_k(x|\mathcal{H}_{k-1})\mathrm{d}x$ can be estimated unbiasedly by $\sum_{x \in (B_k \cup Y_k) \cap W} \nabla\zeta_k(x|\mathcal{H}_{k-1})/\left[\zeta_k(x|\mathcal{H}_{k-1}) + \rho_0(x)\right]$. Crucially, after replacing the integral with the estimate, the resulting approximate score function

$$\mathbf{e}_{\mathrm{b},k}(\boldsymbol{\theta}_{\mathrm{b}}) = \sum_{x \in (B_k \cup Y_k) \cap W} \left[\frac{\nabla\zeta_k(x|\mathcal{H}_{k-1})}{\zeta_k(x|\mathcal{H}_{k-1})} 1[x \in B_k] - \frac{\nabla\zeta_k(x|\mathcal{H}_{k-1})}{\zeta_k(x|\mathcal{H}_{k-1}) + \rho_0(x)} \right] \tag{6}$$

is still conditionally centered. As explained in Waagepetersen (2008) and Baddeley et al. (2014), (6) is formally equivalent to a logistic regression score function and parameter estimates can be obtained using standard `glm` software.

3.2 *Composite likelihood for deaths at time k*

For the mortality process D_k we ignore possible dependencies between deaths and use the Bernoulli composite log likelihood function

$$\sum_{x \in X_{k-1}^{(1)} \cap W} \left\{ I_k(x) \log p_k(x|\mathcal{H}_{k-1}) + [1 - I_k(x)] \log[1 - p_k(x|\mathcal{H}_{k-1})] \right\}$$

$$= \sum_{x \in X_{k-1}^{(1)} \cap W} \left[I_k(x)\eta_k(x|\mathcal{H}_{k-1}) - \log\left\{ 1 + \exp\left[\eta_k(x|\mathcal{H}_{k-1})\right]\right\} \right]$$

with conditionally centered composite score function

$$\mathbf{e}_{\mathrm{d},k}(\boldsymbol{\theta}_{\mathrm{d}}) = \sum_{x \in X_{k-1}^{(1)} \cap W} \nabla\eta_k(x|\mathcal{H}_{k-1}) \left[I_k(x) - p_k(x|\mathcal{H}_{k-1}) \right], \tag{7}$$

where $\nabla\eta_k(x|\mathcal{H}_{k-1})$ denotes the gradient of $\eta_k(x|\mathcal{H}_{k-1})$ with respect to $\boldsymbol{\theta}_{\mathrm{d}}$.

3.3 *Conditional likelihoods and estimating functions based on all data*

Log composite likelihoods based on all generations of recruits and deaths are obtained by adding the one generation log composite likelihoods derived in the previous paragraphs. This results in estimating functions $\mathbf{e}_{\mathrm{o}}(\boldsymbol{\theta}_{\mathrm{o}}) = \sum_{k=1}^K \mathbf{e}_{\mathrm{o},k}(\boldsymbol{\theta}_{\mathrm{o}})$, $\mathrm{o} = \mathrm{b}, \mathrm{d}$. An estimate $\widehat{\boldsymbol{\theta}}_{\mathrm{o}}$ of $\boldsymbol{\theta}_{\mathrm{o}}$ is obtained by solving $\mathbf{e}_{\mathrm{o}}(\boldsymbol{\theta}_{\mathrm{o}}) = 0$. Conditional centering of (6) and (7), $\mathbb{E}[\mathbf{e}_{\mathrm{b},k}(\boldsymbol{\theta}_{\mathrm{b}})|\mathcal{H}_{k-1}] = 0$ and $\mathbb{E}[\mathbf{e}_{\mathrm{d},k}(\boldsymbol{\theta}_{\mathrm{d}})|\mathcal{H}_{k-1}] = 0$, implies that $\mathbf{e}_{\mathrm{b}}(\cdot)$ and $\mathbf{e}_{\mathrm{d}}(\cdot)$ are unbiased, $\mathbb{E}\mathbf{e}_{\mathrm{o}}(\boldsymbol{\theta}_{\mathrm{o}}) = 0$, $\mathrm{o} = \mathrm{b}, \mathrm{d}$.

4. Approximate covariance matrices of parameter estimates

According to standard estimating function theory (e.g. Sørensen, 1999, and Section 5), the approximate covariance matrix of $\widehat{\boldsymbol{\theta}}_{\mathrm{o}}$ is given by the inverse of the Godambe matrix $\mathbf{S}_{\mathrm{o}}(\boldsymbol{\theta}_{\mathrm{o}})\mathbf{V}_{\mathrm{o}}(\boldsymbol{\theta}_{\mathrm{o}})^{-1}\mathbf{S}_{\mathrm{o}}(\boldsymbol{\theta}_{\mathrm{o}})^{\mathsf{T}}$, where $\mathbf{V}_{\mathrm{o}}(\boldsymbol{\theta}_{\mathrm{o}}) = \mathbb{V}\mathrm{ar}\mathbf{e}_{\mathrm{o}}(\boldsymbol{\theta}_{\mathrm{o}})$ and $\mathbf{S}_{\mathrm{o}}(\boldsymbol{\theta}_{\mathrm{o}}) = -\mathbb{E}\frac{\mathrm{d}}{\mathrm{d}\boldsymbol{\theta}_{\mathrm{o}}^{\mathsf{T}}}\mathbf{e}(\boldsymbol{\theta}_{\mathrm{o}})$ are the variability and sensitivity matrices.

By iterated conditioning, $\mathbf{S}_{\mathrm{o}}(\boldsymbol{\theta}_{\mathrm{o}}) = -\sum_{k=1}^{K}\mathbb{E}\mathbb{E}\left[\frac{\mathrm{d}}{\mathrm{d}\boldsymbol{\theta}_{\mathrm{o}}^{\mathsf{T}}}\mathbf{e}_{\mathrm{o},k}(\boldsymbol{\theta}_{\mathrm{o}})|\mathcal{H}_{k-1}\right]$. Moreover, conditional centering of $\mathbf{e}_{\mathrm{o},k}(\boldsymbol{\theta}_{\mathrm{o}})$ and $\mathbf{e}_{\mathrm{o},k'}(\boldsymbol{\theta}_{\mathrm{o}'})$ implies that $\mathbb{C}\mathrm{ov}\left[\mathbf{e}_{\mathrm{o},k}(\boldsymbol{\theta}_{\mathrm{o}}), \mathbf{e}_{\mathrm{o},k'}(\boldsymbol{\theta}_{\mathrm{o}'})\right] = 0$, $\mathrm{o} = \mathrm{b}, \mathrm{d}$, whenever $k \neq k'$. This is very appealing since we avoid assumptions regarding the correlation structure across time for the space-time point process. It follows that $\mathbb{V}\mathrm{ar}\,\mathbf{e}_{\mathrm{o}}(\boldsymbol{\theta}_{\mathrm{o}}) = \sum_{k=1}^{K}\mathbb{V}\mathrm{ar}\,\mathbf{e}_{\mathrm{o},k}(\boldsymbol{\theta}_{\mathrm{o}})$, $\mathrm{o} = \mathrm{b}, \mathrm{d}$, and again by conditional centering, $\mathbb{V}\mathrm{ar}\,\mathbf{e}_{\mathrm{o},k}(\boldsymbol{\theta}_{\mathrm{o}}) = \mathbb{E}\mathbb{V}\mathrm{ar}\left[\mathbf{e}_{\mathrm{o},k}(\boldsymbol{\theta}_{\mathrm{o}})|\mathcal{H}_{k-1}\right]$.

To estimate the approximate covariance matrix of $\widehat{\boldsymbol{\theta}}_{\mathrm{o}}$ we thus need estimates of the conditional expectations $-\mathbb{E}\left[\frac{\mathrm{d}}{\mathrm{d}\boldsymbol{\theta}_{\mathrm{o}}^{\mathsf{T}}}\mathbf{e}_{\mathrm{o},k}(\boldsymbol{\theta}_{\mathrm{o}})|\mathcal{H}_{k-1}\right]$ and the conditional variances $\mathbb{V}\mathrm{ar}\left[\mathbf{e}_{\mathrm{o},k}(\boldsymbol{\theta}_{\mathrm{o}})|\mathcal{H}_{k-1}\right]$. For the estimation of the conditional variances we assume that spatial correlation decays as a function of distance for recruits and deaths at each time k given the past $\mathcal{H}_{k-1}$.

4.1 *Conditional expectation and variance for recruits*

By the first order Campbell formula used for B_k,

$$-\mathbb{E}\left[\frac{\mathrm{d}}{\mathrm{d}\boldsymbol{\theta}_{\mathrm{b}}^{\mathsf{T}}}\mathbf{e}_{\mathrm{b},k}(\boldsymbol{\theta}_{\mathrm{b}})|\mathcal{H}_{k-1}\right] = \int_{W}\frac{\nabla\zeta(x|\mathcal{H}_{k-1})\left[\nabla\zeta(x|\mathcal{H}_{k-1})\right]^{\mathsf{T}}\rho_0(x)}{\zeta(x|\mathcal{H}_{k-1})\left[\zeta(x|\mathcal{H}_{k-1}) + \rho_0(x)\right]}\mathrm{d}x,$$

which by the first order Campbell formula for $B_k \cup Y_k$ can be estimated unbiasedly by

$$\sum_{x\in B_k\cup Y_k}\frac{\nabla\zeta(x|\mathcal{H}_{k-1})\left[\nabla\zeta(x|\mathcal{H}_{k-1})\right]^{\mathsf{T}}\rho_0(x)}{\zeta(x|\mathcal{H}_{k-1})\left[\zeta(x|\mathcal{H}_{k-1}) + \rho_0(x)\right]^2} = \sum_{x\in B_k\cup Y_k}\frac{h_k(x)h_k(x)^{\mathsf{T}}}{\zeta_k(x|\mathcal{H}_{k-1})\rho_0(x)}$$

where $h_k(x) = \nabla\zeta_k(x|\mathcal{H}_{k-1})\rho_0(x)/[\zeta_k(x|\mathcal{H}_{k-1}) + \rho_0(x)]$.

By the first and second order Campbell formulas, the variance of (6) is

$$\mathbb{V}\mathrm{ar}\left[\mathbf{e}_{\mathrm{b},k}(\boldsymbol{\theta}_{\mathrm{b}})|\mathcal{H}_{k-1}\right] = \mathbf{S}_{\mathrm{b}}(\boldsymbol{\theta}_{\mathrm{b}}) + \int_{W^2}h_k(x)h_k(x')^{\mathsf{T}}\left[g_{B_k}(x, x') - 1\right]\mathrm{d}x\mathrm{d}x',$$

where g_{B_k} is the pair correlation function of B_k given $\mathcal{H}_{k-1}$. Following Coeurjolly and Guan (2014) we estimate the last term in the variance by $\sum_{\substack{x,x'\in B_k\cup Y_k: \\ x\neq x'}} k(x, x')\frac{h_k(x)h_k(x')^{\mathsf{T}}}{\zeta_k(x|\mathcal{H}_{k-1})\zeta_k(x'|\mathcal{H}_{k-1})}\phi(x)\phi(x')$,

where $k((u,m),(u',m')) = 1[\|u-u'\| \leqslant \omega]$ is a uniform kernel function depending on a truncation distance ω to be chosen by the user, and $\phi(x) = 1$ if $x \in B_k$ and $\phi(x) = -\rho_0(x)/\zeta_k(x|\mathcal{H}_{k-1})$ if $x \in Y_k$.

The underlying assumption of the estimator is that correlation vanishes for large spatial lags in the sense that $g_{B_k}(x,x') \approx 1$ when the spatial distance between x and x' is large. Hence, the kernel function eliminates pairs of distant points which are uncorrelated (meaning g_{B_K} close to one) and only add noise to the estimate. The unknown regression parameters appearing in $\zeta_k(\cdot|\mathcal{H}_{k-1})$ and $h_k(\cdot)$ are replaced by their composite likelihood estimates.

Crucially, we avoid specifying a model for the pair correlation function g_{B_k} which makes our variance estimate less prone to model misspecification. In contrast to non-parametric kernel estimates of g_{B_k} we also avoid assuming isotropy. If isotropy for g_{B_k} is preferred, shape-constrained non-parametric estimators (Hessellund et al., 2022; Xu et al., 2023) may be plugged in for g_{B_k} in $\mathbb{V}\mathrm{ar}\big[\mathbf{e}_{\mathrm{b},k}(\boldsymbol{\theta}_{\mathrm{b}})|\mathcal{H}_{k-1}\big]$.

4.2 *Conditional expectation and variance for death score*

When conditioning on $\mathcal{H}_{k-1}$, the index set $X_{k-1}^{(1)}$ for the sum in (7) becomes non-random. Therefore, by standard computations for expectations and variances of sums,

$$-\mathbb{E}\Big[\frac{\mathrm{d}}{\mathrm{d}\boldsymbol{\theta}_{\mathrm{d}}^{\mathsf{T}}}\mathbf{e}_{\mathrm{d},k}(\boldsymbol{\theta}_{\mathrm{d}})\Big|\mathcal{H}_{k-1}\Big] = \sum_{x \in X_{k-1}^{(1)} \cap W} \nabla\eta_k(x|\mathcal{H}_{k-1})\nabla\eta_k(x|\mathcal{H}_{k-1})^{\mathsf{T}}\frac{p_k(x|\mathcal{H}_{k-1})^2}{\exp_k(\eta_k(x|\mathcal{H}_{k-1}))} \quad \text{and}$$

$$\mathbb{V}\mathrm{ar}(\mathbf{e}_{\mathrm{d},k}(\boldsymbol{\theta}_{\mathrm{d}})|\mathcal{H}_{k-1}) = \sum_{x,x' \in X_{k-1}^{(1)} \cap W} \nabla\eta_k(x|\mathcal{H}_{k-1})\left(\nabla\eta_k(x'|\mathcal{H}_{k-1})\right)^{\mathsf{T}} \mathbb{C}\mathrm{ov}[I_k(x), I_k(x')|\mathcal{H}_{k-1}].$$

In the spirit of Conley (2010), we estimate the conditional variance by

$$\sum_{x,x' \in X_{k-1}^{(k)}} \nabla\eta_k(x|\mathcal{H}_{k-1})\nabla\eta_k(x'|\mathcal{H}_{k-1})^{\mathsf{T}} k(x,x')\left[I_k(x) - p_k(x|\mathcal{H}_{k-1})\right]\left[I_k(x') - p_k(x'|\mathcal{H}_{k-1})\right],$$

where as in the previous section, $k(\cdot,\cdot)$ is a uniform kernel used to avoid contributions from pairs of distant points in X_{k-1}, and where the unknown parameters in $p_k(\cdot|\mathcal{H}_{k-1})$ and $\nabla\eta_k(\cdot|\mathcal{H}_{k-1})$ are replaced by their composite likelihood estimates. Similar to the previous

section we avoid modeling of the conditional covariance $\mathbb{C}\text{ov}[I_k(x), I_k(x')|\mathcal{H}_{k-1}]$ and just need that the covariance vanishes when the spatial distance between x and x' increases.

5. Asymptotic distribution of parameter estimates

The key elements in establishing asymptotic normality are a first order Taylor expansion of the estimating function and asymptotic normality of the estimating function. In addition various regularity conditions are needed (e.g. Sørensen, 1999). Given asymptotic normality of the estimating function, the further conditions and derivations needed are quite standard (see for example Waagepetersen and Guan, 2009, for a case with all details provided). In the following we give a sketch of the asymptotic approach. We do not present the standard technical assumptions and just focus on the essential peculiarities for our setting.

In the space-time context, several asymptotic regimes are possible. One option is increasing K, i.e. accumulating information over time. Another is increasing spatial domain W. A combination of these is also possible (Jalilian et al., 2024). In our setting, K is moderate (seven for the specific data example) and increasing domain asymptotics therefore seems more relevant than increasing K. Consider a sequence of observation windows W_n, $n = 1, 2, \ldots$ and add the subindex n to $\mathbf{e}_\text{b}(\boldsymbol{\theta}_\text{b})$ and $\mathbf{e}_\text{d}(\boldsymbol{\theta}_\text{d})$ when the observation window W_n is considered. We further divide $\mathbb{R}^2$ into unit squares $C(z) = [z_1 - 1/2, z_1 + 1/2[\times[z_2 - 1/2, z_2 + 1/2[$ for $z = (z_1, z_2) \in \mathbb{Z}^2$. Then $\mathbf{e}_{\text{o},n}(\boldsymbol{\theta}) = \sum_{z \in \mathbb{Z}^2} \sum_{k=1}^{K} E_{\text{o},k,n}(z)$, $\text{o} = \text{b}, \text{d}$, where

$$E_{\text{b},k,n}(z) = \sum_{x \in (B_k \cup Y_k) \cap [C(z) \times \mathcal{M}] \cap W_n} \left[\frac{\nabla \zeta_k(x|\mathcal{H}_{k-1})}{\zeta_k(x|\mathcal{H}_{k-1})} 1[x \in B_k] - \frac{\nabla \zeta_k(x|\mathcal{H}_{k-1})}{\zeta_k(x|\mathcal{H}_{k-1}) + \rho_0(x)} \right]$$

and

$$E_{\text{d},k,n}(z) = \sum_{x \in X_{k-1}^{(1)} \cap [C(z) \times \mathcal{M}] \cap W_n} \nabla \eta_k(x|\mathcal{H}_{k-1}) \left(I_k(x) - p_k(x|\mathcal{H}_{k-1}) \right)$$

are the contributions to $\mathbf{e}_{\text{b},n}(\boldsymbol{\theta}_\text{b})$ and $\mathbf{e}_{\text{d},n}(\boldsymbol{\theta}_\text{d})$ arising from the intersections of time k recruits and deaths with $[C(z) \times \mathcal{M}] \cap W_n$.

For fixed K, $\mathbf{e}_{\text{o},n}(\boldsymbol{\theta}_\text{o})$ can be viewed as a sum of purely spatially indexed variables $E_{\text{o},n}(z) =$

$\sum_{k=1}^{K} E_{\mathrm{o},k,n}(z)$. It is, however, difficult to control the spatial dependence structure of these variables since spatial dependence may propagate over time. Instead, for $n \to \infty$, we invoke case (i) of the central limit theorem established in Jalilian et al. (2024) for the sequence of conditionally centered random fields $E_{\mathrm{o},k,n} = \{E_{\mathrm{o},k,n}(z)\}_{z \in \mathbb{Z}^2}$, $k = 1, \ldots, K$.

Regarding spatial dependence it then suffices to assume for each $k = 1, \ldots, K$, weak spatial dependence (α-mixing) of B_k and D_k conditional on $\mathcal{H}_{k-1}$. For instance, such conditional weak dependence is trivially satisfied for B_k if B_k is a Poisson process conditional on $\mathcal{H}_{k-1}$ and could also be established if for example B_k is a Poisson-cluster process conditional on $\mathcal{H}_{k-1}$. Note that even in the simple case of B_k being conditionally a Poisson process and deaths being conditionally independent, the aggregated processes $E_{\mathrm{o},n} = \{E_{\mathrm{o},n}(z)\}_{z \in \mathbb{Z}^2}$, $\mathrm{o} = \mathrm{b}, \mathrm{d}$ (with $E_{\mathrm{o},n}(z)$ defined above), have non-trivial spatial dependence structures. We refer to Jalilian et al. (2024) for further technical details and discussion of assumptions.

According to the central limit theorem, $\mathbf{V}_{\mathrm{o},n}^{-1/2}(\boldsymbol{\theta}_{\mathrm{o}}^*)\mathbf{e}_n(\boldsymbol{\theta}_{\mathrm{o}}^*)$ converges to a standard Gaussian vector where $\boldsymbol{\theta}_{\mathrm{o}}^*$ denotes the true value of $\boldsymbol{\theta}_{\mathrm{o}}$, $\mathrm{o} = \mathrm{b}, \mathrm{d}$. Using standard arguments, the convergence in distribution of $\mathbf{V}_{\mathrm{o},n}(\boldsymbol{\theta}_{\mathrm{o}}^*)^{-1/2}\mathbf{S}_{\mathrm{o},n}(\boldsymbol{\theta}_{\mathrm{o}}^*)(\widehat{\boldsymbol{\theta}}_{\mathrm{o},n} - \boldsymbol{\theta}_{\mathrm{o}}^*)$ to a standard Gaussian vector is obtained. Hence, $(\widehat{\boldsymbol{\theta}}_{\mathrm{o},n} - \boldsymbol{\theta}_{\mathrm{o}}^*)$ is approximately Gaussian distributed with covariance matrix $\mathbf{S}_{\mathrm{o},n}(\boldsymbol{\theta}_{\mathrm{o}}^*)^{-1}\mathbf{V}_{\mathrm{o},n}(\boldsymbol{\theta}_{\mathrm{o}}^*)\mathbf{S}_{\mathrm{o},n}(\boldsymbol{\theta}_{\mathrm{o}}^*)^{-1}$. Plugging in our estimates for $\mathbf{S}_{\mathrm{o},n}(\boldsymbol{\theta}_{\mathrm{o}}^*)$ and $\mathbf{V}_{\mathrm{o},n}(\boldsymbol{\theta}_{\mathrm{o}}^*)$, we obtain standard errors and confidence intervals for the various parameters. The sensitivity $\mathbf{S}_{\mathrm{o},n}(\boldsymbol{\theta}_{\mathrm{o}}^*)$ and the variance $\mathbf{V}_{\mathrm{o},n}(\boldsymbol{\theta}_{\mathrm{o}}^*)$ are roughly proportional to the spatial window size $|\widetilde{W}_n|$ so the parameter estimation variance is asymptotically inversely proportional to $|\widetilde{W}_n|$.

6. Simulation study

We conduct a simulation study to assess the performance of our methodology and to benchmark it (Section 6.1) against existing methodology for space-time point and binary processes. To emulate the expanding window asymptotics we consider two observation windows

$\widetilde{W}_1 = [0, 500] \times [0, 250]$ and $\widetilde{W}_2 = [0, 1000] \times [0, 500]$. We simulate two types of tree species, $p = 2$, and estimate the parameters for the first species. For sake of simplicity we disregard effects of marks which are just fixed at an arbitrary value 1. The covariate vector for recruit intensities and death probabilities (Section 2.1) is specified as $\mathbf{Z}_{k-1}(u) = (Z^{(1)}(u), Z^{(2)}(u))^\intercal$, $u \in \widetilde{W}_2$, where $Z^{(1)}(u)$ and $Z^{(2)}(u)$ are zero mean Gaussian random fields (Figure 1 in the supplementary material).

For the influence of the existing trees on recruits, we deviate a bit from the description in Section 2.2 and for both species let $\mathbf{c}_{k-1}(x)$ have components $c^l_{k-1}(x) = \exp\{-[d(x, X^{(l)}_{k-1})]/\psi)^2\}$, $l = 1, 2$, as in (3) with $\psi = 6$. For the death probabilities we for both species let $\mathbf{d}_{k-1}(x)$ have components $d^l_{k-1}(x) = \sum_{(u',m') \in X^{(l)}_{k-1} \backslash \{(u,m)\}} \frac{m'}{m} \exp[-(\|u - u'\|/\kappa)^2]$, $l = 1, 2$, as in (4) with $m = m' = 1$ and $\kappa = 10$ so that the practical range of influence of an existing tree is less than 20m.

We simulate $K = 10$ generations of recruits and deaths. For each time step $1 \leqslant k \leqslant K$, the recruits for both species are simulated from a log Gaussian Cox process (Møller et al., 1998) with intensity function given by (1) with intercept $\beta_{\mathrm{b0},k} = \beta_{\mathrm{b0}} = -6.32$, $(\beta_{\mathrm{b1}}, \beta_{\mathrm{b2}}) = (0, 0.1)$, and log-pairwise interaction function $\log g_{B_k}$ given by the Matérn covariance function (supplementary Section 1) with variance, smoothness and correlation scale parameters $\sigma^2 = 1, \nu = 1.75$ and $\xi = 4$. For the first species $(\gamma_{\mathrm{b1}}, \gamma_{\mathrm{b2}}) = (0.1, -2)$, while for the second species $(\gamma_{\mathrm{b1}}, \gamma_{\mathrm{b2}}) = (-2, 0.1)$. The initial point patterns $X^{(1)}_0$ and $X^{(2)}_0$ are generated from the same log Gaussian Cox process but with $\gamma_{\mathrm{b1}} = \gamma_{\mathrm{b2}} = 0$. For the death indicators we for time $1 \leqslant k \leqslant K$ use correlated logistic models. For both species we let $\beta_{\mathrm{d0},k} = \beta_{\mathrm{d0}} = -0.25$, $(\beta_{\mathrm{d1}}, \beta_{\mathrm{d2}}) = (0.25, 0)$, $(\gamma_{\mathrm{d1}}, \gamma_{\mathrm{d2}}) = (-0.25, 0.25)$, specify $\eta_k(\cdot|\mathcal{H}_{k-1})$ by (2), and proceed as follows for $l = 1, 2$.

(1) First, we simulate a zero mean Gaussian random field U_k, with a Matérn covariance function with parameters $\sigma^2 = 1, \nu = 0.5$ and $\xi = 7$.

(2) Then, for $x = (u, m) \in X_{k-1}^{(l)}$, we compute $p_k(u) = \Phi^{-1}[U_k(u)]$ with Φ the standard normal cumulative distribution function and the logistic variable $\tau(u) = \log \left[\frac{p_k(u)}{1 - p_k(u)} \right]$.

(3) Finally, $I_k(x) = 1[\tau(u) \leqslant \eta_k(u | \mathcal{H}_{k-1})]$, $x = (u, m) \in X_{k-1}^{(l)}$.

We generate 1000 bivariate space-time point patterns and compute parameter estimates and estimates of parameter estimate covariance matrices for each simulated space-time pattern. Figure 2 in the supplementary material shows the evolution of the numbers of recruits and deaths for the windows $\widetilde{W}_1$ and $\widetilde{W}_2$.

Figures 3 and 4 in the supplementary material show kernel density estimates of the 1000 simulated regression parameter estimates. For both window sizes the parameter distributions are close to normal and with bias close to zero. Moreover, the kernel density estimates and the numbers in Table 1 show that according to asymptotic theory, parameter estimation variance decreases at a rate inversely proportional to window size (variances four times larger for $\widetilde{W}_1$ than for $\widetilde{W}_2$).

[Table 1 about here.]

Figure 1 shows boxplots of estimates (Section 4) of the variances of recruit parameter estimates for different choices of truncation distances equally spaced between 5 and 155m. For small truncation distances, the estimates are strongly biased downwards and the bias decreases for larger truncation distances. The medians of the variance estimates are stable for truncation distances greater than 30 and the variance increases as the truncation distance increases. The variances of the variance estimates are much reduced when increasing the window from $\widetilde{W}_1$ to $\widetilde{W}_2$ (note the different limits on the y-axes) while the bias relative to the variance seems a bit larger for $\widetilde{W}_2$.

[Figure 1 about here.]

The plots for the death parameters in Figure 5 in the supplementary material are similar to Figure 1 and with similar comments.

The plots in Figure 2 show coverage probabilities over the 1000 simulations of 95% confidence intervals based on the asymptotic normal distribution of the parameter estimates with variances estimated following Section 4. For the recruits parameters, the coverage probabilities are quite close to the nominal 95% over a wide range of truncation distances. For the death parameters, the coverage probabilities are a bit less satisfactory in case of $\widetilde{W}_1$ but quite close to the nominal 95% in case of the bigger window $\widetilde{W}_2$.

[Figure 2 about here.]

Overall, inference based on the asymptotic normal distribution of parameter estimates and the proposed estimates of parameter estimate variances seems reliable at least when the observation window is sufficiently large. Coverage probabilities of confidence intervals are fairly stable across a wide range of truncation distances indicating an appealing robustness to the choice of truncation distance for variance estimation.

6.1 *Comparison with INLA*

The INLA (integrated nested Laplace approximation) package (Rue et al., 2009) has gained huge popularity as a versatile and computationally efficient tool for analysing space-time data. INLA implements Bayesian inference for latent Gaussian random field models and thus requires a full specification of the data generating mechanism. For the recruits we use INLA to fit a log Gaussian Cox process with the same space-time correlation structure and covariates that were used to generate the simulated data sets. More precisely, INLA fits a Poisson log normal model to counts of points within cells of a partition of the observation window. Specifically, we consider counts within disjoint 10×10m squares corresponding to a 50×100 grid on the large window $\widetilde{W}_2$. For the deaths we use INLA to implement a logistic regression with space-time correlated random effects. For the prior distributions we use the

default non-informative priors specified by INLA. More details regarding the INLA method are given in supplementary Section 1.

For the considered space-time setting, INLA is quite time consuming and we therefore only consider the window $\widetilde{W}_2$ and only 100 simulated data sets. The mean computing time (Intel E5-2680 v4, 152GB RAM) for a simulated data set is 21.6 minutes (standard deviation 5.38) for recruits and 19.7 minutes (standard deviation 4.3) for deaths. Our method is much faster with mean computing time 13.7 seconds (standard deviation 3.8) for recruits and 4.63 seconds (standard deviation 2.28) for deaths (including all 11 considered truncation distances). Supplementary Figure 6 compares the composite likelihood estimates with the INLA posterior mean estimates. For 6 out of 10 parameters there is close agreement between the two types of estimates. However, the INLA posterior means for the remaining 4 parameters show considerable bias. As detailed in the supplementary material, this seems to be due to discretization error when a covariate is coarsened to the 50×100 grid. Accordingly, frequentist coverages of nominal 95% posterior intervals differ markedly from 95% for several parameters: the estimated coverage probabilities are 0.00, 0.94, 0.95, 0.90, 0.01 for β_{b0}, β_{b1}, β_{b2}, γ_{b1}, γ_{b2} and 0.77, 0.87, 0.89, 0.48, 0.94 for β_{d0}, β_{d1}, β_{d2}, γ_{d1}, γ_{d2}.

7. The Barro Colorado Island data

[Figure 3 about here.]

We consider tree census data from the 50 ha, $\widetilde{W} = [0, 1000\mathrm{m}] \times [0, 500\mathrm{m}]$, Barro Colorado Island (BCI) study plot (Condit et al., 2019). The first census was conducted in 1983 followed by censuses in 1985, 1990, 1995, 2000, 2005, 2010, and 2015. The censuses include all trees with diameter of breast height $m \geqslant 10\mathrm{mm}$. Figure 3 shows the locations of *Capparis frondosa* trees in the first census and recruits and deaths for the remaining seven censuses. We here ignore that the time-interval between the first and the second census is smaller than for the

remaining censuses. To some extent this is accounted for by the census dependent intercepts. Table 1 in the supplementary material summarizes the numbers of recruits and deaths in each census. The population of *Capparis* trees seems to be declining with a decreasing trend regarding number of recruits and increasing trend regarding number of deaths. A detailed discussion of existing BCI literature on recruitment and mortality is given in the supplementary Section 4.

We employ the log linear recruit intensity function (1) with covariates copper (Cu), potassium (K), phosphorus (P), pH, mineralized nitrogen (Nmin), elevation (dem), slope gradient (grad), convergence index (convi), multi-resolution index of valley bottom flatness (mrvbf), incoming mean annual solar radiation (solar), and topographic wetness index (twi) available on a $5 \times 5 \, \mathrm{m}^2$ grid. For the influence of existing trees we only distinguish between Capparis trees ($l = 1$) and other trees ($l = 2$). For the influence of Capparis on recruits we use (3) with $\psi_1 = 0.25$ and for the influence of Capparis on deaths we use (4) with $\kappa_1 = 5$. For the influence of other trees we compute influence functions of the form (4) for all abundant species other than Capparis with more than 500 trees and with $\kappa_2 = 5$. These influence functions are averaged to get an influence function for other trees that is used both for recruits and deaths.

The left plot in Figure 4 shows non-parametric estimates of pair correlation functions for each point pattern of recruits. Despite the variability between the estimates, all estimates seem to stabilize around 1 after a distance of 55m which we use as the truncation distance for variance estimation. The middle plot shows variograms for the death events. For the deaths we also use truncation distance 55m. There does not appear to be strong spatial correlation between death events and indeed our estimated standard deviations for the death regression parameter estimates are only slightly larger than those (not shown) obtained assuming conditionally independent death events. The right plot shows the intercepts for recruits

that decrease over time and the intercepts for deaths that increase over time. The composite likelihood estimates of the remaining regression parameters for respectively recruits and deaths are given in Table 2 together with p-values and INLA results (see below).

[Figure 4 about here.]

[Table 2 about here.]

We do not attempt a formal investigation of the significance of the various covariates. However, based on the p-values, there is some evidence that recruit intensity is negatively associated with the covariates convi, grad, and solar and positively associated with Nmin and, not surprisingly, presence of existing Capparis trees. Probability of death appears to be positively associated with high level of Cu, pH and solar. However, no dependence on existing trees is detected.

We also analyzed the BCI data set using INLA as described in Section 6.1 with a discretization into 5×5m cells for the recruits. The INLA posterior means for the regression parameters are quite similar to our estimates and this also holds for conclusions regarding significance based on whether zero is contained in the 95% credibility intervals or not (except for Capparis influence on deaths). However, our method is much faster with computing times in seconds 53.2 seconds (recruits) and 12.6 seconds (deaths) compared to 14.6 minutes and 34.9 minutes for INLA.

8. Discussion

Our methodology is essentially free of assumptions regarding the second-order properties of the space-time point process but a risk of misspecification for the intensity and death probability regression models remains. This does not necessarily render regression parameter estimates meaningless. As in Choiruddin et al. (2021) one might define 'least wrong' regression models that minimize a composite likelihood Kullback-Leibler distance to the true

intensity and death probability models. The composite likelihood parameter estimates then estimate the corresponding 'least wrong' regression parameter values.

We have focused on estimation of regression parameters. However, scale parameters in the models for influence of existing trees need to be determined too. If suitable values of these parameters can not be identified from biological insight, one might include these parameters in the composite likelihood estimation although computations become more cumbersome.

We have not provided a theoretically well founded method for truncation distance selection for the variance matrix estimators. Our simulation study, however, indicates robustness to the choice of truncation distance. As in the data example, plots of estimates of pair correlation functions and variograms may give some idea of suitable truncation distances.

Supplementary materials

Web appendices, tables, figures, and code referenced in Sections 6 and 7 are available with this paper at the Biometrics website on Oxford Academic. The code for the simulation study and data example is also available at `https://github.com/FcoCuevas87/CL_ST-PointProcesses`.

Data availability

The Barro Colorado Island data that support the findings in this paper are available at `https://datadryad.org/stash/dataset/doi:10.15146/5xcp-0d46`.

REFERENCES

Baddeley, A., Coeurjolly, J.-F., Rubak, E., and Waagepetersen, R. (2014). Logistic regression for spatial Gibbs point processes. *Biometrika* **101,** 377–392.

Britton, T., Richards, S. A., and Hovenden, M. J. (2023). Quantifying neighbour effects on tree growth: are common 'competition' indices biased ? *Journal of Ecology* **111,** 1270–1280.

Brix, A. and Møller, J. (2001). Space-time multi type log Gaussian Cox processes with a view to modelling weeds. *Scandinavian Journal of Statistics* **28,** 471–488.

Bullock, J., González, L. M., Tamme, R., Götzenberger, L., White, S., Pärtel, M., and Hooftman, D. (2019). A synthesis of empirical plant dispersal kernels. *Journal of Ecology* **105,** 6–19.

Burkhart, H. E. and Tomé, M. (2012). *Indices of Individual-Tree Competition*, pages 201–232. Springer Netherlands, Dordrecht.

Choiruddin, A., Coeurjolly, J.-F., and Waagepetersen, R. (2021). Information criteria for inhomogeneous spatial point processes. *Australian & New Zealand Journal of Statistics* **63,** 119–143.

Coeurjolly, J.-F. and Guan, Y. (2014). Covariance of empirical functionals for inhomogeneous spatial point processes when the intensity has a parametric form. *Journal of Statistical Planning and Inference* **155,** 79–92.

Comita, L. S. and Hubbell, S. P. (2009). Local neighborhood and species' shade tolerance influence survival in a diverse seedling bank. *Ecology* **90,** 328–334.

Condit, R., Pérez, R., Aguilar, S., Lao, S., Foster, R., and Hubbell, S. (2019). Complete data from the Barro Colorado 50-ha plot: 423617 trees, 35 years. *URL https://doi. org/10.15146/5xcp-0d46* .

Conley, T. G. (2010). *Spatial Econometrics*, pages 303–313. Palgrave Macmillan UK, London.

Diggle, P. J. (2013). *Statistical analysis of spatial and spatio-temporal point patterns.* CRC press.

Getzin, S., Wiegand, T., and Hubbell, S. P. (2014). Stochastically driven adult–recruit associations of tree species on Barro Colorado Island. *Proceedings of the Royal Society B: Biological Sciences* **281,** 20140922.

González, J. A., Rodríguez-Cortés, F. J., Cronie, O., and Mateu, J. (2016). Spatio-temporal point process statistics: a review. *Spatial Statistics* **18,** 505–544.

Häbel, H., Myllymäki, M., and Pommerening, A. (2019). New insights on the behaviour of alternative types of individual-based tree models for natural forests. *Ecological modelling* **406,** 23–32.

Hessellund, K. B., Xu, G., Guan, Y., and Waagepetersen, R. (2022). Semiparametric multinomial logistic regression for multivariate point pattern data. *Journal of the American Statistical Association* **117,** 1500–1515.

Hiura, T., Go, S., and Iijima, H. (2019). Long-term forest dynamics in response to climate change in northern mixed forests in Japan: A 38-year individual-based approach. *Forest Ecology and Management* **449,** 117469.

Hubbell, S. P. (2001). *The Unified Neutral Theory of Biodiversity and Biogeography (MPB-32).* Princeton University Press.

Hubbell, S. P., Ahumada, J. A., Condit, R., and Foster, R. B. (2001). Local neighborhood effects on long-term survival of individual trees in a neotropical forest. *Ecological Research* **16,** 859–875.

Jalilian, A., Poinas, A., Xu, G., and Waagepetersen, R. (2024). A central limit theorem for a sequence of conditionally centered and α-mixing random fields. *Bernouilli* To appear.

Johnson, D. J., Condit, R., Hubbell, S. P., and Comita, L. S. (2017). Abiotic niche partitioning and negative density dependence drive tree seedling survival in a tropical

forest. *Proceedings of the Royal Society B: Biological Sciences* **284,** 20172210.

Kohyama, T. S., Kohyama, T. I., and Sheil, D. (2018). Definition and estimation of vital rates from repeated censuses: Choices, comparisons and bias corrections focusing on trees. *Methods in Ecology and Evolution* **9,** 809–821.

May, F., Huth, A., and Wiegand, T. (2015). Moving beyond abundance distributions: neutral theory and spatial patterns in a tropical forest. *Proceedings of the Royal Society B: Biological Sciences* **282,** 20141657.

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. (2017). Some recent developments in statistics for spatial point patterns. *Annual Review of Statistics and Its Application* **4,** 317–342.

Møller, J. and Waagepetersen, R. P. (2003). *Statistical inference and simulation for spatial point processes.* CRC press.

Nathan, R., Klein, E., Robledo-Arnuncio, J. J., and Revilla, E. (2012). Dispersal kernels: review. In *Dispersal Ecology and Evolution.* Oxford University Press.

Rathbun, S. L. and Cressie, N. (1994). A space-time survival point process for a longleaf pine forest in southern Georgia. *Journal of the American Statistical Association* **89,** 1164–1174.

Rue, H., Martino, S., and Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society Series B: Statistical Methodology* **71,** 319–392.

Rüger, N., Huth, A., Hubbell, S. P., and Condit, R. (2009). Response of recruitment to light availability across a tropical lowland rain forest community. *Journal of Ecology* **97,** 1360–1368.

Shen, Y., Santiago, L. S., Ma, L., Lin, G.-J., Lian, J.-Y., Cao, H.-L., and Ye, W.-H. (2013).

Forest dynamics of a subtropical monsoon forest in Dinghushan, China: recruitment, mortality and the pace of community change. *Journal of Tropical Ecology* **29,** 131–145.

Sørensen, M. (1999). On asymptotics of estimating functions. *Brazilian Journal of Probability and Statistics* **13,** 111–136.

Waagepetersen, R. (2008). Estimating functions for inhomogeneous spatial point processes with incomplete covariate data. *Biometrika* **95,** 351–363.

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.

Wiegand, T., Martínez, I., and Huth, A. (2009). Recruitment in tropical tree species: revealing complex spatial patterns. *The American Naturalist* **174,** E106–E140.

Wiegand, T. and Moloney, K. A. (2013). *Handbook of spatial point-pattern analysis in ecology.* CRC press.

Wolf, A. (2005). Fifty year record of change in tree spatial patterns within a mixed deciduous forest. *Forest Ecology and management* **215,** 212–223.

Xu, G., Liang, C., Waagepetersen, R., and Guan, Y. (2023). Semiparametric goodness-of-fit test for clustered point processes with a shape-constrained pair correlation function. *Journal of the American Statistical Association* **118,** 2072–2087.

Zhu, Y., Queenborough, S. A., Condit, R., Hubbell, S., Ma, K., and Comita, L. S. (2018). Density-dependent survival varies with species life-history strategy in a tropical forest. *Ecology letters* **21,** 506–515.

Zuleta, D., Arellano, G., Muller-Landau, H. C., McMahon, S. M., Aguilar, S., Bunyavejchewin, S., Cárdenas, D., Chang-Yang, C.-H., Duque, A., Mitre, D., et al. (2022). Individual tree damage dominates mortality risk factors across six tropical forests. *New Phytologist* **233,** 705–721.
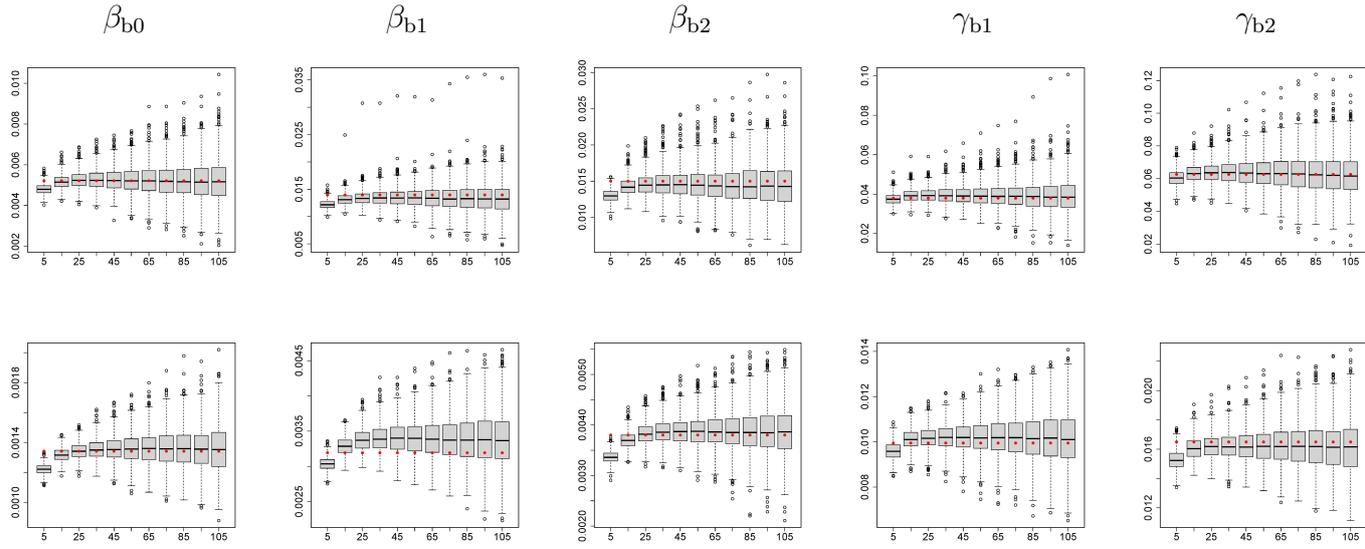
Figure 1. Boxplots of the estimated variances for recruit parameter estimates for $\widetilde{W}_1 = [0, 250] \times [0, 500]$ (upper row) and $\widetilde{W}_2 = [0, 500] \times [0, 1000]$ (lower row) for different truncation distances. The red dots show the empirical variance of the simulated parameter estimates. The parameter β_{b0} is the intercept for recruits, β_{b1} and β_{b2} are regression parameters for spatial covariates, and γ_{b1} and γ_{b2} are regression parameters for the dependence on previous points of same type and different type, respectively.
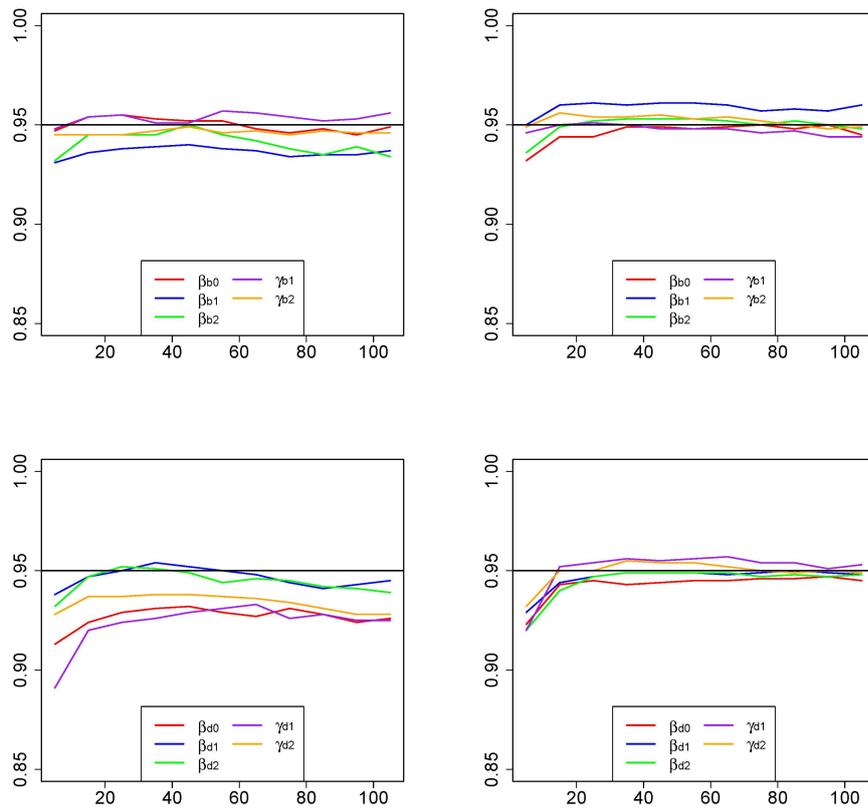
Figure 2. Coverage probabilities for the recruits (top) and the deaths (bottom) regression parameter confidence intervals obtained with varying truncation distances and windows $\widetilde{W}_1 = [0, 250] \times [0, 500]$ (left) and $\widetilde{W}_2 = [0, 500] \times [0, 1000]$ (right). The parameter β_{b0} is the intercept for recruits, β_{b1} and β_{b2} are regression parameters for spatial covariates, and γ_{b1} and γ_{b2} are regression parameters for the dependence on previous points of same type and different type, respectively. The explanation is analogous for the death regression parameters $\beta_{d0}, \beta_{d1}, \beta_{d2}, \gamma_{d1}, \gamma_{d2}$.
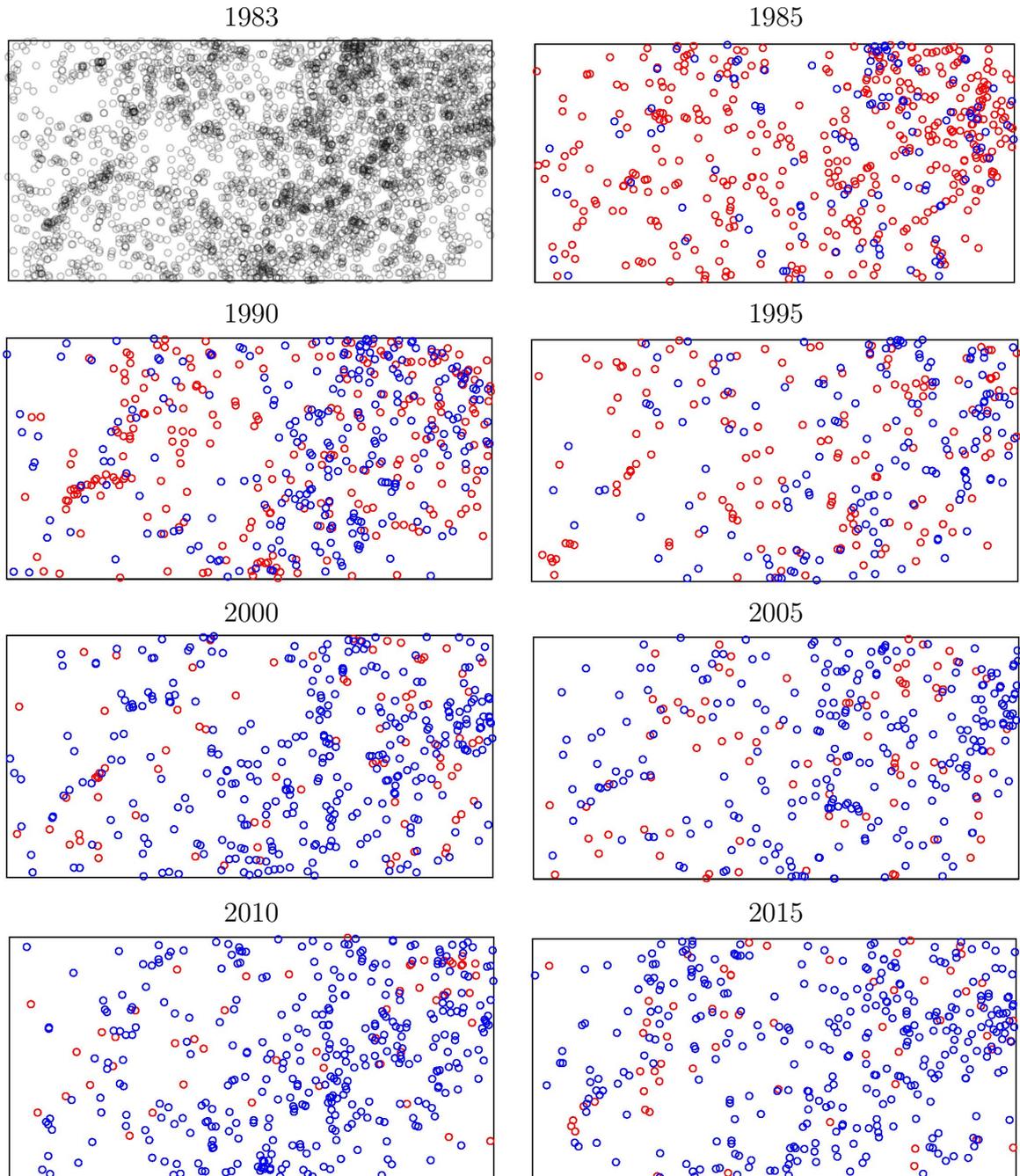
Figure 3. Spatial point patterns of locations of *Capparis frondosa* trees in the eight censuses of the 50 ha plot in Barro Colorado Island (BCI), Panama. Top left plot: trees in first census. Remaining plots: recruits (red) and deaths (blue) relative to previous census.
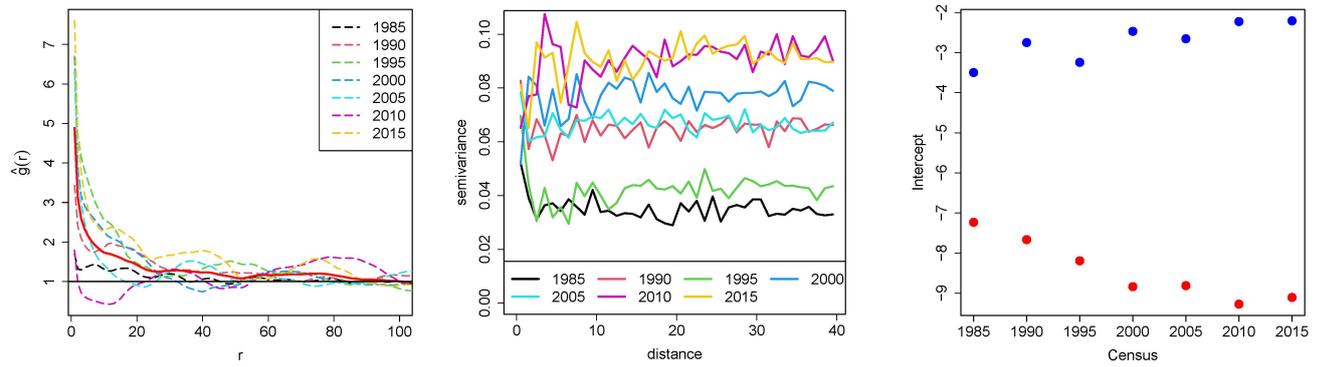
Figure 4. Left: estimated pair correlation functions for each census. Red solid curve shows the average of the estimates. Middle: variograms for death indicators for each census. Right: time dependent intercepts for recruits (blue) and deaths (red).

Table 1

Second and fifth rows: variances of the parameter estimates for the recruits and deaths for $\widetilde{W}_2 = [0, 500] \times [0, 1000]$. Third and sixth rows: ratios of variances for $\widetilde{W}_1 = [0, 250] \times [0, 500]$ and $\widetilde{W}_2$. The parameter β_{b0} is the intercept for recruits, β_{b1} and β_{b2} are regression parameters for spatial covariates, and γ_{b1} and γ_{b2} are regression parameters for the dependence on previous points of same type and different type, respectively. The explanation is analogous for the death regression parameters $\beta_{d0}, \beta_{d1}, \beta_{d2}, \gamma_{d1}, \gamma_{d2}$.

	β_{b0}	β_{b1}	β_{b2}	γ_{b1}	γ_{b1}
Var. $\widetilde{W}_2$	1.3e-03	3.2e-03	3.8e-03	9.9e-03	1.7e-02
Var. $\widetilde{W}_1$/ Var. $\widetilde{W}_2$	3.89	4.37	3.95	3.81	3.79
	β_{d0}	β_{d1}	β_{d2}	γ_{d1}	γ_{d2}
Var. $\widetilde{W}_2$	7.6e-02	7.6e-02	8.4e-02	2.6e-02	2.6e-02
Var. $\widetilde{W}_1$/ Var. $\widetilde{W}_2$	3.95	3.85	3.55	4.05	4.25

Table 2
Results for recruits (left) and deaths (right). Columns 'estm.', 'post. mean', 'p', and '0 inside' contain composite likelihood parameter estimates, INLA posterior means, p-values based on asymptotic normality, and indication of whether 0 is inside INLA 95% posterior credibility interval. Rows 'Infl. others' and 'Infl. Capparis' contain the results for the regression parameters measuring influence of other trees respectively Capparis trees on either Capparis recruits or Capparis deaths.

| | Recruits | | | | Deaths | | | |
	estm.	post. mean	p	0 inside	estm.	post. mean	p	0 inside
convi	-0.01	-0.01	0.00	no	-3e-3	0.00	0.09	yes
Cu	0.03	0.05	0.33	yes	0.10	0.11	0.00	no
dem	0.02	0.02	0.06	yes	4e-3	0.01	0.54	yes
grad	-2.79	-3.25	0.03	no	0.76	0.63	0.32	yes
K	0.00	0.00	0.85	yes	-2e-3	0.00	0.17	yes
mrvbf	-0.07	-0.03	0.24	yes	0.02	0.00	0.64	yes
Nmin	0.03	0.02	0.00	no	1e-3	0.00	0.88	yes
P	-2e-3	0.01	0.96	yes	-0.03	-0.03	0.32	yes
pH	-0.02	-0.01	0.90	yes	0.55	0.60	0.00	no
solar	-5e-3	0.00	0.00	no	3e-3	0.00	0.01	no
twi	0.01	-0.03	0.84	yes	0.03	0.03	0.39	yes
Infl. others	1e-3	0.01	0.93	yes	-0.01	-0.00	0.21	yes
Infl. Capparis	0.64	0.29	0.00	no	2e-4	0.32	0.65	no