JYU DISSERTATIONS 492

Mikko Kuronen

Analysis of Spatial Point Patterns with Anomalies, Covariates and Intractable Likelihoods

UNIVERSITY OF JYVÄSKYLÄ
FACULTY OF MATHEMATICS
AND SCIENCE

JYU DISSERTATIONS 492

Mikko Kuronen

Analysis of Spatial Point Patterns with Anomalies, Covariates and Intractable Likelihoods

Esitetään Jyväskylän yliopiston matemaattis-luonnontieteellisen tiedekunnan suostumuksella
julkisesti tarkastettavaksi yliopiston vanhassa juhlasalissa S212
maaliskuun 8. päivänä 2022 kello 12.

Academic dissertation to be publicly discussed, by permission of
the Faculty of Mathematics and Science of the University of Jyväskylä,
in building Seminarium, auditorium S212 on March 8, 2022 at 12 o'clock noon.

JYVÄSKYLÄN YLIOPISTO
UNIVERSITY OF JYVÄSKYLÄ

JYVÄSKYLÄ 2022

Editors
Matti Vihola
Department of Mathematics and Statistics, University of Jyväskylä
Ville Korkiakangas
Open Science Centre, University of Jyväskylä

ISBN 978-951-39-9020-6 (PDF)
URN:ISBN:978-951-39-9020-6
ISSN 2489-9003

Permanent link to this publication: http://urn.fi/URN:ISBN:978-951-39-9020-6

Abstract

Spatial point patterns arise in a number of applications from different disciplines. They represent locations of objects or events of interest. Such data is analysed and modelled using point process statistics. This work develops new statistical models and methods for challenges encountered in a few specific applications in forestry and medicine. We consider methods for the analysis of datasets that include artefacts or missing data, introduce new point process models, and suggest tests having graphical interpretation. In one of the applications, we develop models for sweat gland activation data, which is important in early screening of diabetes. To this end, we suggest methods to handle erroneously detected points in the data produced by image analysis. We also consider modelling how the locations of tree seedlings are affected by large trees. Here we propose a Bayesian inference method for handling nonlinear covariates in a log Gaussian Cox process. Furthermore, we present an estimator for forest characteristics in data obtained by terrestrial laser scanning. The new estimator accounts for unobserved trees behind other trees. Finally, we suggest a test with a graphical interpretation for including particular covariates in a point process model.

Tiivistelmä

Spatiaalisia pistekuvioita esiintyy useiden tieteenalojen sovelluksissa. Ne edustavat kiinnostavien kohteiden tai tapahtumien sijainteja. Tällaisia pistekuvioaineistoja analysoidaan ja mallinnetaan pisteprosessitilastotieteen avulla. Tässä työssä kehitetään uusia tilastollisia malleja ja menetelmiä metsätalouden ja lääketieteen erityissovellusten haasteisiin. Työssä tutkitaan aineistoja, joissa on virheellisiä tai puuttuvia havaintoja, sekä esitellään uusia pisteprosessimalleja ja ehdotetaan testejä, joilla on graafinen tulkinta. Yhdessä sovelluksessa kehitetään malleja hikirauhasten aktivoitumiselle, jonka ymmärtäminen on tärkeää diabeteksen varhaisessa seulonnassa. Hikirauhaset havaitaan työssä kehitetyllä kuva-analyysimenetelmällä. Menetelmän tuottamat virheelliset pisteet otetaan huomioon hikirauhaskuvioiden analyysissä. Toisessa sovelluksessa mallinnetaan log-gaussisella Coxin prosessilla suurten puiden vaikutusta taimien sijaintiin ja käytetään Bayes-päättelyä epälineaaristen kovariaattien sisällyttämiseksi malliin. Kolmas osajulkaisu esittelee uuden estimaattorin puustotunnuksille maalaserkeilauksella saadusta aineistosta. Uusi estimaattori huomioi puut, jotka jäävät havaitsematta, koska ne ovat muiden puiden takana. Viimeisessä työn osassa esitellään uusia testejä selittävien muuttujien merkitsevyydelle pisteprosessimallissa.

Acknowledgements

I wish to thank my friend and primary supervisor Dr Mari Myllymäki for everything. I am also deeply grateful to my supervisor Prof. Matti Vihola for guidance and to Prof. Aila Särkkä for time.

I am grateful to Dr Helena Henttonen for introducing me to the art of forest mensuration, Dr Adam Loavenbruck for providing expertise on the sweat spot phenomenon and Dr Tomas Mrkvička for discussions on hypothesis testing.

I thank Prof. Juha Heikkinen for constructive comments on the thesis and Dr Kasper Kansanen, Prof. Lauri Mehtätalo and the rest of the spatio-temporal seminar group for useful discussions. I further thank the Forest inventory and planning group, and Natural Resources Institute Finland in general, for providing a working environment, both live and virtual coffee rooms and colleagues.

I am also grateful to Prof. Emer. Antti Penttinen for awakening my interest in spatial statistics and Dr Lasse Leskelä for all the advise during my first PhD attempt. I thank Prof. Emer. Elja Arjas and Prof. Jorge Mateu for their work as pre-examiners and Dr Thordis Thorarinsdottir for agreeing to serve as the opponent. Finally, I wish to thank Dr Sara Taskinen for always believing in me.

I thank Academy of Finland (Project numbers 304212, 306875 and 327211) for financial support and CSC – IT Center for Science, Finland, for computational resources.

Finally, I would like to thank my family and friends for their continuous support.

Helsinki, January 2022

Mikko Kuronen

List of included articles

This thesis consists of an introductory part and the publications listed below.

- I Kuronen, M., Henttonen, H. M., and Myllymäki, M. Correcting for nondetection in estimating forest characteristics from single-scan terrestrial laser measurements. *Canadian Journal of Forest Research*, 49(1):96–103, 2018. doi:10.1139/cjfr-2018-0072
- II Kuronen, M., Myllymäki, M., Loavenbruck, A., and Särkkä, A. Point process models for sweat gland activation observed with noise. *Statistics in Medicine*, 40:2055–2072, 2021. doi:10.1002/sim.8891
- III Kuronen, M., Särkkä, A., Vihola, M., and Myllymäki, M. Hierarchical log Gaussian Cox process for regeneration in uneven-aged forests. *Environmental and Ecological Statistics*, 2021. doi:10.1007/s10651-021-00514-3
- IV Myllymäki, M., Kuronen, M., and Mrkvička, T. Testing global and local dependence of point patterns on covariates in parametric models. *Spatial Statistics*, 42, 100436, 2021. doi:10.1016/j.spasta.2020.100436

Author's contributions

The author of this thesis was the main contributor and writer for the joint articles I, II and III and actively contributed to the research and writing of the joint article IV. The author was solely responsible for implementation of the methods, carrying out the experiments and the visualisations in all of the included articles.

Contents

Chapter 1

Introduction

Spatial point patterns are encountered in many fields of research. The points represent the locations of some objects or events of interest such as the locations of particles or pores in material science, trees or disasters in environmental science, disease occurrences in epidemiology, or fixations of eye movement in behavioural science. Further relevant information of the objects or events is often available in addition to the location, for example the species or size of a tree. In the point pattern context, such information is called a mark associated to a point.

A point process is a statistical model for a point pattern. Similar models and methods of point process statistics can be useful in many different applications. In general, the locations are points in a *d*-dimensional Euclidean space. In this work, we consider only point patterns in the plane, although some of the methods are applicable also in higher dimensional settings.

Typical workflow (Illian et al., 2008) in statistical analysis of point patterns starts with plotting the data and selected summary characteristics. This preliminary examination can reveal such features as regularity or clumpiness of point locations. It is common that a few appropriate models are fitted to the data, for example using software packages such as `spatstat` (Baddeley et al., 2015) or `INLA` (Rue et al., 2009). Again summary characteristics are computed from simulated data generated from the models to assess goodness-of-fit. It is not uncommon to point process statistics that challenges arise in different phases of the analysis, which may be due to limitations in the current methodology or in the methods implemented in software. In the applications discussed here such issues arise because of spurious points which are produced by image analysis, all relevant points have not been observed, the likelihood may be intractable, or the distribution of test statistics is unknown.

The contributions of this work consist of developing new models and methods in order to resolve a number of challenges encountered in specific applications. In modelling sweat spots (Article II) we dealt with points that were caused by errors in image analysis by utilising mixture models as an inferential method. We also tackled computationally demanding likelihood by an efficient implementation. Furthermore, we proposed a new model for sweat glands where the extra points were handled with inference based on approximate Bayesian computation with carefully selected low dimensional summary statistics. When modelling tree seedlings (Article III) the complications were caused by nonlinear point pattern based predictors making the model unsuitable for the available software. Several approximations were used to achieve efficient inference. We also proposed a test (Article IV) that can be used to select covariates in the model fitting stage of the point process workflow. Furthermore, we used the theory behind point processes to introduce a new estimator (Article I) for forest characteristics. The problem here was due to missing data: the trees that were not detected with terrestrial laser scanning because they were behind other trees.

Section 2 introduces background of point process statistics relevant for the developments in this work. In particular, Section 2.3 presents a general formulation for the sequential model likelihood that allows for efficient computation. Section 3 discusses the inference methods that were used in this work. Sections 4, 5, 6 and 7 summarise and elaborate on the main results of the Articles I, II, III and IV, respectively. Finally Section 8 finishes the thesis with discussion.

Chapter 2

Point process statistics

Two dimensional point processes are usually defined on the whole plane, but point patterns are observed within an observation window denoted by $W \subset \mathbb{R}^2$. A spatial point process $X = \{x_i\}$ may be augmented with additional information associated to the objects, called marks, to produce a marked point process $\{(x_i, m_i)\}$.

The intensity of a point process X is a function $\lambda : \mathbb{R}^2 \to \mathbb{R}^+$ that satisfies $\mathbf{E}X(B) = \int_B \lambda(x)dx$ for all (Borel[1]) $B \subset \mathbb{R}^2$, where $X(B)$ stands for the number of points of the process in the set B. The Poisson process with intensity λ is characterised by two properties. The number of points of the process in a set B has a Poisson distribution with mean $\int_B \lambda(s)ds$ and the numbers of points in disjoint sets are independent. A Poisson process with constant intensity is a model of complete spatial randomness.

The rest of this section is organised as follows. Section 2.1 introduces the summary statistics used in Articles II and III. Sections 2.2 and 2.3 introduce the point process models used in Article II. Section 2.4 introduces the log Gaussian Cox process (LGCP) used in Article III. Finally, Section 2.5 introduces point process residuals that were used in Article IV for defining new summary statistics.

---

[1] In this work we implicitly assume all sets and functions to be appropriately measurable.

2.1 Summaries

Features of point patterns are often explored using functional summary statistics. This section presents selected summary statistics that were the most important for this work. The point process is assumed to be stationary and isotropic, that is, invariant under translations and rotations (e.g. Illian et al., 2008, p. 177). Stationarity implies that the process has constant intensity.

A commonly used functional summary is the Ripley's K-function (Ripley, 1977) $K(r) = \frac{1}{\lambda}\mathbf{E}_o(X(B(o,r) \setminus \{o\}))$, where λ is the intensity of the process, $B(o,r)$ is the ball centred at the origin o with radius r and $\mathbf{E}_o$ is the expectation conditional on having a point of the process at o (in the Palm sense, e.g. Illian et al., 2008; Chiu et al., 2013). Due to the stationarity, it suffices to consider the origin only.

Instead of the K-function, its derivative, the pair correlation function

$$g(r) = \frac{K'(r)}{2\pi r} \tag{2.1}$$

can be used. The pair correlation has the following interpretation. Under complete spatial randomness $g(r) = 1$ for all $r \geq 0$. Values $g(r) < 1$ indicate repulsion at distance r and values $g(r) > 1$ indicate attraction at r.

The K-function can be modified to describe association between two point processes X_1 and X_2 by defining $K_{12}(r) = \frac{1}{\lambda_2}\mathbf{E}_{o1}(X_2(B(o,r)))$, where λ_2 is the intensity of X_2 and $\mathbf{E}_{o1}$ is the Palm expectation of having a point of X_1 at o (e.g. Illian et al., 2008). The cross pair correlation function between the two point processes is, analogous to (2.1):

$$g_{12}(r) = \frac{K'_{12}(r)}{2\pi r}. \tag{2.2}$$

Another commonly used summary is the empty space function

$$F(r) = \mathbf{P}(X(b(o,r)) \neq 0), \tag{2.3}$$

which describes the probability that a ball with radius r around the origin contains at least one point of the process.

Summary statistics have several uses in exploration, parameter estimation and evaluation of goodness-of-fit of a model. They were used for parameter estimation in an approximate Bayesian computation method in Article II and for evaluating the goodness-of-fit in Articles II and III.

2.2 Simple sequential inhibition model

The simple sequential inhibition model (e.g. Illian et al., 2008, p. 393) was used in Article II as a building block for a model for observed sweat gland locations. It is an example of a hard-core model where points cannot exist at distances smaller than a specific distance from another point. It has one parameter, the hard-core radius R.

A realisation of the model is constructed by first placing a uniformly distributed random point in a bounded window W. After that points are added one at a time, so that independent uniformly random points are proposed until they satisfy the hard-core condition that no two points are at a distance less than R from each other. This is continued until no more points can be added (there is no space left).

The hard-core property causes the generated patterns to look regular. Lotwick (1981) and Wang (1994) present efficient algorithms for simulating the process. The algorithm of Wang (1994) was implemented in Article II to obtain accurate realisations of SSI process in reasonable time.

2.3 Sequential point processes

The sequential model of Penttinen and Ylitalo (2016) is a point process with a fixed number n of points. The points are assumed to arrive sequentially, conditional on the earlier points. In the general sequential point process, the first point x_1 is assumed to follow the density f_1 and the kth point, $k = 2, 3, \ldots, n$, is assumed to follow the conditional density $f_k(y|\vec{\mathbf{x}}_{k-1})$, where $\vec{\mathbf{x}}_{k-1} = (x_1, x_2, \ldots, x_{k-1})$. The density function for the whole point pattern $(x_1, ..., x_n)$ is then

$$f_1(x_1) \prod_{k=2}^{n} f_k(x_k|\vec{\mathbf{x}}_{k-1}).$$

In the following models, which were considered in Article II, f_1 is the uniform distribution on W and $f_k = f$ for $k = 2, 3, \ldots, n$.

As a first instance we considered the model where the reweighting probability for a new point is dictated by the presence of an earlier 'neighbour', that is, a point at distance less than R. The density is

$$f(y|\vec{\mathbf{x}}_k, R, \theta) \propto \theta \mathbf{1}(M_k(y) < R) + (1-\theta)\mathbf{1}(M_k(y) \geq R), \tag{2.4}$$

where $R > 0$ and $\theta \in [0,1]$ are parameters and $M_k(y) = \min_{i \leq k} d(y, x_i)$ is the Euclidean distance (d) to the nearest earlier neighbour. This model is closely related to the recurrence-based reweighting model of Penttinen and Ylitalo (2016). In Article II we were mainly interested in cases where $\theta < \frac{1}{2}$ (repulsion between points), and especially for $\theta = 0$, when the model is a hard-core model.

We also considered the soft-core model

$$f(y|\vec{\mathbf{x}}_k, R, \kappa) \propto \exp\left(-\sum_{i=1}^{k}\left(\frac{R}{d(y, x_i)}\right)^{2/\kappa}\right), \tag{2.5}$$

with parameters $R > 0$ and $\kappa \geq 0$. This particular form, which was used in Article II, was inspired by a soft-core Gibbs process (Baddeley et al., 2015; Ogata and Tanemura, 1981, 1984).

Both (2.4) and (2.5) admit the hard-core process as a special case, namely with $\theta = 0$ for the first model and $\kappa = 0$ for the second model. Note that both models can be written in the following form:

$$f_{k+1}(y|\vec{\mathbf{x}}_k) \propto H\left(\bigoplus_{i=1}^{k} h(y, x_i)\right), \tag{2.6}$$

where h and H are functions and $\bigoplus$ is a binary operator. The hard-core model (2.4) is obtained by choosing $H(A) = \theta \mathbf{1}(A < R) + (1-\theta)\mathbf{1}(A \geq R)$, $\bigoplus = \min$ and h as the Euclidean distance. The soft-core model (2.5) is obtained by choosing $H(A) = \exp(-A)$, $\bigoplus = +$ and $h(y,x) = \left(\frac{R}{d(y,x)}\right)^{2/\kappa}$.

With the formulation (2.6), the likelihood of both models can be written as follows:

$$Z^{-1} f_1(x_1) \prod_{k=2}^{n} H\left(\bigoplus_{i=1}^{k-1} h(x_k, x_i)\right)$$

with the normalising constant

$$Z = \prod_{k=2}^{n} \int H \left(\bigoplus_{i=1}^{k-1} h(y, x_i) \right) dy. \tag{2.7}$$

This formulation allows for efficient likelihood calculations, as we shall see in Section 5.

Penttinen and Ylitalo (2016) considered more general sequential models which included a heterogeneity term $\alpha(y)$ and a Markov transition kernel $K_k(y, x_k)$ for the transition from the current point x_k to the next point y. If we extend (2.6) by multiplying it with $\alpha(y)K_k(y, x_k)$ and allowing the "sum" to stop at $k-1$ instead of k, then the extended formulation accommodates all the models considered in Penttinen and Ylitalo (2016).

2.4 Log Gaussian Cox process

Log Gaussian Cox process (LGCP) (Møller et al., 1998; Møller and Waagepetersen, 2004) is a Poisson process with random intensity e^Z, where $Z = \{Z(s) : s \in \mathbb{R}^2\}$ is a Gaussian field (i.e. any vector $(Z(s_1), \ldots, Z(s_k))$ is Gaussian). The distribution of a Gaussian field (and the LGCP) is characterised by the mean function $\mu(s) = \mathbf{E}Z(s)$ and the covariance function $C(s_1, s_2) = \mathrm{cov}(Z(s_1), Z(s_2))$. We restrict our attention to stationary and isotropic covariance functions that only depend on the distance between the two locations, that is, $C(s_1, s_2) = C(d(s_1, s_2))$.

Our main interest will be on the Matérn covariance function

$$C(r) = \sigma^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\sqrt{8\nu} \frac{r}{\rho} \right)^{\nu} K_\nu \left(\sqrt{8\nu} \frac{r}{\rho} \right), \quad r > 0, \tag{2.8}$$

where ν is the smoothness parameter, ρ is the range parameter, σ^2 is the variance parameter, and K_ν is the modified Bessel function of the second kind (e.g. Cressie, 1993; Chilés and Delfiner, 1999; Banerjee et al., 2004). The parameter ν of the Matérn covariance function controls the degree of smoothness: a Gaussian random field with the Matérn covariance function will have $\lceil \nu - 1 \rceil$ continuous derivatives (Handcock and Stein, 1993).

2.5 Residuals

Residuals for point processes were introduced in Baddeley et al. (2005) for processes with known conditional intensity and Waagepetersen (2005) further suggested them for processes with known intensity. For simplicity of presentation, here we only consider smoothed residuals for the Poisson process which are defined with respect to the intensity $\lambda(x)$ (which is also the conditional intensity).

Residuals in the point process context are measures (Baddeley et al., 2005). Smoothed residuals, however, are functions and can be defined without first defining the residuals. Smoothed residual at location u for the pattern X and a fitted Poisson process can be defined as

$$s(u) = e(u) \left(\sum_{x \in X} k(u-x) - \int_W k(u-v)\hat{\lambda}(v)dv \right),$$

where k is a smoothing kernel, $e(u)^{-1} = \int_W k(v)dv$ is an edge correction factor (cf. Section 3.4) and $\hat{\lambda}(v)$ is the intensity given by the fitted model.

The interpretation is that a positive smoothed residual at u indicates that there are more observed points around u than predicted by the model. However, care must be taken when interpreting the smoothed residuals since they depend on the user specified smoothing kernel.

Chapter 3

Inference methods

Likelihood based methods are generally desirable also in the point process context. However, the likelihood can be difficult to compute for point process models. This is particularly the case in many latent variable models. One option to handle the likelihood is to use Laplace approximation (Tierney and Kadane, 1986). For our LGCP model (Article III) we used the Laplace approximation to marginalise the latent Gaussian field, which itself was approximated by a Gaussian Markov random field. Sections 3.1 and 3.2 explain these approximations.

For some point process models the likelihood is intractable to the extent that approximation is difficult or impossible. Several estimation methods have been developed which can accommodate such models, for instance minimum contrast methods and the pseudo-likelihood. In Article II we used approximate Bayesian computation based on simulated summaries, which is described in Section 3.3.

Section 3.4 introduces the edge correction that was used to account for partially observed covariates of the LGCP model. Preliminary data analysis and goodness-of-fit assessments are integral parts of inference. Section 3.5 introduces the Monte Carlo test, Section 3.6 explains a multiple testing correction used with functional summary statistics and Section 3.7 explains posterior predictive check based on the same principle as the Monte Carlo test.

3.1 Laplace approximation

Laplace approximation (e.g. Tierney and Kadane, 1986) can be used to compute marginal likelihoods by using optimization. It is used for example in the popular integrated nested Laplace approximation framework of Rue et al. (2009). Laplace approximation is based on the assumption that the likelihood has approximately normal shape near the maximum. Let $L(x;\theta)$ be the likelihood of the latent variables x. The Laplace approximation states that

$$\int L(x;\theta)dx \approx \sqrt{\frac{(2\pi)^n}{\det(-H_\theta(\hat{x}_\theta))}} L(\hat{x}_\theta;\theta),$$

where H_θ and $\hat{x}_\theta$ are the Hessian and maximiser of $\log L(x;\theta)$ and n is the dimension of x.

3.2 Gaussian random field approximations

Assume that Z is a Gaussian random field in $\mathbb{R}^2$ with mean zero and the Matérn covariance function (2.8) with $\nu = 0, 1$ or 2. Lindgren et al. (2011) have shown that Z can be approximated by a Gaussian Markov random field on a regular grid with a sparse precision matrix. In general, the sparsity of the precision matrix depends on the smoothness parameter ν. The precision matrix element is non-zero only when $|i-k|+|j-l| \leq \nu+1$.

For example if $\nu = 2$, which was used in Article III, then the precision matrix element for the grid cells (i,j) and (k,l) with $|i-k|+|j-l| \leq 3$ is

$$Q_{(i,j),(k,l)} = \sigma^{-2} 4\pi\nu(a-4)^\nu A_{|i-k|,|j-l|},$$

and 0 otherwise. Here

$$A = \begin{pmatrix} a(a^2+12) & -3(a^2+3) & 3a & -1 \\ -3(a^2+3) & 6a & -3 & \\ 3a & -3 & & \\ -1 & & & \end{pmatrix}$$

and $a = \kappa^2 + 4$, $\kappa = \sqrt{8\nu}/\rho$, where ρ is the range parameter in multiples of grid cell width and σ^2 is the variance.

3.3 Approximate Bayesian computation

Approximate Bayesian computation (ABC) (e.g. Sunnåker et al., 2013) can be viewed as the Bayesian version of the minimum contrast method, which has been used in the point process context before (e.g. Illian et al., 2008; Diggle, 2003; Stoyan and Stoyan, 1994; Møller and Waagepetersen, 2004). The minimum contrast method can be characterised as the method of moments for functional summary characteristics.

The ABC approach differs from the minimum contrast method in several ways. As a Bayesian method it aims for producing an approximate posterior distribution for the parameters instead of a point estimate. The minimum contrast method is based on a known formula for the expected value of the summary characteristic, whose discrepancy with respect to the corresponding data summary is minimised. In contrast, ABC is based on random samples of summary statistics simulated from the model with different parameter values.

The ABC methods sample from the approximate posterior $\pi_\epsilon(\theta) \propto \text{prior}(\theta) L_\epsilon(\theta)$, where the approximate likelihood is

$$L_\epsilon(\theta) = \mathbf{P}(\|s(Y_\theta) - s(y)\| \leq \epsilon). \tag{3.1}$$

Here Y_θ is simulated from the assumed model with parameter θ, y is the observed data, s is the user specified summary and ϵ is a tolerance parameter, balancing between approximation accuracy (smaller ϵ) and computational efficiency (larger ϵ).

We obtained samples from the ABC-posterior based on the approximate likelihood (3.1) by using MCMC as explained in Vihola and Franks (2020). The method selects the threshold ϵ automatically and the only user supplied parameter is the summary function s.

3.4 Edge corrections for a covariate based on a point pattern

Typically only a part of the point pattern of interest is observed and there is interaction between the points in the observation window and the nearby points outside the window. This interaction can cause so called edge effects in simulation and estimation. For example, summary functions typically measure different kinds of interactions between points and, due to the unobserved points, a naive estimator will be biased. To estimate summary functions with less bias, different edge corrections are used (e.g. Stoyan and Stoyan, 1994; Baddeley and Gill, 1997).

The simplest edge correction is the minus sampling (e.g. Illian et al., 2008), that is, restricting the attention to W^-, a smaller window where the neighbourhoods of all points are known. This is wasteful and will lead to discarding possibly a large amount of data. A better situation would be if the point pattern was observed in a larger window W^+, that is, neighbourhoods would be observed for all points in W. This is called plus sampling (e.g. Illian et al., 2008). Unfortunately, plus sampling is often infeasible.

In Article III, we modelled a point pattern $\boldsymbol{y}$ using another point pattern $\boldsymbol{x}$ as a covariate, both observed in the window W. The pattern $\boldsymbol{x}$ was influencing the intensity of $\boldsymbol{y}$ through a computed covariate field f. The covariate field value $f(s, \boldsymbol{x})$ at a location $s \in W$ depended on the points near s. Thus some form of edge correction was necessary, and we adopted the one proposed by Kühlmann-Berenzon et al. (2005).

The edge correction is based on a simplifying assumption that the point pattern $\boldsymbol{x}$ is a realisation of homogeneous Poisson process X restricted to the observation window W. Under this assumption we use the expected contribution of the unobserved points to define the Poisson corrected covariate value as

$$\tilde{f}(s, \boldsymbol{x}) = f(s, \boldsymbol{x}) + \mathbf{E} f(s, X_{W^c}),$$

where X_{W^c} is the restriction of X to W^c, the complement of W.

If f is a sum over the points of the process X, then the expectation can be computed using Campbell's theorem (e.g. Chiu et al., 2013). This is the case with the C shown in (6.1), which was used in Article III. For this function, the resulting integrals could be efficiently evaluated using the Fast Fourier Transform (e.g. Oppenheim et al., 1999; Frigo and Johnson, 2005).

3.5 Monte Carlo testing

In point process statistics the distributions of test statistics are usually not available in an accessible form (e.g. Illian et al., 2008). Thus Monte

Carlo testing is commonly used. The Monte Carlo test is applicable when it is possible to simulate under the null hypothesis. The test statistic T_1 is computed from the data and $n-1$ test statistics $T_2, T_3, \ldots, T_n$ are simulated under the null hypothesis. Suppose that significance test with size α is required. Then n has to be chosen such that $m = \alpha n$ is an integer. The hypothesis is rejected if the observed test statistic is among the m most extreme values of the n test statistics. The test is exact in the sense that the type I error is precisely α (e.g. Marriott, 1979).

Model fitting involves finding parameter values and when testing hypothesis with those parameter values the hypothesis is said to be composite. The simple Monte Carlo test discussed above will not be exact in general for such a composite hypothesis. An adjusted test was introduced in Dao and Genton (2014), see also Baddeley et al. (2017). The simple but possibly biased test is still often used in practice since the adjusted test can be computationally prohibitive.

Since the summaries are usually functional in the point process context, that is, instead of T_i, we have, $T_i(x)$, the definition of extremeness is more complicated than with real valued summaries. Traditionally a simple scalarisation such as an integral or a maximum of the difference between the functional summary and a reference level has been used (e.g. Diggle, 1979; Ripley, 1979). This can be sufficient if a simple decision is enough. However, it is often necessary to understand why the hypothesis was rejected. In this case global envelopes, which will be discussed next, can be useful.

3.6 Global envelope

With functional test statistics, global envelope test (Myllymäki et al., 2017; Myllymäki and Mrkvička, 2020) can be used to produce an envelope that has the property that the significance of the test is equivalent to having the observed statistic not stay inside the envelope.

Let $T_i(x), i = 1, \ldots, n$ be the realisations of the functional test statistics from which global extreme rank envelope is to be constructed. First, local ranks of the test statistics are computed as $r_i(x) = \min(r_i^a(x), r_i^d(x))$, where

$$r_i^a(x) = 1 + \sum_j \mathbf{1}(T_j(x) < T_i(x)) + \frac{1}{2}\sum_{j \neq i} \mathbf{1}(T_j(x) = T_i(x))$$
$$r_i^d(x) = 1 + \sum_j \mathbf{1}(T_j(x) > T_i(x)) + \frac{1}{2}\sum_{j \neq i} \mathbf{1}(T_j(x) = T_i(x))$$

are the mid-ranks in ascending and descending order. The extreme rank of a functional test statistic can be computed as $R_i = \min_x r_i(x)$. For instance, the 95%-critical rank is defined as the 5%-quantile $R^c = \max\{R : \frac{1}{n}\sum_j \mathbf{1}(R_j < R) \leq 0.05\}$. Finally the global envelope is defined as

$$T_{low}(x) = \min\nolimits^{R^c}\{T_1(x), T_2(x), \dots, T_n(x)\}$$
$$T_{upp}(x) = \max\nolimits^{R^c}\{T_1(x), T_2(x), \dots, T_n(x)\},$$

where $\min^k$ and $\max^k$ denote kth smallest and largest values, respectively.

3.7 Posterior predictive check

Posterior predictive checks (e.g. Gelman et al., 2013) can be used to assess the fit of the model in the Bayesian setting. Posterior predictive checks are similar to hypothesis testing in practice, but in theory they are quite different. Roughly speaking, if the model fit is good, the summary statistic of the data should be consistent with the posterior predictive distribution of the summary statistic.

Samples from the posterior predictive distribution are generated as follows. Let $\theta_i, i = 1, 2, \dots, n$ be samples from the posterior. Simulate data X_i from the studied model with parameters θ_i and compute the summary statistic T_i based on X_i. Then $T_i, i = 1, 2, \dots, n$ is a sample from the posterior predictive distribution of the summary statistic.

Global envelopes can be used to visualise the posterior predictive distribution of functional summaries. We used this approach for model assessment in Articles II and III.

Chapter 4

Estimation of forest characteristics from a sample plot measured with terrestrial laser scanning

Terrestrial laser scanning (TLS) is a technique where a laser scanner measures distance from a single point to the nearest visible obstacle in a large number of directions (Liang et al., 2016). There is great interest to replace some of the traditional field measurements in forest inventories by TLS (Liang et al., 2016). In the forest inventory context, extracting single tree characteristics from the 3D point cloud obtained by TLS is an interesting research topic in itself (e.g. Liang et al., 2012; Olofsson et al., 2014; Olofsson and Holmgren, 2016; Pitkänen et al., 2019). Here we assume that a 2D map of trees with diameter at breast height has been extracted from a single scan.

Single scan TLS cannot see behind trees or other obstacles as illustrated in Figure 4.1 A. This limits the usefulness of the technique in forest inventories (Liang et al., 2016). Several corrections for the nondetection have been considered. For example Ducey and Astrup (2013) and Astrup et al. (2014) proposed distance sampling methods. Seidel and Ammer (2014) corrected for the shadowing based on the visible area which is shown in Figure 4.1 A. Olofsson and Olsson (2018) investigated the same correction for different detection rules. Detection rules model the necessity to detect more than a single point on the stem of the tree, for example, the whole tree could be required or only the center point of the tree. Olofsson and Olsson (2018) observed a large bias with some detection rules even when the trees in the forest were assumed to follow a Poisson process.

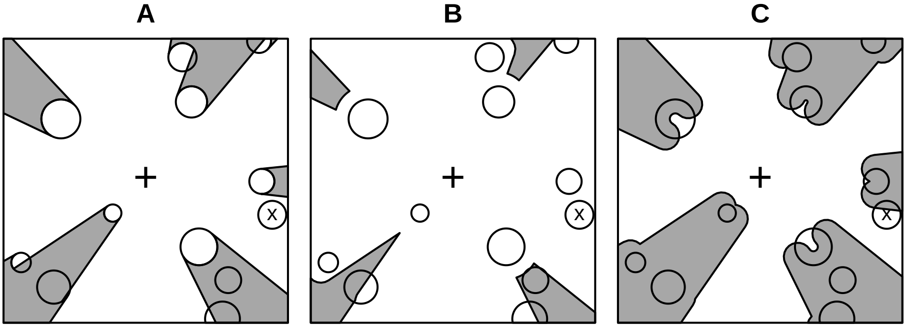


Figure 4.1: The trees (circles), scanner (+) and a test tree (x). A tree with size similar to the test tree is visible to the scanner if its center is in the white region. Panels A, B and C correspond to the detection rules 2, 1 and 4, respectively.

In Article I we proposed an estimator where the visible area is computed taking into account the detection rule. For example if the complete stem is required for detection then the area shadowed by each tree is larger than when only the center point of the tree is required to be seen (see Figure 4.1 A and C). Moreover the shadowing also depends on the size of the tree to be detected.

We considered four detection rules which were also considered in Olofsson and Olsson (2018). Let $V(X)$ denote the visible region seen from the scanner, as in Figure 4.1 A, corresponding to forest (marked point pattern) X.

1. Visible detection: A tree at location x with diameter d is detected if any part of it is visible

$$v(x, d, X) = \mathbf{1}(B(x, d/2) \cap V(X) \neq \emptyset).$$

2. Center detection: A tree is detected if its center point x is seen

$$v(x, d, X) = \mathbf{1}(x \in V(X)).$$

3. Proportional detection: This is a weighted version of the detection rule 1. Trees are weighted by the proportion of their circumference that is visible. Here we use the circle $\partial B(x, d/2)$ to define the visible

proportion. In this case,

$$v(x, d, X) = \frac{\mathrm{len}(\partial B(x, d/2) \cap V(X))}{\mathrm{len}(\partial B(0, d/2))},$$

where len is the boundary length.

4. Complete detection: A tree is detected if it is completely visible

$$v(x, d, X) = \mathbf{1}(B(x, d/2) \subset V(X)).$$

For detection rule 2 our estimator coincides with the estimator of Olofsson and Olsson (2018). Figure 4.1 shows detected trees for detection rules 1, 2 and 4.

Consider a marked point pattern X (the trees) in W (the forest stand) with marks: diameter d and the mark of interest m. If the interest is in the tree density, the mark of interest is 1. The aim is to estimate

$$T = \frac{1}{|W|} \sum_{(x,d,m) \in X} m\mathbf{1}(x \in W),$$

where $|W|$ is the size of the window W. The proposed estimator for T is

$$\hat{T} = \sum_{(x,d,m) \in X} m \frac{v(x, d, X)}{w(d, X)},$$

where $w(d, X) = \int_W v(y, d, X)dy$ and v is the detection rule. We showed in Article I that $\mathbf{E}\hat{T} \geq \mathbf{E}T$ if X is a Poisson process and argued that the bias can be small.

A simulation study was executed to evaluate the performance also in non-Poisson cases and to compare with the estimator proposed by Olofsson and Olsson (2018). We found that the proposed estimator works consistently well with different detection rules, solving the problem with the earlier estimator that different detection rules could cause large biases, even in the Poisson case. Moreover, the proposed estimator was found to have positive bias with regular patterns and negative with clustered patterns.

Chapter 5

Analysis of sweat gland patterns

The dynamic sweat test was presented by Provitera et al. (2010) to help diagnose neuropathy, which could be caused by diabetes, for example. In this test, sweating is stimulated by placing a patch with pilocarpine gels on the test site, foot or calf. Then, the test site is dried and painted with iodine solution. Finally, a camera is placed on the skin and a video is recorded for 60 seconds. The accumulation of sweat is visible on the video. Typically subjects with neuropathy have less sweat produced and smaller number of active sweat glands than healthy controls.

The data analysed in Article II have been collected by Dr. William R. Kennedy's group at the University of Minnesota by using the dynamic sweat test. The sample consists of 5 healthy controls, 5 subjects who had reported having symptoms of neuropathy (MNA), and 5 with diagnosed neuropathy (MNA Diagnosed). The size of the video frame is 2592×1944 pixels corresponding to 17.5×13 mm^2. The video shows the sweat produced as dark areas (Figure 5.1).

In Article II we sought spatial characteristics that could characterise the healthy and diseased patterns. The first step was to extract point patterns from the video data. To extract the coordinates of the individual sweat glands, that is the point patterns, from the videos (see Figure 5.1), several image analysis steps were needed. Instead of a frame-by-frame approach, we applied an algorithm based on the detection of a change point to the time series of grey scale values for each individual pixel. This pixel-by-pixel approach suits to the video sequences, where the sweat accumulates and does not dry once it has appeared, much better than going

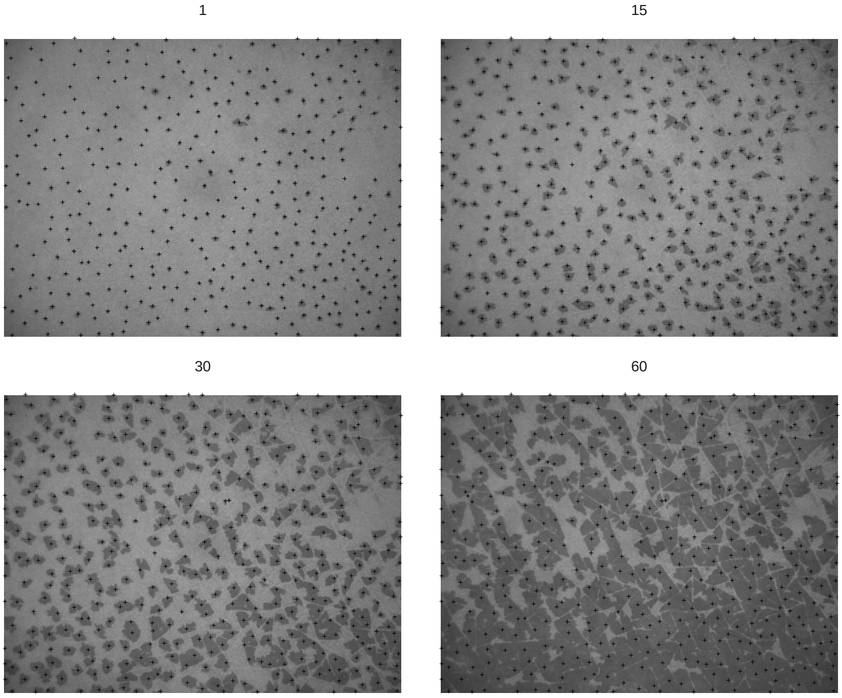


Figure 5.1: A sequence of snapshots at 1 sec (top left), 15 sec (top right), 30 sec (bottom left), and 60 sec (bottom right) of one control subject with extracted gland locations (+). The size of the image is 17.5×13 mm^2.

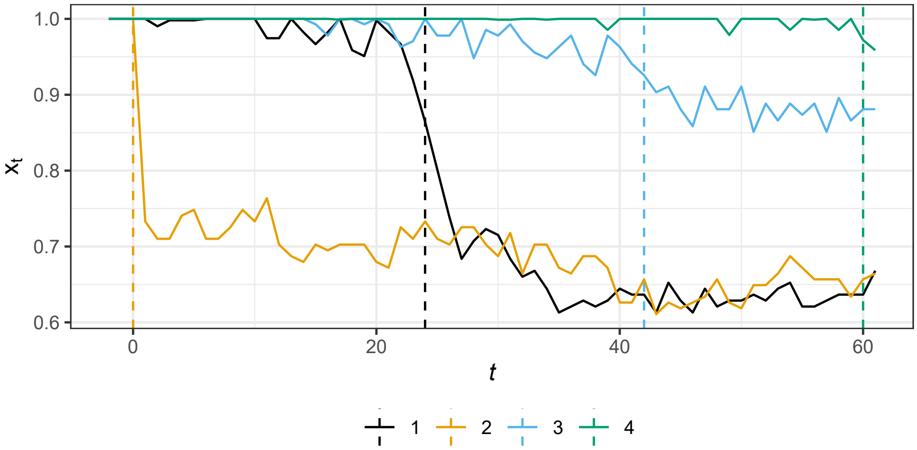


Figure 5.2: Time series for four pixels with estimated jump locations (frames) marked by dashed lines.

through the videos frame by frame, because the sweat gland locations are the easiest to detect at times when the sweat first appears.

Testing for a change point is a well studied problem in statistics (e.g. Hinkley, 1970; Yao and Davis, 1986; James et al., 1987; Pettitt, 1979). Here, however, the problem is not purely of statistical nature. There appeared to be some jumps possibly due to the changing lightning conditions that were not real jumps caused by sweat accumulation but clear enough to be detected by a statistical change point test. In Figure 5.2, the series number 3 (blue curve) shows an example of such a spurious change point. We used a similar principle as in the statistical change point tests to locate the most likely change point. To decide if the change point was caused by sweat filling a pixel we applied a simple threshold instead of a statistical test. Only changes that were larger than the threshold were accepted. The threshold for each individual video was chosen from a range of values by looking at the final result for each threshold and selecting the one that gave the best result visually.

More precisely, the change point was defined as the integer value $1 \leq t < T$ that minimises $f_t = s^2_{1,t} + s^2_{t+1,T}$, where T is the length of the time series and $s^2_{i,j}$ is the sample variance of $x_i, x_{i+1}, \ldots, x_j$. The main difference compared to the methods presented in the literature (e.g. Hinkley, 1970) is that we minimised the sum of sample variances instead of the sums of squared residuals.

The method was originally developed in Article II as a tool to find the change point, but it can be also seen as a maximum a posterior estimate for a Bayesian model. The model is the following. Let t be a potential change point, μ_1 and μ_2 the means of the process before and after the change point and $x_1, x_2, \ldots, x_T$ the time series. The change point model is

$$x_i \sim \begin{cases} N(\mu_1, t/2), & \text{for } i \leq t \\ N(\mu_2, (T-t)/2), & \text{for } i > t. \end{cases}$$

Additionally a U-shaped prior is needed for t

$$\text{prior}(t) \propto t^{t/2}(T-t)^{(T-t)/2}.$$

This prior quite strongly prefers change points in the beginning or end of the series. Other parameters had improper uniform priors.

We first modelled the activation of sweat glands using sequential point processes (Section 2.3). The first try was the hard-core process (2.4). We quickly observed that a hard-core model was too hard, producing too sharp changes in the pair-correlation functions, and instead tried the soft-core model (2.5). We fitted the models with maximum likelihood. Since the density of the soft-core process (2.5) was known only up to a constant factor, we computed the normalisation factor similarly as Penttinen and Ylitalo (2016).

In particular, we applied numerical integration with regular grid of integration points to the integrals (2.7). Following the formulation in Section 2.3, the numeric approximation of $\log Z$ can be written as

$$\sum_{k=2}^{n} \log \sum_{j=1}^{J} w_j G \left(\bigoplus_{i=1}^{k-1} g(y_j, x_i) \right),$$

where y_j are the integration points, and w_j are the integration weights. Since $g(y_j, x_i)$ are independent of k, their computation can be reused. This reduces the number of evaluations of $g(y_j, x_i)$ and the operator $\bigoplus$ for each integration point from quadratic to linear. This is a key to more efficient inference.

For some sweat gland patterns, the soft-core model gave results that were not completely satisfactory. On the left of Figure 5.3, the simulated pair correlation functions are notably different from the observed pair correlation function. The problems were associated with the patterns that had non-zero pair correlation well before the steep upward slope. Careful inspection of the videos revealed that the non-zero pair correlation was mainly caused by erroneously detected points close to other points.

Our next sequential model was an attempt to accommodate such anomalies by a mixture model where the density of the next point was a mixture of soft-core and uniform components,

$$f(y|\vec{\mathbf{x}}_k, R, \kappa, \theta) = (1-\theta) f_{\mathrm{SC}}(y|\vec{\mathbf{x}}_k, R, \kappa) + \frac{\theta}{|W|},$$

where f_{SC} is the soft-core model (2.5) and θ is the mixture proportion. The middle panel in Figure 5.3 shows the pair correlation function of this model. The upward slopes of the simulated and observed pair correlation functions are located reasonably close to each other. Thus, the model with the mixture component was superior compared to the soft-core model. However, the uniform distribution was not a realistic model for the erroneously detected points. Visual inspection revealed that the erroneous points were always points that were close to other points. For this reason the mixture proportions were not usefully related to the proportion of errors, and it is better to regard the mixture model as an estimation method for the soft-core model in the presence of erroneous points.

The mixture model was much better but still not quite satisfactory. There was still a clear difference in the steepness of the slope of the simulated and observed pair correlation functions (Figure 5.3, middle). Our next model was a completely different generative model for the sweat gland formation and activation processes.

In the generative model the formation of sweat glands and their activation were modelled separately. We started with a simple sequential inhibition (Section 2.2) process with hard-core parameter R. To get some softness in the process we added independent isotropic identically distributed Gaussian perturbation with standard deviation σ to each point. To model the activation we further applied an independent thinning with probability $(1-p)$.

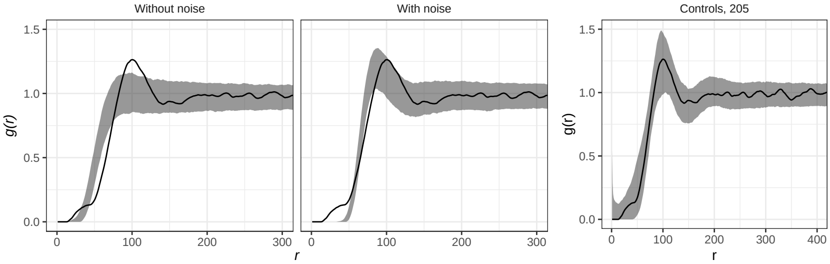


Figure 5.3: Empirical pair correlation functions (black lines) for subject 205 in the end of the video recording together with 95% global envelopes (grey areas) constructed from 25000 simulations from the soft-core model estimated without (left) and with (middle) noise and from the posterior predictive distribution of the generative model (right). Here, $r = 100$ corresponds to approximately 0.7 mm.

Since the likelihood of the proposed generative model is not tractable, but the model is relatively straightforward to simulate, we used approximate Bayesian computation (ABC) inference. For such a regular point process model, it was natural to use summary statistics based on the pair correlation function g, illustrated in Figure 5.3. Intuitively, the location of the upward slope of the pair correlation function would be informative about the hard-core parameter R. The steepness of the slope would be informative about the noise component σ. Further the activation probability p is closely related to the amount of empty space. To capture the location and steepness of the upward slope of the pair correlation function (2.1) we selected the smallest distances $r_1, r_2 \geq 10$ pixels (approximately 0.07 mm) where $g(r_1) = 0.75$ and $g(r_2) = 1$ as summaries. For the empty space we chose $r_3 = F^{-1}(0.5)$, where F is the empty space function (2.3). These choices helped avoiding the effect of erroneous detection of points close to other points.

In Figure 5.3 (right) the simulated pair correlation functions are very close to the observed one, even though the simulation uses the posterior predictive distribution instead of a point estimate for the model parameters. We also inspected the posterior predictive distributions of the empty space function and the K-function. These summaries did not show evi-

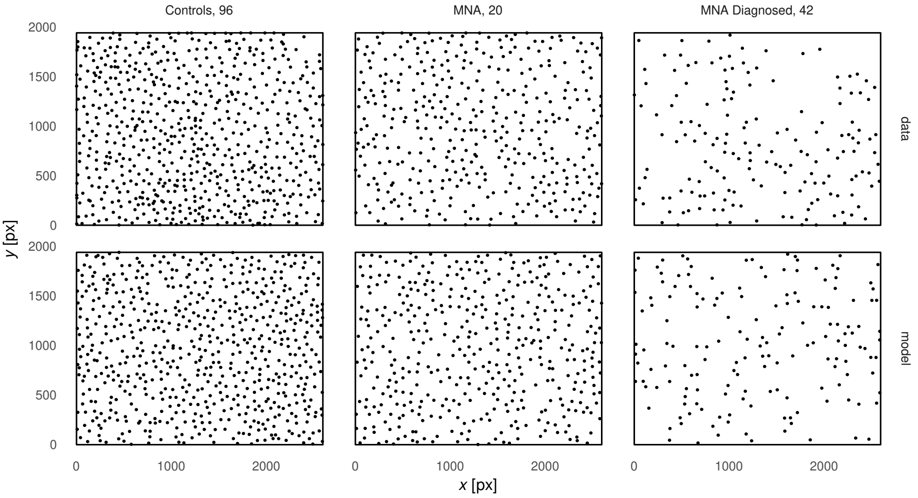


Figure 5.4: The original point patterns (top) and patterns generated from the corresponding posterior predictive distributions of the generative model (bottom) for one subject from each group (96, 20 and 42). The size of the window is 17.5 × 13 mm^2.

dence of substantial misfit. Figure 5.4 shows examples of simulated and observed point patterns, which look very similar. Even some empty regions similar to those observed in Provitera et al. (2010) are present in the simulated patterns. The activation probability was higher for controls than other groups agreeing with earlier studies where controls were found to have higher density of sweat spots than other groups.

Chapter 6

Modelling the regeneration patterns

Recently there has been a lot of interest in continuous cover forestry (Kusinen et al., 2019; Juutinen et al., 2018). The basic idea of continuous cover forestry is that clear cuts are not performed and thus a forest always looks more or less like a forest that has trees from all size classes. However there are a lot of management decisions that have to be made. One important question is what kind of management would ensure a successful regeneration. This motivates the research of Article III: We aimed for modelling the effect of large trees on the seedlings to build understanding of the regeneration in a continuous cover forest managed by selection cutting.

We modelled the intensity of new seedlings in an uneven-aged forest given the locations of large trees. The data consist of 14 sample plots in Southern Finland (Figure 6.1) (Eerikäinen et al., 2007; Eerikäinen et al., 2014; Saksa and Valkonen, 2011). We used a log Gaussian Cox process (LGCP) since the seedling patterns are clustered. Following Pommerening and Grabarnik (2019) we modelled the effect of the large trees X as an influence field. The influence field

$$C(s;X) = \sum_{[x,m]\in X} c(\|s-x\|, m). \tag{6.1}$$

is a superposition of kernels c attached to each tree. We used the influence kernel

$$c(h,m) = m^{\alpha} \exp\left(-\left(\frac{h}{\theta m^{\delta}}\right)^2\right) \tag{6.2}$$

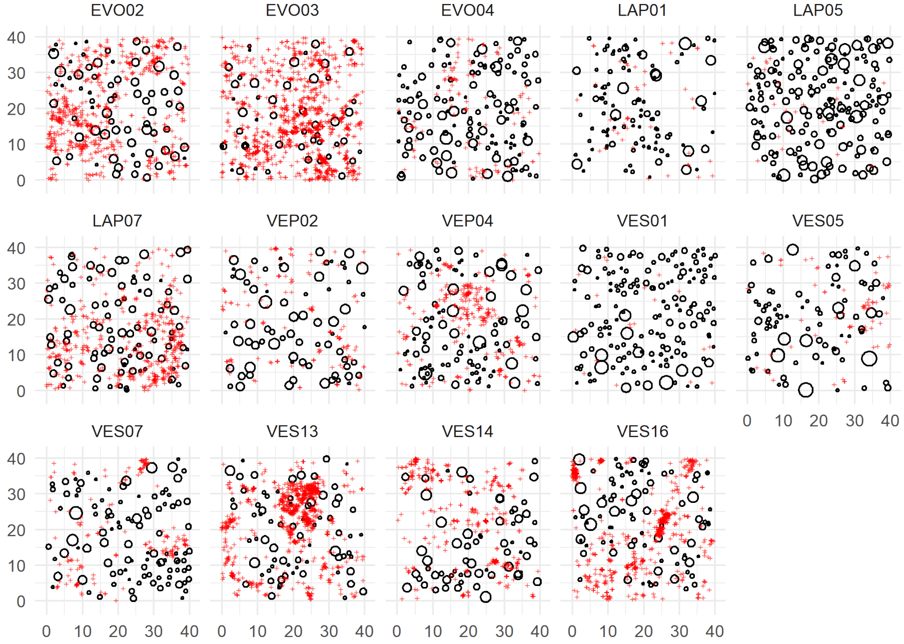


Figure 6.1: Trees with diameter at breast height (dbh) at least 7 cm (open circles with radii relative to the dbh of the tree) and new seedlings (red crosses) in areas of size 40 m × 40 m. The headings give abbreviations for the plot locations and numbers.

where $\theta > 0$, $\delta \geq 0$, and $\alpha \geq 0$ are parameters.

The model for the seedlings was a LGCP. The mean function of the Gaussian process was

$$\beta_0 + \beta_1 C(s; X),$$

where β_0 and β_1 are parameters, and the covariance function was the Matérn covariance function (2.8).

We modelled the sample plots as independent replicates of the same LGCP but with plot specific intercepts β_{0k} to accommodate varying seedling intensities. We discretised the problem using a regular grid. Since the discretised random field has quite high dimension and the covariate is nonlinear there were no readily available software to do the inference.

Instead, we used the Laplace approximation (Section 3.1) and GMRF (Section 3.2) to deal with the high dimension of the latent random field, and MCMC to estimate the few remaining parameters. Since the large tree pattern was only observed in the same window as the seedlings we applied the Poisson correction (Section 3.4) for the influence field.

We conducted a simulation experiment to assess the performance of the Poisson correction, comparing it to no correction and plus sampling, which is the ideal case. The large tree patterns were generated from either a Poisson process or a hard-core process (see Article III for details). If the influence of the large trees was modest, the edge effect was small, but when the influence had somewhat long range, the Poisson correction produced results that were much closer to the plus sampling than without any correction.

We tried four different influence kernel models for our tree data, and one model without influence kernel. Three of the models were mark dependent. The first model was (6.2) and the two others were (6.2) with either α or δ set to zero. The fourth model was mark independent, that is, (6.2) with both α and δ set to zero. The last model had no influence kernel.

We constructed 95% global envelopes (Section 3.6) based on posterior predictive distributions (Section 3.7) for several summary statistics (see Article III for details). Figure 6.2 shows the global envelopes for the cross pair correlation function (2.2) between the trees and seedlings for the complete model with three parameters (dotted lines), the mark independent model (grey shade) and the model with no influence (dashed lines). All mark dependent influence models produced very similar envelopes. The other summary functions, that only described the seedling pattern, were very similar for all models, see Article III for the figures. Based on the envelope tests using the influence kernel had a strong effect on the model fit, while using the mark was not important. According to the final model with mark independent influence kernel, the seedlings had a preference for locations with no large trees in the immediate vicinity.

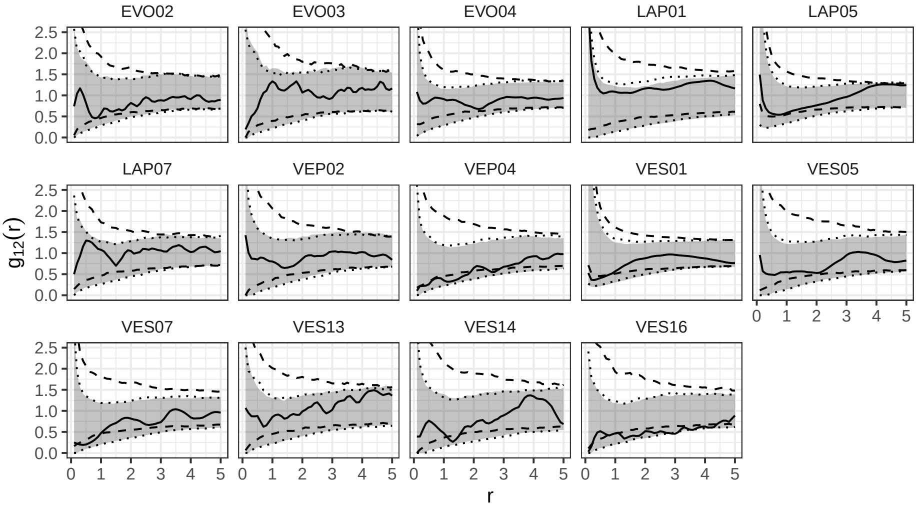


Figure 6.2: Empirical cross pair correlation functions (solid line) between trees and seedlings together with the 95% global envelopes constructed from 10 000 simulations from the posterior predictive distribution of the fitted LGCP models for the 14 plots in Figure 6.1 with mark independent (grey shade), mark dependent (dotted lines), and no (dashed lines) influence.

Chapter 7

Testing for a covariate effect on a point pattern

It can be useful to explain the intensity of a point pattern using a covariate field. If there are multiple covariates the question arises which of the covariates should be included in the model. In Article IV we proposed two new spatial test statistics for testing the hypothesis that a particular covariate should be included in a model. They both give information on the locations where the covariate has a significant effect on the point pattern intensity. The first test statistic directly compares the residuals of the competing models while the second one, loosely speaking, compares the gradients of the residuals.

For simplicity of presentation we assume that there is an interesting covariate field C and a nuisance covariate field U. Moreover we assume that we have observed a point pattern X which is modelled as the inhomogeneous Poisson process with intensity

$$\lambda(u) = \exp\left(\beta C(u) + \gamma U(u)\right), \tag{7.1}$$

where β and γ are regression coefficients. The question of interest is whether the covariate C should be included in the model, or not. There are many alternatives for performing such a test, for example, the likelihood ratio test (Baddeley et al., 2015, p. 372), and the Wald test (Baddeley et al., 2015; Coeurjolly and Rubak, 2013; Waagepetersen and Guan, 2009), which are both based on asymptotic distributions of the test statistics.

We considered two functional test statistics. The first proposed test statistic is

$$F(u) = s_0(u)^2 - s_1(u)^2,$$

where s_1 is the smoothed residual field from the fitted model with intensity (7.1) and s_0 is the smoothed residual field from the fitted null model with the following intensity

$$\lambda(u) = \exp\left(\gamma U(u)\right).$$

The second proposed test statistic is

$$S(u) = \frac{\int_{B(u,R)\cap W}(s_0(x) - \bar{s}_0(u))^2 dx}{\int_{B(u,R)\cap W}(s_1(x) - \bar{s}_1(u))^2 dx}, \tag{7.2}$$

where R is a smoothing parameter, in general different from the one used in the smoothed residuals, and

$$\bar{s}_i(u) = \frac{1}{|B(u,R)\cap W|}\int_{B(u,R)\cap W} s_i(x)dx$$

is the mean value of s_i within $B(u,R)\cap W$.

The following result, which was not reported in Article IV, consolidates the interpretation of $S(u)$ as the gradient of the residuals when R is small.

Theorem 1. *If s_0 and s_1 are differentiable at u, an interior point of W, then*

$$\lim_{R\to 0} S(u) = \frac{\|\nabla s_0(u)\|^2}{\|\nabla s_1(u)\|^2}.$$

Proof. To see this let B_R stand for the ball at the origin with radius R. Since s_i is differentiable at u, the function $f_R : B_1 \to \mathbb{R}$ defined as

$$f_R(x) = \frac{s_i(u + Rx) - s_i(u)}{R}$$

converges uniformly to $f_0(x) = \nabla s_i(u)\cdot x$ as $R \to 0$. Uniform convergence allows us to interchange the order of integration and limit and thus

$$\begin{aligned}\frac{\bar{s}_i(u) - s_i(u)}{R} &= \frac{1}{|B_R|}\int_{B_R} \frac{s_i(u+x) - s_i(u)}{R} dx \\ &= \frac{1}{|B_1|}\int_{B_1} \frac{s_i(u+Rx) - s_i(u)}{R} dx \\ &\to \frac{1}{|B_1|}\int_{B_1} \nabla s_i(u) \cdot x \, dx \\ &= \frac{1}{|B_1|}\int_{B_1} \|\nabla s_i(u)\| x_1 \, dx,\end{aligned}$$

where the last equality follows from a change of variables by a rotation to align the first coordinate with the vector $\nabla s_i(u)$. The last integral is zero because of anti-symmetry of x_1. Now the numerator/denumerator of (7.2), when scaled appropriately, satisfies

$$\begin{aligned}&\frac{1}{R^2}\frac{|B_1|}{|B_R|}\int_{B_R} (s_i(u+x) - \bar{s}_i(u))^2 dx \\ &= \frac{1}{R^2}\int_{B_1} (s_i(u+Rx) - \bar{s}_i(u))^2 dx \\ &= \frac{1}{R^2}\int_{B_1} (s_i(u+Rx) - s_i(u) + s_i(u) - \bar{s}_i(u))^2 dx \\ &\to \int_{B_1} (\nabla s_i(u) \cdot x)^2 dx \\ &= \|\nabla s_i(u)\|^2 \int_{B_1} (x_1)^2 dx. \qquad \square\end{aligned}$$

Since the null distributions of these test statistics are unknown, we employ a Monte Carlo test as discussed in Section 3.5. First the null model is fitted to the data. A number n of simulations are generated from the fitted null model. Test statistics are computed for the data and the simulated patterns. A global envelope (Section 3.6) is constructed from the test statistics. If the test statistic of the data is outside the envelope then the null hypothesis is rejected.

The proposed tests easily generalise to Gibbs and Cox processes, multiple covariates and multiple nuisance covariates. In Article IV we conducted a simulation experiment to study the empirical performance of the proposed test statistics compared to the Wald test. It was found out that the performance of the proposed tests was comparable to the Wald test in the investigated cases.

Chapter 8

Discussion

This work originated from applied problems in forestry and medicine. Various challenges arose in modelling which demanded development of methodology, new models and employing a range of inference methods. Challenges were caused by anomalies such as spurious or missing points, effects of covariates in a point process model, nonlinear covariates and likelihoods that were difficult to handle.

In Article I we developed an estimator for forest characteristics based on a tree map obtained by a single terrestrial laser scan. The estimator was found to be clearly better than the earlier method of Olofsson and Olsson (2018). Later, another estimator was developed by Kansanen et al. (2021). Their simulation study, comparing all the three methods above, confirmed the findings in Article I and found little differences between our and their estimators. However, their estimator is unbiased for the Poisson case and even has an estimator for the variance.

The strategy of accounting for the spurious points in the inference worked in the case of sweat gland activation data in Article II. Based on the small data sample we were able to confirm earlier findings. However, we found no evidence against the hypothesis that the sweat glands activate independently. It might still be that there is some patterning to which the used summary functions are blind. The next step in studying the sweat gland activation would be to implement several activation strategies and evaluate the performance of the summary functions in classifying the activation patterns. If no good summary function would be found, a new summary should be developed that is more sensitive to the differences in activation patterns. Certainly, it would be interesting to study these issues for larger numbers of subjects, taking into account also subject related covariates, like age and body mass index.

In Article III we investigated seedling patterns in forests managed under continuous cover forestry, by using a LGCP with influence field induced by large trees, a nonlinear covariate. Based on the model assessment, using the influence field had a strong effect on the model fit, while the tree size played a minor role. Our findings with the final model with size independent influence kernel suggested that the seedlings preferred locations with no large trees in the immediate vicinity. We used the LGCP model as a general model for all the sample plots. During the model fitting the sample plots with very few seedlings were somewhat problematic. This was partially the reason for using the same covariance parameters for all sample plots. A possible reason for the problems is that the Laplace approximation struggles with very sparse data. Although there are multiple methods for dealing with a LGCP, a robust method for working with nonlinear covariates is missing. It is also clear that for a sample plot with very few seedlings a simpler model, like Poisson process, would be sufficient.

In Article IV we introduced two new test statistics for testing for the effect of a covariate in a parametric point process model. The introductory part of this thesis complemented Article IV by providing a small scale interpretation for the second test statistic (7.2). An advantage of the new tests is that they produce a map of significant regions and in this way provide local information, which can be useful as a model diagnostic. Another advantage is that the proposed tests could be used in situations where the Wald test is not applicable as such, for example with cluster processes estimated using the composite likelihood. Although they work also for more complex processes than the Poisson process, the computation effort required also increases significantly.

Bibliography

Astrup, R., Ducey, M. J., Granhus, A., Ritter, T., and von Lüpke, N. (2014). Approaches for estimating stand-level volume using terrestrial laser scanning in a single-scan mode. *Canadian Journal of Forest Research 44*(6), 666–676. doi: 10.1139/cjfr-2013-0535.

Baddeley, A. and Gill, R. D. (1997). Kaplan-Meier estimators of distance distributions for spatial point processes. *The Annals of Statistics 25*(1), 263–292.

Baddeley, A., Hardegen, A., Lawrence, T., Milne, R. K., Nair, G., and Rakshit, S. (2017). On two-stage Monte Carlo tests of composite hypotheses. *Computational Statistics & Data Analysis 114*, 75–87. doi: 10.1016/j.csda.2017.04.003.

Baddeley, A., Rubak, E., and Turner, R. (2015). *Spatial Point Patterns: Methodology and Applications with R.* London: Chapman and Hall/CRC Press.

Baddeley, A., Turner, R., Møller, J., and Hazelton, M. (2005). Residual analysis for spatial point processes (with discussion). *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 67*(5), 617–666. doi: 10.1111/j.1467-9868.2005.00519.x.

Banerjee, S., Carlin, B. P., and Gelfand, A. E. (2004). *Hierarchical Modeling and Analysis for Spatial Data* (1 ed.). Boca Raton: Chapman & Hall/CRC.

Chilés, J.-P. and Delfiner, P. (1999). *Geostatistics: Modeling Spatial Uncertainty.* New York: Wiley.

Chiu, S. N., Stoyan, D., Kendall, W. S., and Mecke, J. (2013). *Stochastic Geometry and its Applications* (3 ed.). Chichester: Wiley.

Coeurjolly, J.-F. and Rubak, E. (2013). Fast covariance estimation for innovations computed from a spatial Gibbs point process. *Scandinavian Journal of Statistics 40*(4), 669–684. doi: 10.1111/sjos.12017.

Cressie, N. A. C. (1993). *Statistics for Spatial Data* (Revised ed.). Wiley Series in Probability and Mathematical Statistics. New York: Wiley.

Dao, N. A. and Genton, M. G. (2014). A Monte Carlo adjusted goodness-of-fit test for parametric models describing spatial point patterns. *Journal of Computational and Graphical Statistics 23*, 497–517. doi: 10.1080/10618600.2012.760459.

Diggle, P. J. (1979). On parameter estimation and goodness-of-fit testing for spatial point patterns. *Biometrics 35*, 87–101.

Diggle, P. J. (2003). *Statistical Analysis of Spatial Point Patterns* (2 ed.). London: Arnold.

Ducey, M. J. and Astrup, R. (2013). Adjusting for nondetection in forest inventories derived from terrestrial laser scanning. *Canadian Journal of Remote Sensing 39*(5), 410–425. doi: 10.5589/m13-048.

Eerikäinen, K., Valkonen, S., and Saksa, T. (2014). Ingrowth, survival and height growth of small trees in uneven-aged Picea abies stands in southern Finland. *Forest Ecosystems 1:5*. doi: 10.1186/2197-5620-1-5.

Eerikäinen, K., Miina, J., and Valkonen, S. (2007). Models for the regeneration establishment and the development of established seedlings in uneven-aged, Norway spruce dominated forest stands of southern Finland. *Forest Ecology and Management 242*(2), 444–461. doi: 10.1016/j.foreco.2007.01.078.

Frigo, M. and Johnson, S. G. (2005). The design and implementation of FFTW3. *Proceedings of the IEEE 93*(2), 216–231. Special issue on "Program Generation, Optimization, and Platform Adaptation".

Gelman, A., Carlin, J., Stern, H., Dunson, D., Vehtari, A., and Rubin, D. (2013). *Bayesian Data Analysis* (3 ed.). Chapman and Hall/CRC.

Handcock, M. S. and Stein, M. L. (1993). A Bayesian analysis of kriging. *Technometrics 35*(4), 403–410. doi: 10.2307/1270273.

Hinkley, D. V. (1970). Inference about the change-point in a sequence of random variables. *Biometrika 57*(1), 1–17. doi: 10.2307/2334932.

Illian, J., Penttinen, A., Stoyan, H., and Stoyan, D. (2008). *Statistical Analysis and Modelling of Spatial Point Patterns* (1 ed.). Chichester: John Wiley & Sons, Ltd.

James, B., James, K. L., and Siegmund, D. (1987). Tests for a change-point. *Biometrika 74*(1), 71–83. doi: 10.2307/2336022.

Juutinen, A., Ahtikoski, A., Mäkipää, R., and Shanin, V. (2018). Effect of harvest interval and intensity on the profitability of uneven-aged management of Norway spruce stands. *Forestry: An International Journal of Forest Research 91*(5), 589–602. doi: 10.1093/forestry/cpy018.

Kansanen, K., Packalen, P., Maltamo, M., and Mehtätalo, L. (2021). Horvitz-Thompson–like estimation with distance-based detection probabilities for circular plot sampling of forests. *Biometrics 77*(2), 715–728. doi: 10.1111/biom.13312.

Kühlmann-Berenzon, S., Heikkinen, J., and Särkkä, A. (2005). An additive edge correction for the influence potential of trees. *Biometrical journal 47*, 517–526. doi: 10.1002/bimj.200310157.

Kuusinen, N., Valkonen, S., Berninger, F., and Mäkelä, A. (2019). Seedling emergence in uneven-aged Norway spruce stands in Finland. *Scandinavian Journal of Forest Research 34*(3), 200–207. doi: 10.1080/02827581.2019.1575976.

Liang, X., Kankare, V., Hyyppä, J., Wang, Y., Kukko, A., Haggrén, H., Yu, X., Kaartinen, H., Jaakkola, A., Guan, F., Holopainen, M., and Vastaranta, M. (2016). Terrestrial laser scanning in forest inventories. *ISPRS Journal of Photogrammetry and Remote Sensing 115*, 63–77. doi: 10.1016/j.isprsjprs.2016.01.006. Theme issue 'State-of-the-art in photogrammetry, remote sensing and spatial information science'.

Liang, X., Litkey, P., Hyyppa, J., Kaartinen, H., Vastaranta, M., and Holopainen, M. (2012). Automatic stem mapping using single-scan terrestrial laser scanning. *IEEE Transactions on Geoscience and Remote Sensing 50*(2), 661–670. doi: 10.1109/TGRS.2011.2161613.

Lindgren, F., Rue, H., and Lindström, J. (2011). An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 73*(4), 423–498. doi: 10.1111/j.1467-9868.2011.00777.x.

Lotwick, H. W. (1981). *Spatial Stochastic Point Processes*. Ph. D. thesis, Unversity of Bath.

Marriott, F. H. C. (1979). Barnard's Monte Carlo tests: How many simulations? *Journal of the Royal Statistical Society. Series C (Applied Statistics) 28*(1), 75–77. doi: 10.2307/2346816.

Møller, J., Syversveen, A. R., and Waagepetersen, R. P. (1998). Log Gaussian Cox processes. *Scandinavian Journal of Statistics 25*(3), 451–482. doi: 10.1111/1467-9469.00115.

Møller, J. and Waagepetersen, R. P. (2004). *Statistical Inference and Simulation for Spatial Point Processes* (1 ed.). Boca Raton: Chapman & Hall/CRC.

Myllymäki, M. and Mrkvička, T. (2020). GET: Global envelopes in R. arXiv:1911.06583 [stat.ME].

Myllymäki, M., Mrkvička, T., Seijo, H., Grabarnik, P., and Hahn, U. (2017). Global envelope tests for spatial processes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 79*, 381–404. doi: 10.1111/rssb.12172.

Ogata, Y. and Tanemura, M. (1981). Estimation of interaction potentials of spatial point patterns through the maximum likelihood procedure. *Annals of the Institute of Statistical Mathematics 33*(2), 315–338. doi: 10.1007/BF02480944.

Ogata, Y. and Tanemura, M. (1984). Likelihood analysis of spatial point patterns. *Journal of the Royal Statistical Society: Series B (Methodological) 46*(3), 496–518. doi: 10.1111/j.2517-6161.1984.tb01322.x.

Olofsson, K. and Holmgren, J. (2016). Single tree stem profile detection using terrestrial laser scanner data, flatness saliency features and curvature properties. *Forests 7*(9). doi: 10.3390/f7090207.

Olofsson, K., Holmgren, J., and Olsson, H. k. (2014). Tree stem and height measurements using terrestrial laser scanning and the RANSAC algorithm. *Remote Sensing 6*(5), 4323–4344. doi: 10.3390/rs6054323.

Olofsson, K. and Olsson, H. (2018). Estimating tree stem density and diameter distribution in single-scan terrestrial laser measurements of field plots: a simulation study. *Scandinavian Journal of Forest Research 33*(4), 365–377. doi: 10.1080/02827581.2017.1368698.

Oppenheim, A. V., Schafer, R. W., and Buck, J. R. (1999). *Discrete-Time Signal Processing (2nd Ed.)*. USA: Prentice-Hall, Inc.

Penttinen, A. and Ylitalo, A.-K. (2016). Deducing self-interaction in eye movement data using sequential spatial point processes. *Spatial Statistics 17*, 1–21. doi: 10.1016/j.spasta.2016.03.005.

Pettitt, A. N. (1979). A non-parametric approach to the change-point problem. *Journal of the Royal Statistical Society. Series C (Applied Statistics) 28*(2), 126–135. doi: 10.2307/2346729.

Pitkänen, T. P., Raumonen, P., and Kangas, A. (2019). Measuring stem diameters with TLS in boreal forests by complementary fitting procedure. *ISPRS Journal of Photogrammetry and Remote Sensing 147*, 294–306. doi: 10.1016/j.isprsjprs.2018.11.027.

Pommerening, A. and Grabarnik, P. (2019). *Individual-based Methods in Forest Ecology and Management* (1 ed.). Springer.

Provitera, V., Nolano, M., Caporaso, G., Stancanelli, A., Santoro, L., and Kennedy, W. R. (2010). Evaluation of sudomotor function in diabetes using the dynamic sweat test. *Neurology 74*(1), 50–56. doi: 10.1212/WNL.0b013e3181c7da4b.

Ripley, B. D. (1977). Modelling spatial patterns. *Journal of the Royal Statistical Society, Series B 39*(2), 172–212.

Ripley, B. D. (1979). Tests of 'randomness' for spatial point patterns. *Journal of the Royal Statistical Society, Series B 41*, 368–374.

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

Saksa, T. and Valkonen, S. (2011). Dynamics of seedling establishment and survival in uneven-aged boreal forests. *Forest Ecology and Management 261*(8), 1409–1414. doi: 10.1016/j.foreco.2011.01.026.

Seidel, D. and Ammer, C. (2014). Efficient measurements of basal area in short rotation forests based on terrestrial laser scanning under special consideration of shadowing. *iForest - Biogeosciences and Forestry 7*(4), 227–232. doi: 10.3832/ifor1084-007.

Stoyan, D. and Stoyan, H. (1994). *Fractals, Random Shapes and Point Fields: Methods of Geometrical Statistics* (1 ed.). Wiley.

Sunnåker, M., Busetto, A. G., Numminen, E., Corander, J., Foll, M., and Dessimoz, C. (2013). Approximate Bayesian computation. *PLOS Computational Biology 9*(1), 1–10. doi: 10.1371/journal.pcbi.1002803.

Tierney, L. and Kadane, J. B. (1986). Accurate approximations for posterior moments and marginal densities. *Journal of the American Statistical Association 81*(393), 82–86. doi: 10.2307/2287970.

Vihola, M. and Franks, J. (2020). On the use of approximate Bayesian computation Markov chain Monte Carlo with inflated tolerance and post-correction. *Biometrika 107*(2), 381–395. doi: 10.1093/biomet/asz078.

Waagepetersen, R. (2005). Comment on ‘Residual analysis for spatial point processes (with discussion)’ by A. Baddeley et al. *Journal of the Royal Statistical Society, Series B (Statistical Methodology) 67*(5), 617–666.

Waagepetersen, R. and Guan, Y. (2009). Two-step estimation for inhomogeneous spatial point processes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 71*(3), 685–702. doi: 10.1111/j.1467-9868.2008.00702.x.

Wang, J.-S. (1994). A fast algorithm for random sequential adsorption of discs. *International Journal of Modern Physics C 05*(04), 707–715. doi: 10.1142/S0129183194000817.

Yao, Y.-C. and Davis, R. A. (1986). The asymptotic behavior of the likelihood ratio statistic for testing a shift in mean in a sequence of independent normal variates. *Sankhyā: The Indian Journal of Statistics, Series A (1961-2002) 48*(3), 339–353.

ORIGINAL PAPERS

I

CORRECTING FOR NONDETECTION IN ESTIMATING FOREST CHARACTERISTICS FROM SINGLE-SCAN TERRESTRIAL LASER MEASUREMENTS

by

Kuronen, M., Henttonen, H. M., and Myllymäki, M.

Canadian Journal of Forest Research, 49(1):96–103, 2018.

doi:10.1139/cjfr-2018-0072

NRC Research Press
ARTICLE

Correcting for nondetection in estimating forest characteristics from single-scan terrestrial laser measurements

Mikko Kuronen, Helena M. Henttonen, and Mari Myllymäki

Updated online 17 December 2018: The license for this article has been changed to the CC BY 4.0 license. The PDF and HTML versions of the article have been modified accordingly.

Abstract: A problem in the single-scan setup of terrestrial laser scanning is that some trees are shaded by others and therefore not detected in the scan. A basic estimator for forest characteristics such as tree density or basal area is based on the visible area of a scanner. However, simply compensating for nondetection by the visible area may result in considerable bias even in Poisson forests, especially if the detection of a tree depends on its size. We propose a new estimator that is a generalization of the visible area based estimator. Most importantly, the new estimator allows different detection rules; for example, full or partial visibility of a tree can be required for detection. By a simulation study, it is shown to work adequately in different types of simulated and empirical forests with different detection rules.

Key words: terrestrial laser scanning, nondetection, single scan, spatial point pattern, tree density estimator.

Résumé : Un problème inhérent à l'utilisation du balayage laser terrestre avec un seul balayage vient du fait que certains arbres sont cachés par d'autres et ne sont donc pas détectés dans l'analyse. Un estimateur de base des caractéristiques de la forêt, comme la densité ou la surface terrière, est basé sur la zone visible d'un balayage. Cependant, une simple compensation de la non-détection par la zone visible peut entraîner un biais important même dans les forêts répondant à une distribution de Poisson, surtout si la détection d'un arbre dépend de sa taille. Nous proposons un nouvel estimateur qui est une généralisation de l'estimateur basé sur la zone visible. Plus important encore, le nouvel estimateur permet d'utiliser différentes règles de détection; par exemple, la visibilité requise pour la détection d'un arbre peut être complète ou partielle. Une étude par simulation, a montré qu'il fonctionne correctement dans différents types de forêts, simulées et empiriques, avec différentes règles de détection. [Traduit par la Rédaction]

Mots-clés : balayage laser terrestre, non détection, balayage unique, configuration spatiale de points, estimateur de la densité d'arbres.

1. Introduction

There is great interest in replacing at least part of traditional field measurements in forest inventories by terrestrial laser scanning (TLS). As soon as the best practices with TLS become known, it is expected that TLS will be used operationally in forest inventories (Liang et al. 2016). Meanwhile, there are still many open questions about its use.

TLS can be used in single- and multi-scan setups for field plot inventories. An advantage of the multi-scan situation, where the TLS instrument is positioned at several locations in the field plot, is that, at least in principle, all of the trees in the field plot can be detected (increasing the number of scans). In such a situation, the estimation of forest characteristics is rather straightforward after the tree characteristics have been extracted from the TLS point cloud, which certainly is a challenging problem on its own (see, e.g., Liang et al. 2012; Olofsson et al. 2014; Olofsson and Holmgren 2016). On the other hand, the fieldwork of the multi-scan setup can be rather time consuming, thus expensive, demanding not only that the scans be done at several locations, but also co-registration of the different scans. The main advantage of the single-scan setup is indeed that it is fast to position the scanner only in one location, typically in the middle of the field plot, which makes it an interesting technique for operational forestry (e.g., Astrup et al. 2014). The downside of the cost efficiency is that some of the trees in the field plot are not detected or they are only partly detected due to lack of visibility. The nondetection rate of trees obviously depends on the plot size and forest type, and there are several studies that have considered nondetection rates empirically: Liang et al. (2016, tables 2 and 3) provides summaries of the accuracy of stem detection and plot-level estimation of mean diameter (diameter at breast height, dbh) of the single-scan method in previous studies. Because of nondetection, the estimation of forest characteristics needs careful treatment.

The current study deals with estimation of forest characteristics after the tree stem coordinates and dbhs of trees have been extracted from a TLS point cloud. Some studies have already presented models for compensating for the nondetection problem in the single-scan setup. Jupp et al. (2005), Strahler et al. (2008), and Lovell et al. (2011), besides extensive work with TLS data including identifying tree characteristics, compensated for nondetected trees using gap probabilities under the Poisson forest assumption. Ducey and Astrup (2013) and Astrup et al. (2014) used distance sampling methods. On the other hand, Seidel and Ammer (2014) and Olofsson and Olsson (2018) compensated for the nondetected or visible area of the plot. More precisely, Olofsson and Olsson

Received 28 February 2018. Accepted 13 June 2018.

M. Kuronen, H.M. Henttonen, and M. Myllymäki. Natural Resources Institute Finland, Latokartanonkaari 9, FI-00790 Helsinki, Finland.
Corresponding author: Mikko Kuronen (email: mikko.kuronen@luke.fi).

Can. J. For. Res. **49**: 96–103 (2019) dx.doi.org/10.1139/cjfr-2018-0072

(2018) took the region visible from the scanner as the sampling window, which may be regarded as the most obvious correction for nondetection.

Our focus is solely on further development of the latter correction type, which may be attractive due to their simplicity. Olofsson and Olsson (2018) studied the performance of the simple visible area based estimator and found that if only a small part of a tree is required for the tree to be detected, i.e., to be included in the sample, then the simple estimator is seriously biased even in the Poisson forest case. The finding was similar for a detector that required full visibility of the tree. Olofsson and Olsson (2018) concluded that a careful choice of the detection rule is needed to obtain unbiased estimates even in Poisson forests. However, the detectors leading to unbiased results may not be the most reasonable ones in practice.

We propose a new estimator for forest characteristics such as tree density and basal area from a set of tree stem coordinates and dbhs extracted from a single terrestrial laser scan. This estimator allows, in principle, the use of any detector. It can be seen as a refined visible area based estimator. We study its accuracy and precision in simulated and empirical data using four different detectors.

2. Materials and methods

2.1. Visible area

Similarly as in Olofsson and Olsson (2018), we assume that nondetection of trees in a single terrestrial laser scan is caused only by the tree trunks and that the cross sections of tree trunks are perfect circles. This refers to an "easy" forest with no understorey vegetation or tree branches affecting the visibility. Therefore, given a set of tree coordinates and the tree dbhs, the visible area of a scanner can be determined simply by geometric rules. Figure 1 illustrates the visible area and shows that some trees are completely or partly shaded by other trees in a single terrestrial laser scan (also see Olofsson and Olsson 2018). From here on, we will denote the set (or area) that is visible from the scanner by $V(X)$, where X will be explained below.

In practice, the visible area needs to be estimated from TLS data. In addition, to use the estimators considered in this paper, we assume that location and dbh of each detected tree (by a certain detection rule) have been extracted from TLS data.

2.2. Estimator

We first define the proposed estimator in terms of a marked point pattern (see, e.g., Illian et al. 2008) $X = \{(x_i, d_i, m_i)\}$ of trees that are in a forest area $W \subset \mathbb{R}^2$. Here x denotes the location of a tree (stem), d is the dbh, and m is the mark for which total per unit area is to be estimated. For example, the marks $m = \pi d^2/4$ and $m = 1$ lead to the estimation of basal area (per hectare) and tree density, respectively. Using this notation, the variable of interest, i.e., the total per unit area (T), is

$$T = \frac{1}{|W|} \sum_{(x,d,m)\in X} m\mathbf{1}(x \in W)$$

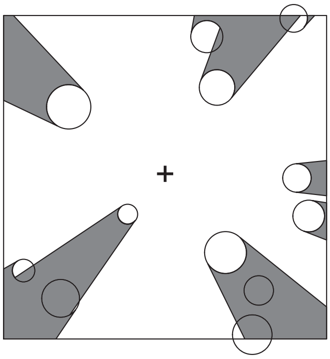
where $|W|$ is the size of the area W and $\mathbf{1}(\ldots)$ is the indicator function that takes the value 1 if the condition in the parenthesis is true and 0 otherwise.

We propose the estimator

$$\hat{T} = \sum_{(x,d,m)\in X} m\frac{v(x,d,X)}{w(d,X)} \tag{1}$$

where $v(x, d, X)$ is the detection function and

Fig. 1. Illustration of the visible area of a scan restricted to the square window. The white area represents the visible region for the scan centered at +; the grey area is not visible.



$$w(d,X) = \int_W v(y,d,X)\mathrm{d}y \tag{2}$$

is the weight. Note that the sum in eq. 1 is over all trees of the marked point pattern X in W. However, the detection function $v(x, d, X)$ determines the contribution of a tree, and this contribution is positive only for detected trees. For undetected trees, $v(x, d, X) = 0$. Further, note that the contribution $v(x, d, X)$ of a tree depends not only on the location (x) and dbh (d) of the tree, but also on the whole marked pattern of trees (X) because the other trees affect whether or not the tree is detected. For some detection functions, the weight (eq. 2) has an intuitive interpretation as the area of the set where a tree with dbh d would be detected. Examples of detection functions will be given below in section 2.2.1.

Estimator (1) is motivated by a heuristic calculation in a Poisson process setting that is explained in Appendix A. There it is found that the estimator has positive bias in the Poisson process case,

$$\mathbf{E}\hat{T} \geq \mathbf{E}T$$

There appears to be a somewhat close relationship between our estimator (1) and the Horvitz–Thompson estimator (Horvitz and Thompson 1952), which multiplies the sample values by the inverse of the inclusion probability. To see this, consider the case in which the scanner is located randomly in the forest area W, the tree locations are fixed, and the detection function v obtains only the values 0 and 1. The inclusion probability of a tree at location x and having dbh d is then

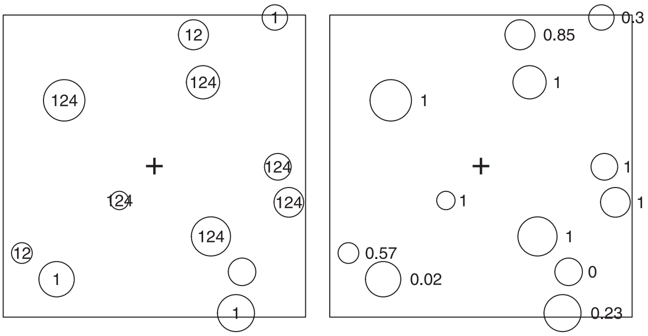
$$\frac{1}{|W|}\int_W v_s(x,d,X)\mathrm{d}s \tag{3}$$

where s is the scanner location and v_s is the detection function for a scanner located at s. The problem here is, of course, that eq. 3 depends, in general, on the trees that are not seen. The difference between the Horvitz–Thompson estimator and our estimator (1) is that in eq. 3, the integral is over the scanner location and the tree is fixed, whereas in eq. 2, the integral is over the tree location and the scanner location is fixed.

2.2.1. Detection functions

We consider four different detection functions for which it is necessary first to introduce some notation. Recall that $V(X)$ denotes the set (or area) that is visible from the scanner and Fig. 1

Fig. 2. (Left) Illustration of the detectors Visible (1), Center (2), and Complete (4) for a scan centered at +. The labels indicate which detectors detect which trees. The tree without label is not detected by any of these detectors. (Right) Illustration of the detector Proportional. The number next to the circle is the proportion that is visible.



illustrates how we define the visible region for a scan. Following Chiu et al. (2013), we use the following notation for a set A and a real number $r \geq 0$, $A_{\oplus r} = A \oplus B(o, r)$ for morphological dilation, $A_{\ominus r} = A \ominus B(o, r)$ for morphological erosion, ∂A for the boundary of A, and $l(A)$ for the length of A whenever it makes sense. Here, $B(x, r)$ is a disc centered at x and having radius r.

We study four detectors also considered in Olofsson and Olsson (2018). Figure 2 illustrates these different detection rules.

1. Visible detection: a tree is detected if any part of it is visible. Then

$$v(x, d, X) = \mathbf{1}[B(x, d/2) \cap V(X) \neq \emptyset]$$

Because this detection function corresponds to

$$v(x, d, X) = \mathbf{1}[x \in V(X)_{\oplus d/2}]$$

the weight function is simply

$$w(d, X) = |W \cap V(X)_{\oplus d/2}|$$

2. Center detection: a tree is detected if its center point x is seen. Then

$$v(x, d, X) = \mathbf{1}[x \in V(X)]$$

and

$$w(d, X) = |W \cap V(X)|$$

3. Proportional detection: this is a weighted version of the detector Visible. Trees are weighted by the proportion that is visible. Here we use the circle $\partial B(x, d/2)$ to define the visible proportion. In this case,

$$v(x, d, X) = \frac{l[\partial B(x, d/2) \cap V(X)]}{l[\partial B(0, d/2)]}$$

It turns out that if the scanner is not too close to the boundary of W, $V(X)_{\oplus d/2} \subset W$, the weight function has a simple form, namely

$$w(d, X) = |W \cap V(X)|$$

which follows from noting that

$$\begin{aligned}
&\int_W l[\partial B(x, r) \cap A]\mathrm{d}x \\
&= \int_{\mathbb{R}^2} l[\partial B(x, r) \cap A]\mathrm{d}x \\
&= \int_{\mathbb{R}^2}\int_{\partial B(0,r)} \mathbf{1}(x + y \in A)\mathrm{d}l(y)\mathrm{d}x \\
&= \int_{\partial B(0,r)} |A|\mathrm{d}l(y) \\
&= 2\pi r|A|
\end{aligned}$$

4. Complete detection: a tree is detected if it is completely visible. Then

$$v(x, d, X) = \mathbf{1}[B(x, d/2) \subset V(X)]$$

and

$$w(d, X) = |W \cap V(X)_{\ominus d/2}|$$

Detectors 1–4 were also considered in Olofsson and Olsson (2018) in the estimator

$$\hat{T}' = \sum_{(x,d,m)\in X} m\frac{v(x, d, X)}{|V(X)|} \tag{4}$$

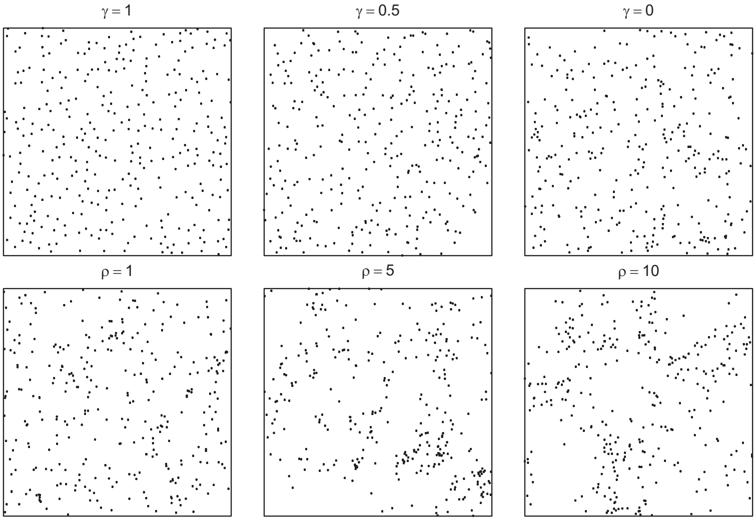
By a simulation study, Olofsson and Olsson (2018) concluded that estimator $\hat{T}'$ was approximately unbiased for the detectors Center and Proportional, whereas it was clearly biased for the detectors Visible and Complete. We note that estimator (4) coincides with our estimator in the case of the detectors Center and Proportional (up to how the stem visibility ratio is defined) in the case in which $V(X) \subset W$.

2.3. Data for the simulation experiment

Our simulation study had three parts to study the behaviour of estimator (1): (*i*) the case of Poisson-distributed stem positions and equal dbhs for all stems; (*ii*) simulated spatial patterns of the trees where the stem coordinates exhibit different degrees of regularity or clustering and independent stem dbhs; and (*iii*) empirical data. In case *i*, we also included comparison with estimator (4). Sections 2.3.1 and 2.3.2 below describe the Poisson, regular, and clustered models. Section 2.3.3 presents the empirical data.

For the simulated spatial patterns (*i* and *ii*), the simulation window was set to $W = [0, 40\text{ m}] \times [0, 40\text{ m}]$. The parameter values of

Fig. 3. (Top) Simulated patterns of Strauss processes with interaction parameters $\gamma = 0$, 0.5, and 1. (Bottom) Simulated patterns of log Gaussian Cox processes with correlation ranges $\rho = 1$, 5, and 10.



the models regarding tree density and dbh were selected to be realistic for Finnish forests.

2.3.1. *Poisson forests*

Poisson-distributed stem positions were obtained by first simulating the number of stems n from the Poisson distribution with mean $\lambda|W|$ and then generating the locations of the n points uniformly in W. We considered the properties of estimators (1) and (4) by simulation for varying λ from 150 to 2500 trees per hectare. In each simulation, a fixed dbh was used. We used the following values of dbh: 5, 10, 20, and 45 cm.

2.3.2. *Regular and clustered forests with independent stem dbhs*

The model for regular forests was a Strauss process with an interaction radius of 2 m and an interaction parameter γ controlling the strength of interaction (see, e.g., Illian et al. 2008). This process is a model for spatial inhibition, ranging from a strong "hard core" inhibition ($\gamma = 1$) to a completely random pattern ($\gamma = 0$). We fixed the number of trees to $n = 160$ in our simulation window corresponding to 1000 trees per hectare. Figure 3 (top row) shows three examples of simulated patterns with $\gamma = 0$, 0.5, and 1.

The model for clustered forests was the log Gaussian Cox process (Møller et al. 1998; Illian et al. 2008) in which the density of the points is given by $\exp[Z(s)]$, $s \in \mathbb{R}^2$, where Z is an underlying Gaussian process. We used the Matérn covariance function as the covariance function of the Gaussian process. The smoothness and variance parameters were fixed to 2 and 1, respectively. To obtain different degrees of clustering, we varied the correlation range ρ of the Matérn covariance function between 1 and 10 m. The correlation range was defined as the distance where correlation drops to approximately 0.05. In our simulations, we fixed the number of trees to the same $n = 160$ in W as used in the regular process and simulated the tree locations according to $\exp[Z(s)]$ in W. Figure 3 (bottom row) shows three examples of simulated patterns with $\rho = 1$, 5, and 10.

In both cases, four different Weibull distributions were used for dbh. We selected basal areas 3, 12, 20, and 35 $m^2 \cdot ha^{-1}$ and chose thereafter the parameters of the Weibull distribution using the method of Siipilehto and Mehtätalo (2013). The mean dbhs corresponding to the selected basal areas were 6, 12, 15, and 21 cm.

The simulations were performed using R (R Core Team 2017) libraries rstrauss (https://github.com/antiphon/rstrauss), spatstat (Baddeley et al. 2015), and RandomFields (Schlather et al. 2015).

2.3.3. *Empirical data*

The empirical data consisted of observations and measurements for 30 field plots located in southern Finland with a size of approximately 32 m × 32 m, previously used by Tomppo et al. (2017). The locations of these plots were selected subjectively from forests in which estimation of stem volume and other forest resource characteristics with airborne laser scanning data usually leads to large root-mean-square errors (RMSEs) (Tomppo et al. 2017). Each plot was entirely within one forest stand. The plots were distributed into development classes as follows: one in an advanced seedling stand, 17 in young thinning stands, 10 in advanced thinning stands, and 2 in mature stands. The minimum dbh of a measured tree was 2.5 cm. The maximum observed dbh was 49.0 cm. Tree densities in the empirical data ranged from 737 to 8231 stems per hectare. The range of plot basal areas was from 10 to 40 $m^2 \cdot ha^{-1}$.

The field plots in the empirical data had different degrees of regularity or clustering. Figure 4 shows examples of four patterns with obvious differences in the stem coordinate distributions. The degree of clustering or regularity was described using $L_{\max} = L(\hat{r}) - \hat{r}$, where $\hat{r}$ maximizes $|L(r) - r|$ on the chosen interval of distances r (which was chosen here to be from 0 to approximately one-fourth of the side of the observation window) and L is the square root transformation of Ripley's K function (Besag 1977). Positive values of $L_{\max}$ indicate clustering of the corresponding pattern, whereas negative values are an indication of regularity.

2.4. Simulation study

In the simulation experiment, we positioned the TLS scanner in the center of the simulation window W but found a new location for the scanner if its location was closer than 1 m away from the

Fig. 4. Four field plots from the empirical data. The patterns from left to right represent a regular, random, and clustered pattern of tree locations and a pattern with small-scale clustering and larger scale regularity.

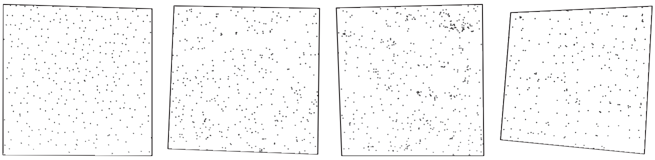
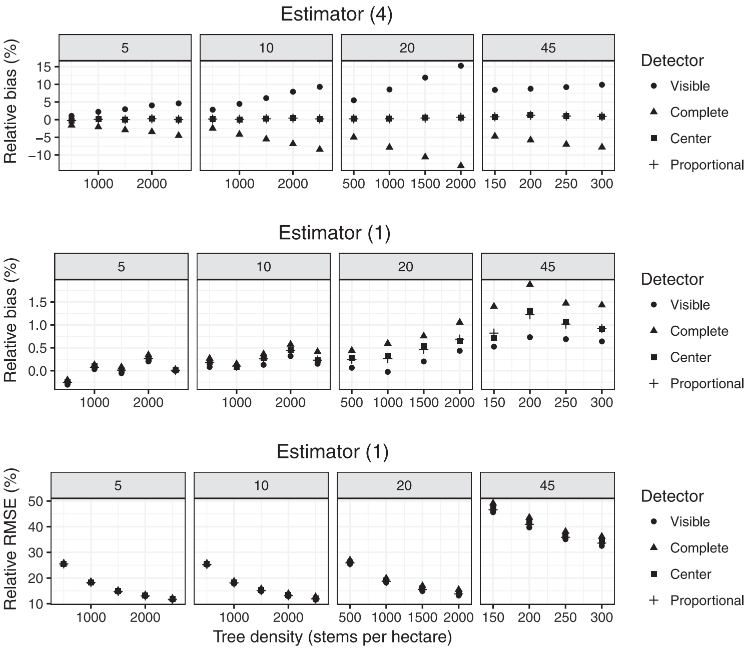


Fig. 5. Relative biases of estimator (4) (top row), and relative biases (middle row) and relative RMSEs (bottom row) of estimator (1) with different detectors for the Poisson forest with different tree densities (given on the *x* axis) and with dbhs of 5, 10, 20, and 45 cm (given in the heading).



nearest tree. The search operated on a grid with 1 m spacing taking the nearest suitable grid location.

We generated 10 000 replicates of each considered Poisson forest model and 1000 replicates of each regular and clustered model. To obtain replicates for the empirical data, we simulated approximately 1000 TLS scans on each plot. The scan locations were at first placed on a regular grid, locations outside the plot windows were discarded, and then the remaining locations were moved as in the previous cases if the scan location was too close to a tree.

We estimated the tree density using estimator (1) and each of the detectors (Visible, Center, Proportional, and Complete) specified in section 2.2.1. For the empirical data, we also estimated the basal area. We set a maximum radius of 10 m for the TLS plot.

For the Poisson forests, we compared the estimates to the intensity λ of the Poisson process, and in the regular and clustered cases, we used the fixed tree density of 1000 trees per hectare as the reference value. The relative bias was calculated as

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\hat{Y}_i - Y}{Y}$$

where $\hat{Y}_i$ is the estimated value for replicate i, Y is the reference value, and n is the number of replicates. The relative RMSE was calculated as

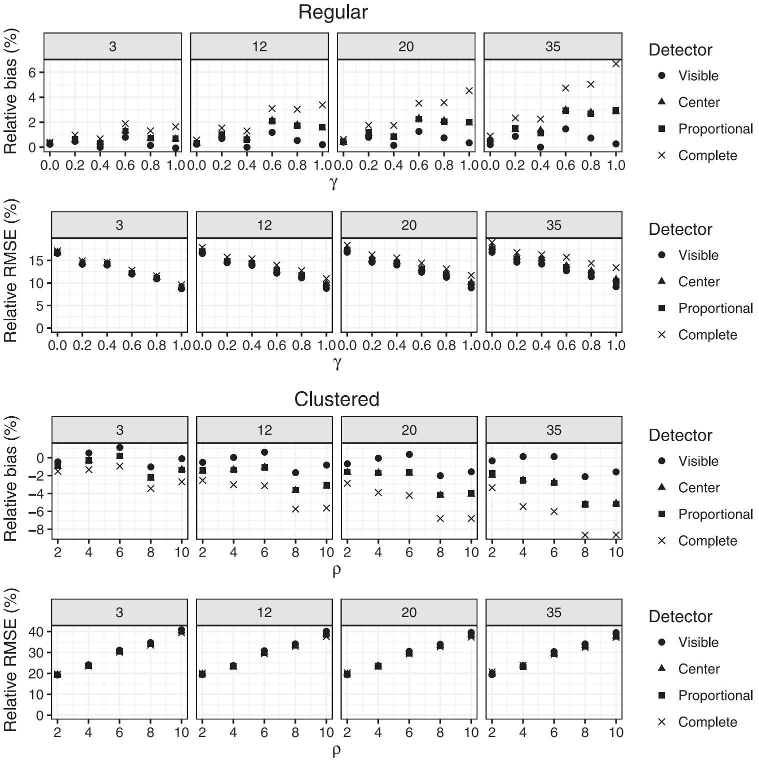
$$\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\frac{\hat{Y}_i - Y}{Y}\right)^2}$$

3. Results

3.1. Poisson forests

Figure 5 shows the relative biases of estimator (4) and the relative biases and relative RMSEs of estimator (1) for the tree density for the four different detectors. The Monte Carlo errors for the relative biases of estimator (1) were less than 0.5 percentage points. First, by comparing the top and middle rows of Fig. 5, we see that the biases of the proposed estimator were clearly smaller than the biases of estimator (4) for the detectors Visible and Complete. Second, the detectors generally led to different biases also for estimator (1), even though the differences were smaller than when estimator (4) was used. The detector Visible led to the lowest

Fig. 6. Relative biases and relative RMSEs for tree density in regular forests (top) and clustered forests (bottom) with different interaction parameters γ, correlation ranges ρ, and basal areas 3, 12, 20, and 35 m²·ha⁻¹. The legend gives the different detectors considered.



biases of estimator (1), whereas the detector Complete led to the largest biases. The relative bias also tended to increase along increasing tree density and dbh for all detectors.

Regarding estimator (1), the detector Visible led also to the lowest relative RMSEs, although the differences of the different detectors in RMSEs were rather small (see Fig. 5, bottom). It should be noted that the relative RMSEs were largest for low tree densities, but the (non-relative) RMSEs increased with decreasing visibility, i.e., increasing tree density. All in all, the biases were less than 2% for all detectors when estimator (1) was used, and the contribution of the bias to the RMSE was rather small.

3.2. Regular and clustered forests with independent dbhs

Figure 6 shows the relative biases and RMSEs of estimator (1) for the tree density for the regular and clustered cases, respectively. The Monte Carlo standard errors for the biases were less than 0.6 percentage points for the regular cases and less than 1.3 percentage points for the clustered cases.

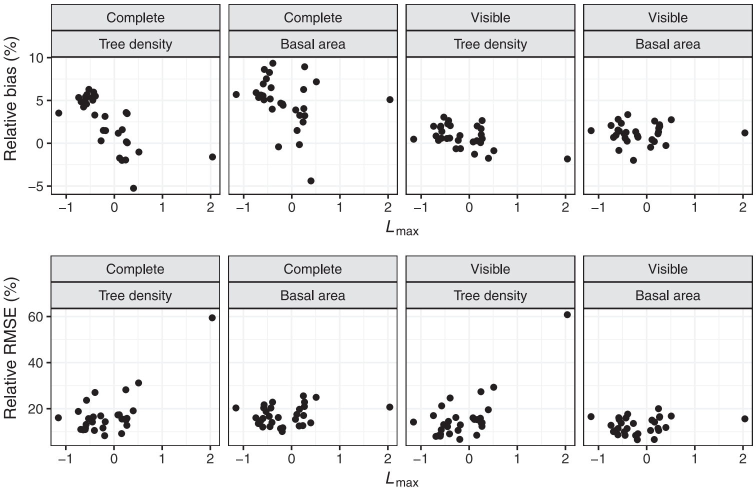
For the regular patterns of tree locations, the biases were positive and tended to increase with the degree of regularity (Fig. 6, top row). On the other hand, for the clustered tree locations, the biases were negative and increased in absolute value with the degree of clustering (Fig. 6, third row). In both cases, the absolute values of relative bias increased with increasing basal area. The detectors had rather different biases. The detector Visible that detects the largest number of the trees had the smallest biases, as in the case of Poisson forests. Considering the four detectors, the absolute values of relative bias were less than 7% for the regular patterns and less than 9% for the clustered patterns. On the other hand, for the detector Visible, the absolute values of relative bias were all less than 2.5%. (In the case of Poisson patterns, they were less than 1% for the detector Visible.) Thus, the biases were somewhat larger for the regular and clustered patterns than for the Poisson patterns.

For the regular forests, the relative RMSEs decreased with increasing regularity (Fig. 6, second row), whereas for the clustered forests, the relative RMSEs increased clearly with increasing degree of clustering (Fig. 6, bottom row). The differences in relative RMSEs between different basal areas were small.

3.3. Empirical data

Figure 7 shows the relative biases and relative RMSEs for tree density and basal area for the empirical data with the detectors Visible and Complete as a function of L_{max}. The performance of the detectors Proportional and Center were between the detectors Visible and Complete, as in our previous experiments. Similarly as for the regular and clustered simulated forests, we observed positive biases for the regular patterns of the empirical data and negative biases for the clustered patterns when the detector Complete was used. For the detector Visible, however, the pattern was not that clear and the absolute values of bias were less than 3%. The relative RMSEs were rather similar for the two detectors and did not clearly depend on L_{max}, except for the very clustered pattern (with $L_{max} \approx 2$), where the RMSEs for the tree density were clearly larger than for the other patterns. This clustered pattern

Fig. 7. Relative biases (top row) and relative RMSEs (bottom row) in the empirical data for the Visible and Complete detectors with respect to $L_{\max}$, which is a measure of regularity (negative values) and clustering (positive values). See section 2.3.3 for a more detailed explanation of $L_{\max}$.



was one of the advanced thinning stands, which by visual investigation was somewhat inhomogeneous.

4. Discussion

We proposed the new estimator (1) for correcting the effect of nondetection of the trees in the single-scan setup of TLS. This estimator allows for different detectors, e.g., those considered by Olofsson and Olsson (2018). A key observation is that, instead of scaling by the visible area of the scanner in the estimator for each detector, the scaling must account for the chosen detector. See Fig. 5 for a comparison of the biases of the corrected and uncorrected estimators in a Poisson forest.

In our simulation study with simulated and empirical forests, the new estimator worked adequately. It appeared that it would be best to use the detector Visible, i.e., to consider a tree detected if any part of it is seen. Our study is, however, based on the theoretical setting where the tree dbh is always obtained exactly, no matter how small a part of the tree is seen. In practice, the measurement accuracy is likely to decrease as the distance to the measured tree grows (see, e.g., Lovell et al. 2011). The accuracy may also depend on the proportion of the visible part of the tree. Thus, other detectors may be preferred over the detector Visible in practice. We considered other detectors, and estimator (1) allows for further ones as well, as long as one is able to calculate the corresponding weight (eq. 2).

We only considered a few detection functions to keep the simulation study reasonably sized. We think that the chosen detectors illustrate the variation in performance quite well. As pointed out also by Olofsson and Olsson (2018), small trees could be more difficult to detect than large trees. In that case, it might be reasonable to consider the detector

$$v(x, d, X) = \mathbf{1}\left\{\frac{l[\partial B(x, d/2) \cap V(X)]}{\pi d} \geq f(d)\right\}$$

where f is a function that gives the proportion of the tree that is required for detection. So far, though, we have not been able to compute a good enough approximation for the weight (eq. 2) corresponding to this detector.

In principle, it appears possible to use something other than the dbh to determine the nondetection caused by a tree. One could use the tree shape or even a three-dimensional tree model, for example. An interesting question would be whether or not that could improve estimation.

We believe that obstacles such as stones or understorey vegetation could be handled in the estimation by including them in the scanner visible area $V(X)$. The proposed estimators should work if the obstacles are independent of the trees (suggested by our heuristic calculations of Appendix A). Unfortunately, in practice, the branches of trees are probably affecting the visibility of trees the most, and they would need a different treatment, definitely being strongly dependent on the locations of trees.

Another point worth studying would be to include the measurement accuracy in the estimator. The dbhs (or other properties) of more distant trees may be harder to estimate from a TLS point cloud than those of trees close to the scanner; however, it is not immediately clear how to extend the proposed estimator to handle that case.

The final question is, of course, how to apply the results in practice. Our recommendation is that the applied detection algorithm should be studied and an approximate detection function chosen. An estimator taking the used detection algorithm into account could have smaller bias than the simple visible area based estimator.

Acknowledgements

The authors have been financially supported by the Academy of Finland (Project Nos. 304212, 295100, and 306875). The authors thank Erkki Tomppo for providing the empirical data, Lauri Mehtätalo and Kasper Kansanen for useful discussions on the topic, and the anonymous reviewer, whose comments and criticism helped clarify the presentation of the paper.

References

Astrup, R., Ducey, M.J., Granhus, A., Ritter, T., and von Lüpke, N. 2014. Approaches for estimating stand-level volume using terrestrial laser scanning in a single-scan mode. Can. J. For. Res. **44**(6): 666–676. doi:10.1139/cjfr-2013-0535.

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

Besag, J.E. 1977. Discussion on 'Modelling spatial patterns' by B.D. Ripley. J. R. Stat. Soc., Ser. B, **39**(2): 193–195.

Chiu, S.N., Stoyan, D., Kendall, W.S., and Mecke, J. 2013. Stochastic geometry and its applications. 3rd ed. Wiley.

Ducey, M.J., and Astrup, R. 2013. Adjusting for nondetection in forest inventories derived from terrestrial laser scanning. Can. J. Remote Sens. **39**(5): 410–425.

Horvitz, D.G., and Thompson, D.J. 1952. A generalization of sampling without replacement from a finite universe. J. Am. Stat. Assoc. **47**(260): 663–685. doi:10.1080/01621459.1952.10483446.

Illian, J., Penttinen, A., Stoyan, H., and Stoyan, D. 2008. Statistical analysis and modelling of spatial point patterns. 1st ed. John Wiley & Sons.

Jupp, D.L., Culvenor, D., Lovell, J.L., and Newnham, G. 2005. Evaluation and validation of canopy laser radar (lidar) systems for native and plantation forest inventory. CSIRO Mar. Atmos. Res. Tech. Rep. 020. CSIRO, Canberra, Australia.

Liang, X., Litkey, P., Hyyppä, J., Kaartinen, H., Vastaranta, M., and Holopainen, M. 2012. Automatic stem mapping using single-scan terrestrial laser scanning. IEEE Trans. Geosci. Remote Sens. **50**(2): 661–670. doi:10.1109/TGRS.2011.2161613.

Liang, X., Kankare, V., Hyyppä, J., Wang, Y., Kukko, A., Haggrén, H., Yu, X., Kaartinen, H., Jaakkola, A., Guan, F., Holopainen, M., and Vastaranta, M. 2016. Terrestrial laser scanning in forest inventories. ISPRS J. Photogramm. Remote Sens. **115**: 63–77. doi:10.1016/j.isprsjprs.2016.01.006.

Lovell, J.L., Jupp, D.L., Newnham, G.J., and Culvenor, D.S. 2011. Measuring tree stem diameters using intensity profiles from ground-based scanning lidar from a fixed viewpoint. ISPRS J. Photogramm. Remote Sens. **66**(1): 46–55. doi:10.1016/j.isprsjprs.2010.08.006.

Møller, J., Syversveen, A.R., and Waagepetersen, R.P. 1998. Log Gaussian Cox processes. Scand. J. Stat. **25**(3): 451–482. doi:10.1111/1467-9469.00115.

Olofsson, K., and Holmgren, J. 2016. Single tree stem profile detection using terrestrial laser scanner data, flatness saliency features and curvature properties. Forests, **7**(9): 207. doi:10.3390/f7090207.

Olofsson, K., and Olsson, H. 2018. Estimating tree stem density and diameter distribution in single-scan terrestrial laser measurements of field plots: a simulation study. Scand. J. For. Res. **33**(4): 365–377. doi:10.1080/02827581.2017.1368698.

Olofsson, K., Holmgren, J., and Olsson, H. 2014. Tree stem and height measurements using terrestrial laser scanning and the ransac algorithm. Remote Sens. **6**(5): 4323–4344. doi:10.3390/rs6054323.

R Core Team. 2017. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Schlather, M., Malinowski, A., Menck, P.J., Oesting, M., and Strokorb, K. 2015. Analysis, simulation and prediction of multivariate random fields with package RandomFields. J. Stat. Softw. **63**(8): 1–25. doi:10.18637/jss.v063.i08.

Seidel, D., and Ammer, C. 2014. Efficient measurements of basal area in short rotation forests based on terrestrial laser scanning under special consideration of shadowing. iForest, **7**(4): 227–232. doi:10.3832/ifor1084-007.

Siipilehto, J., and Mehtätalo, L. 2013. Parameter recovery vs. parameter prediction for the Weibull distribution validated for Scots pine stands in Finland. Silva Fenn. **47**(4): 1057. doi:10.14214/sf.1057.

Strahler, A.H., Jupp, D.L., Woodcock, C.E., Schaaf, C.B., Yao, T., Zhao, F., Yang, X., Lovell, J., Culvenor, D., Newnham, G., Ni-Miester, W., and Boykin-Morris, W. 2008. Retrieval of forest structural parameters using a ground-based lidar instrument (Echidna®). Can. J. Remote Sens. **34**(Suppl. 2): S426–S440. doi:10.5589/m08-046.

Tomppo, E., Kuusinen, N., Mäkisara, K., Katila, M., and McRoberts, R.E. 2017. Effects of field plot configurations on the uncertainties of ALS-assisted forest resource estimates. Scand. J. For. Res. **32**(6): 488–500. doi:10.1080/02827581.2016.1259425.

Appendix A. Motivation for the estimator

Estimator (1) is motivated by a heuristic calculation in a Poisson process setting. This setting means that the tree locations *x* follow the Poisson process with expected tree density $\lambda(s)$, $s \in \mathbb{R}^2$, and the dbhs *d* are independent. The additional marks *m* can depend on *d* (or be defined in terms of *d*), but mutually, the *m*s are assumed to be independent as well. We denote the joint mark distribution of *d* and *m* by *F*. Under this setting, we can calculate the expectation of estimator (1) by applying the Mecke theorem (Schneider and Weil 2008, theorem 3.2.5):

$$\begin{aligned}\mathbf{E}\widehat{T} &= \mathbf{E}\sum_{(x,d,m)\in X} m\frac{v(x,d,X)}{w(d,X)}\\ &= \int \mathbf{E}m\frac{v[x,d,X\cup\{(x,d,m)\}]}{w[d,X\cup\{(x,d,m)\}]}\mathrm{d}\lambda(x)\mathrm{d}F(d,m)\\ &= \int \mathbf{E}m\frac{v(x,d,X)}{w[d,X\cup\{(x,d,m)\}]}\mathrm{d}\lambda(x)\mathrm{d}F(d,m)\end{aligned}$$

where in the last equation, we assumed that

$$v[x,d,X\cup\{(x,d,m)\}] = v(x,d,X)$$

which corresponds to assuming that a tree cannot block itself.

Unfortunately, it is not possible to calculate the expectation, but we can obtain an approximation by replacing $w[d, X\cup\{(x,d,m)\}]$ by $w(d,X)$, namely then

$$\begin{aligned}\mathbf{E}\hat{T} &\approx \int \mathbf{E}m\frac{v(x,d,X)}{w(d,X)}\mathrm{d}\lambda(x)\mathrm{d}F(d,m)\\ &= \mathbf{E}\int m\frac{\int v(x,d,X)\mathrm{d}\lambda(x)}{w(d,X)}\mathrm{d}F(d,m)\end{aligned}$$

Now if we define the weight (eq. 2) and assume that the Poisson process is homogeneous, i.e., $\lambda(s) = \lambda$ for all $s \in \mathbb{R}^2$ (the expected tree density is constant), we arrive at

$$\mathbf{E}\hat{T} \approx \int m\lambda\mathrm{d}F(d,m) = \lambda\mathbf{E}m$$

Although we cannot say anything about the size of the error made by the approximation, it appears to give valuable insight, which is shown in the simulation study.

For many detection functions $v(x,d,X\cup\{y\}) \le v(x,d,X)$, that is, adding a tree decreases visibility. For the weight function, this implies that $w[d,X\cup\{(x,d,m)\}] \le w(d,X)$. Then $\mathbf{E}\hat{T} \ge \lambda\mathbf{E}m$.

Reference

Schneider, R., and Weil, W. 2008. Stochastic and integral geometry. Springer-Verlag, Berlin, Heidelberg. doi:10.1007/978-3-540-78859-1.

II

POINT PROCESS MODELS FOR SWEAT GLAND ACTIVATION OBSERVED WITH NOISE

by

Kuronen, M., Myllymäki, M., Loavenbruck, A., and Särkkä, A.

Statistics in Medicine, 40:2055–2072, 2021.

doi:10.1002/sim.8891

Received: 27 April 2020 | Revised: 7 November 2020 | Accepted: 6 January 2021

DOI: 10.1002/sim.8891

RESEARCH ARTICLE

Statistics in Medicine WILEY

Point process models for sweat gland activation observed with noise

Mikko Kuronen[1] | **Mari Myllymäki**[1] | **Adam Loavenbruck**[2] | **Aila Särkkä**[3]

[1]Natural Resources Institute Finland (LUKE), Helsinki, Finland

[2]Department of Neurology, Kennedy Laboratory, University of Minnesota, Minneapolis, Minnesota

[3]Department of Mathematical Sciences, Chalmers University of Technology and the University of Gothenburg, Gothenburg, Sweden

Correspondence
Mikko Kuronen, Natural Resources Institute Finland (LUKE), Helsinki, Finland.
Email: mikko.p.kuronen@gmail.com

Funding information
Academy of Finland, Grant/Award Numbers: 295100, 306875, 32711; Swedish Foundation for Strategic Research, Grant/Award Number: SSF AM13-0066; Swedish Research Council, Grant/Award Number: VR 2013-5212

The aim of this article is to construct spatial models for the activation of sweat glands for healthy subjects and subjects suffering from peripheral neuropathy by using videos of sweating recorded from the subjects. The sweat patterns are regarded as realizations of spatial point processes and two point process models for the sweat gland activation and two methods for inference are proposed. Several image analysis steps are needed to extract the point patterns from the videos and some incorrectly identified sweat gland locations may be present in the data. To take into account the errors, we either include an error term in the point process model or use an estimation procedure that is robust with respect to the errors.

KEYWORDS

Bayesian inference, independent thinning, peripheral neuropathy, point pattern, sequential point process, soft-core inhibition

1 | INTRODUCTION

Sweating is critical in human evolution in maintaining ability to thermoregulate in a wide range of climates and activity levels. Neurologic control, headquartered in the hypothalamis, is therefore tightly regulated and concordantly disruptive in pathologic states such as peripheral neuropathy. Assessment of sudomotor (sweat) function has long been used in clinical and research settings for detection of neurologic dysfunction.[1,2] Minor's starch iodine test was originally described in 1928,[3] where tincture of iodine was applied to the skin, air dried, and then powdered with corn starch. Sweating is stimulated with increasing room temperature or use of pilocarpine. As sweat flows from pores, iodine is diluted and the solution absorbed by the starch powder, turning dark black from yellow. Normally the entire skin surface should be able to sweat in response to sufficient stimuli, and absence of sweating in an area of the body is indicative of loss of neurologic function in that area.

Peripheral neuropathy is a disease state of peripheral nerves, the segment of the nervous system which extends from the brain and spinal cord to various targets in the body, such as muscles, sensory receptors, and autonomically controlled organs. Peripheral neuropathy occurs in etiologically diverse conditions which cause damage or dysgenesis of peripheral nerves. The most common causes include diabetes, toxicity such as in chemotherapy and excessive alcohol consumption, and vitamin deficiencies.[4] The resulting nerve damage causes various combinations of muscle weakness, pain, numbness, and autonomic dysfunction.

Autonomic nerves are often the earliest to be affected in peripheral neuropathy,[5-7] and penetrate all parts of the body, including digestive tract, liver, kidneys, bladder, genitals, lungs, pupils, heart, and skin. Skin is the largest organ in the body, and contains a vast network of the distal ends of sensory and autonomic nerves over the entire body surface. These distal ends of nerves are especially susceptible to systemic disease. Because sweating is neutrally controlled and modulated, and can be measured at the skin surface, it can be used to reflect alterations in the underlying nerves.

Currently, there are several tests used in clinical practice to evaluate sudomotor function.[8-10] Thermoregulatory sweat testing[11] measures percentage of body surface area sweating elicited by a heated, humidified sauna. Sweat imprints and silastic molds[12,13] measure the density and distribution of activated sweat glands in an area of skin. Quantitative sudomotor axon reflex testing (QSART) is likely the most widely used autonomic test of sweating,[14,15] comparing against robust normative databases the total volume of sweat produced by 1 cm^2 areas of skin at standardized sites.

The sensitive sweat test (SST) enhanced Minor's starch iodine test with closeup time lapse imaging, and software analysis.[4,16-18] The critical feature of the test is a rigid, transparent video screen which limits sweat to an essentially two-dimensional space. As sweat exits ducts, it dilutes the iodine painted on the skin onto starch coated plastic film. The imaged result is a field of sharply demarcated, dark sweat spots on a white background, expanding centripetally around the opening of each duct (Figure 1). The area of each spot is therefore a measurement of the volume of sweat produced by each gland. Sub-nanoliter volumes of liquid were measured by pipette and shown to create reproducible sweat spot areas. Similarly, tracking the increase in sweat spots' areas between timelapse frames measures the rate of sweating from each duct at the nanoliter level. Of note, blackened areas of film do not return to white during the test—sweat spots can only enlarge. The videos therefore provide several measurable physiologic data points, including coordinates and relative locations of all sweat spots, the second by second volumes of sweat (nanoliters) and flow rate of each sweat gland (nL/minute), total number of activated sweat glands, density of activated sweat glands (glands/cm^2), total sweat volume (nL), and total sweat rate (nL/minute).

Using the dynamic sweat test, a significant reduction of sweating was observed in diabetic subjects in the distal leg but not in forearm.[16] The study included measurements taken from the forearm of 14 diabetic subjects and 14 age- and sex-matched healthy controls and from distal leg of seven diabetes patients and seven controls. In a larger study,[4] 178 healthy controls and 20 neuropathy subjects were tested, most of them at the hand, thigh, calf, and foot, some only at calf and foot, and it was concluded that neuropathy subjects had lower sweat rates per sweat gland, lower total amount of sweat, and lower sweat gland density. It was also observed visually that the sweat patterns of the diabetic subjects were less regular than the healthy patterns.[16] To quantify this visual observation, the spatial sweat patterns that the videos provide should be investigated more carefully. Spatial analysis can provide more detailed information on the sweat patterns and tools for revealing additional abnormalities that may appear when the sweat rate and the total amount of sweat are still within normal ranges. However, up to now, no spatial analysis of the sweat patterns to quantify this observation has been performed.

In this article, the locations of sweat spots or glands are regarded as realizations of spatial point processes. Our main emphasis is to develop suitable methodology for analyzing the spatial structure of the sweat gland patterns and activation extracted from video sequences. As the data are non-standard in the point process literature, some special treatment is needed.

To extract the coordinates of the individual sweat spots, that is, the points of the point patterns, from the videos (see Figure 1), several image analysis steps are needed. As a non-standard step, we apply an algorithm based on the detection of a change point to the sequence of gray scale values in each pixel. This pixel-by-pixel approach suits to the video sequences, where the sweat does not dry once it has appeared, much better than going through the videos frame by frame, because the sweat gland locations are best detected at times where sweat first appears. However, even though we perform careful analysis of the videos, there are some spots that are incorrectly recorded as two spots due to, for example, wrinkles in the skin. It is not straightforward to remove these errors automatically and doing it manually can be very time consuming. Therefore, they need to be taken into account in the analysis of the point patterns.

Some studies of point patterns observed with errors or noise can be found in the literature. For example, in the area of minefield detection, one first detects a minefield and then classifies each observed point in the minefield either as mine or as noise. The observation window is typically divided into two parts, the minefield as a region with a higher intensity containing both mines and noise and a low intensity area containing only noise.[19] The points can then be classified in a

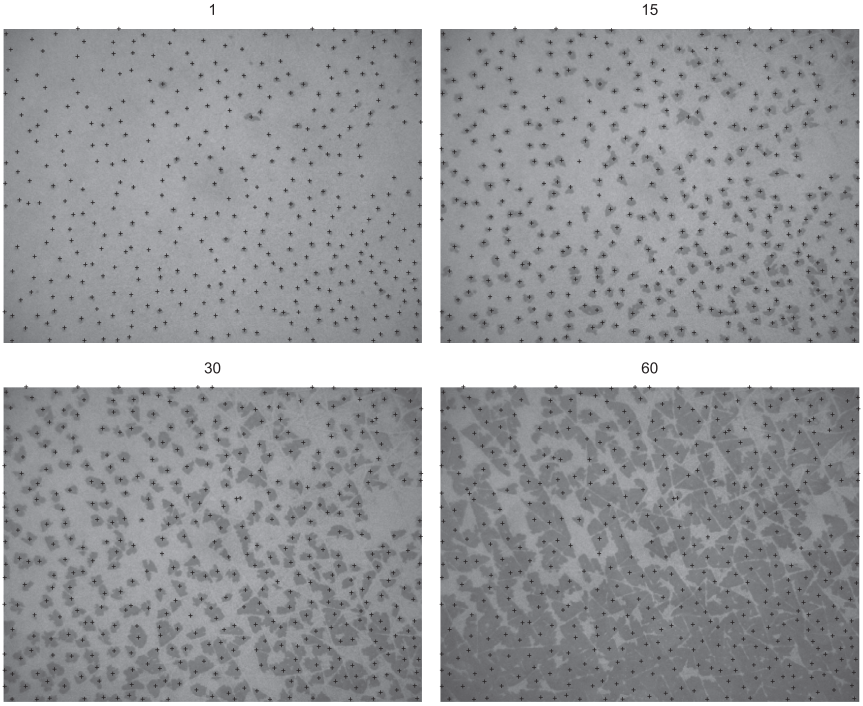


FIGURE 1 A sequence of snapshots (1 second (top left), 15 seconds (top right), 30 seconds (bottom left), and 60 seconds (bottom right)) of one control subject with extracted gland locations (+)

Bayesian set-up where a posterior probability for each point being a mine is derived.[20,21] In a similar Bayesian framework, classification of points of a superposition of a Strauss process and Poisson noise has been considered.[22,23] A likelihood of an imperfect observation given the true point process, where the imperfect observation can be due to random thinning, displacement, censoring of the displaced points, or superposition of additional points[24] and Bayesian analysis for similar data[25] can also be found in the literature. Furthermore, a Bayesian framework for estimating the intensity of a noisy point process, where the noise is either due to points within the sampling window but regarded as being outside and/or points outside the window which were incorrectly regarded as points inside the window, is available[26] as well as descriptive statistics for noisy spatial point processes, where the noise is perturbation of points.[27]

Here, we suggest two different ways to model the activation of sweat glands and to take noise into account in the analysis, by either including an error term in the model or using an estimation procedure that is robust with respect to the errors. We pay special attention to incorrectly recorded close pairs of points since they can cause problems for the analysis of regular point patterns such as our data.

We propose two models for the sweat gland patterns that are different in their nature. In the first model, the activation of individual sweat glands is modeled by a sequential point process, where sweat spots appear conditionally on the pattern so far. The other model is more physiologically motivated, a generative point process model, where the activated sweat gland pattern is modeled as a thinning of the underlying true (unobserved) sweat gland pattern which is modeled first. While the likelihood of the sequential model is tractable, it has been considered computationally costly to evaluate.[28] Here, we propose an efficient way to perform traditional likelihood-based inference for a certain type of sequential models, which makes also likelihood-based Bayesian inference feasible. The likelihood of the generative model is not easily tractable and, therefore, we employ approximate Bayesian computation (ABC) to estimate the model parameters. When some noise points are present, the sequential model is replaced by a mixture model having the sequential point process and an error point process as its components. In the generative model, the summary statistics in the ABC approach are chosen such that they are robust with respect to the errors.

The rest of this article is organized as follows. We first describe the extraction of the points from the videos and the preliminary analysis of the data in Section 2. Then, we present the sequential and generative models together with a description of the inference methods in Sections 3 and 4, respectively. Further, the methodology is illustrated by analyzing a set of video sequences taken from the right foot of 15 subjects, either controls or subjects with suspected or diagnosed neuropathy. The models are fitted separately to each subject. Section 5 is left for further discussion. We carry out all computations in Julia[29] while we use R[30] mainly for plotting.

2 | DATA AND PRELIMINARY DATA ANALYSIS

2.1 | Description of data

The data have been collected by Dr. Kennedy's group at the University of Minnesota by using the dynamic sweat test they have presented.[16] First, sweating is stimulated by placing a patch with pilocarpine gels on the test site, foot or calf. Then, the test site is dried and painted with iodine solution. Finally, a camera is placed on the skin and a video is recorded at the rate of 1 frame/sec for 60 seconds. The size of the frame was 2592×1944 pixels corresponding to 17.5×13 mm^2. Videos were taken from the feet and/or calves from 121 healthy controls without known neuropathy or known risk factors for neuropathy, as well as 72 subjects who had reported having symptoms of neuropathy, 20 of whom had well characterized neuropathy (diagnosis based on neurologic examination and nerve conduction studies). Therefore, the subjects were divided into three groups: controls, subjects with suspected neuropathy (MNA), and subjects with diagnosed neuropathy (MNA Diagnosed).

In this study, we have access to five videos from the right foot from each of the three groups. Based on earlier studies, it was clear that the number of activated sweat glands is an important predictor for the condition, controls having higher density than subjects in the neuropathy groups. The five videos were selected based on the point density of the pattern so that different groups have overlapping densities. A sequence of snapshots (1 second, 15 seconds, 30 seconds, and 60 seconds) of one control subject is shown in Figure 1. Here, we study the patterns of activated sweat glands at the end of videos as realizations of spatial point processes. The complete video is needed to extract the gland locations, because these can be obtained most precisely at their first occurrence (see Section 2.2).

2.2 | Video processing with change point detection

Extracting the locations and sizes of the sweat spots required several video processing steps: transforming the video into sweat/not sweat binary video, splitting the sweat part of the video into the sweat produced by individual sweat glands, and finally extracting the point pattern of gland locations. As is commonly done in data pre-processing, some trial and error was necessary before a satisfactory result was obtained. In the following, we describe the final choices in more detail.

The first step consisted of a background correction, finding change points, and finding and applying a threshold to the change points. The background was first estimated by kernel smoothing using a Gaussian kernel with $\sigma = 100$ pixels to the first frame of the video. The smoothing bandwidth of 100 pixels, which is a bit more than the minimum interpoint distance of sweat spots, was large enough to remove the sweat spots, but small enough to keep the background fluctuations. Since the first frame had only small amount of sweat, the resulting image mostly mimicked the lighting conditions. For example, the corners of the frame were darker than the middle. Next, the grayscale values g_t of each pixel at frame t were divided by the estimated lightning intensity l of the pixel and the time series of these scaled grayscale values were considered to find the pixels that belong to the wet area. More precisely, a time series was constructed for each pixel as follows: Let $x_{-2} = x_{-1} = x_0 = 1$ and $x_t = g_t/l$ for $t = 1, \dots, T$. The change point of the time series $x_{-2}, x_{-1}, x_0, x_1, \dots, x_T$ was defined as the integer value $t \geq 1$ that minimizes $f(t) = s^2_{-2,t} + s^2_{t+1,T}$, where

$$s^2_{j,k} = \frac{1}{k-j+1}\sum_{i=j}^{k} x_i^2 - \left(\frac{1}{k-j+1}\sum_{i=j}^{k} x_i\right)^2$$

is the sample variance of $x_j, \dots, x_k$.

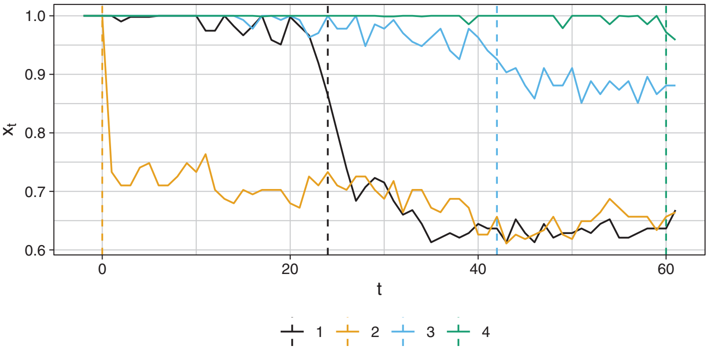


FIGURE 2 Time series for four pixels with estimated jump locations (frames) marked by dashed lines [Colour figure can be viewed at wileyonlinelibrary.com]

Testing for a change point is a well known problem in statistics.[31-34] Here, however, the problem is not of purely statistical nature. There appeared to be some jumps possibly due to the lightning conditions that were not real jumps but clear enough to be detected by a statistical change point test. In Figure 2, the series number 3 (blue curve) is an example of such a change point. We used a similar principle as in the statistical change point tests to locate the most likely change point. Instead of formally testing for the change point, we only accepted those change points where the change was large enough.

The time series and estimated change points for four pixels are shown in Figure 2. Since each pixel, even the ones that do not belong to the wet area, obtained a change point, thresholding on the difference of sample means was used to filter out small changes. A per video threshold was found by trial and error evaluating the point patterns that resulted from the whole video processing visually. By looking at the video it was quite easy to see the emerging sweat glands and how well the detected points matched them. In Figure 2, the two largest jumps, 1 and 2, passed the threshold. The third jump, although it clearly is a jump, did not pass the threshold. The resulting binary video frames were post processed with a morphological closing to fill in some small gaps.

The sweat area in the first frame was segmented into the sweat produced by individual glands. Starting with the second frame, each new sweat pixel was assigned to the closest spot in the previous frame. The distance was measured as the shortest path through the new sweat area. Several filtering steps were applied in various stages of the process to account for pixels that belonged to spots that were too small to be sweat.

Finally, we extracted a point pattern with coordinates for each gland. To obtain an ordered point pattern we used the frame of the first appearance, and for those spots that arrived at the same frame, we used spot size as a surrogate for the time, where larger ones were assumed to have appeared earlier. An example of extracted point patterns in the video can be seen in Figure 1. Figure 3 shows the final point patterns of all subjects.

2.3 | Spatial summary functions

To analyze the spatial structure of the activated sweat gland patterns, we used two different commonly used spatial summary functions, the pair correlation function g and the empty space function F. If the underlying point process is stationary and isotropic, these summary functions are functions of distance r only.

The pair correlation function g describes the (pairwise) interaction between the points.[35] It gives a scaled measure that describes how likely two points are to occur at distance r from each other. For a completely spatially random point pattern, $g(r) = 1$ for all r. Values $g(r) < 1$ indicate inhibition and values $g(r) > 1$ attraction at distance r. Thus, the pair correlation function can recognize clustering and regularity at different spatial scales, and is especially useful in describing strongly inhibitive point patterns. To estimate the pair correlation function, we used a traditional kernel smoothing method with Epanechnikov kernel and the recommended bandwidth $0.15/\sqrt{\hat{\lambda}}$, where $\hat{\lambda}$ is the intensity estimated from the point pattern.[36] Formally,

$$\hat{g}(r) = \frac{1}{2\pi r \hat{\lambda}^2} \sum_{i=1}^{n} \sum_{j=1; j \neq i}^{n} w(x_i, x_j) k(r - \|x_i - x_j\|),$$

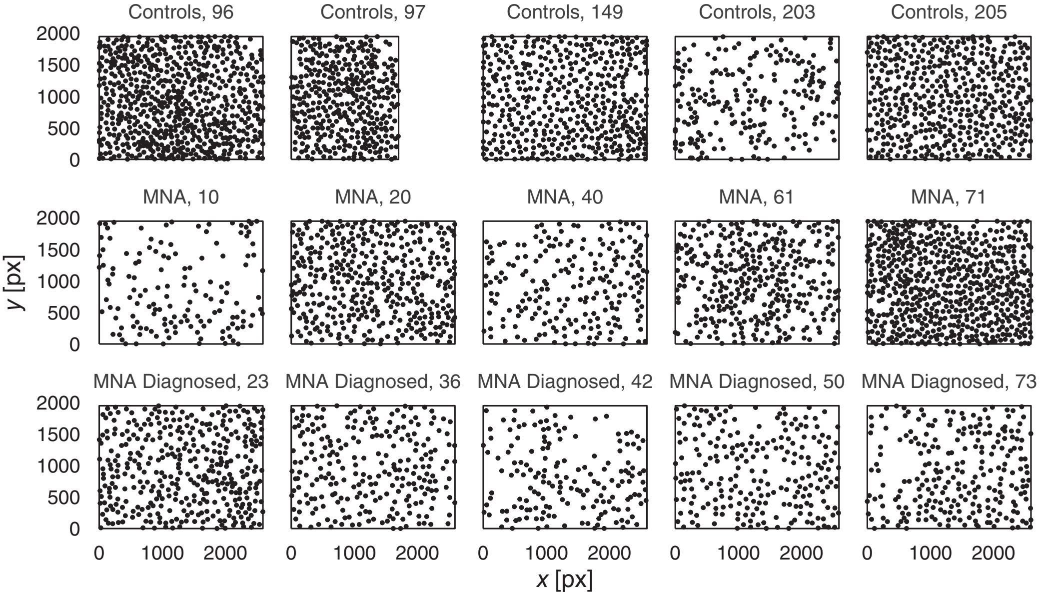


FIGURE 3 Point patterns extracted from the videos for each subject of the three groups (group and subject number shown on label)

where n is the observed number of points in the observation window, k is the chosen kernel function, $\|x_i - x_j\|$ is the distance between the points x_i and x_j, and $w(x_i, x_j)$ is an edge correction weight. Here, we used the translation edge correction.[36]

The empty space function $F(r)$ is related to the probability that an arbitrary location s in the observation window has an empty disk of radius r, $b(s, r)$, around it. It is defined as

$$F(r) = 1 - \mathbf{P}(\text{the number of } x_i \text{ in } b(s, r) = 0).$$

The empty space function can be estimated using a number of sample points in the observation window W. Typically, a grid is used. Let m be the number of sample points and d_i the distance from the sample point i to its nearest point in the point pattern. Then, an estimator for the empty space function is

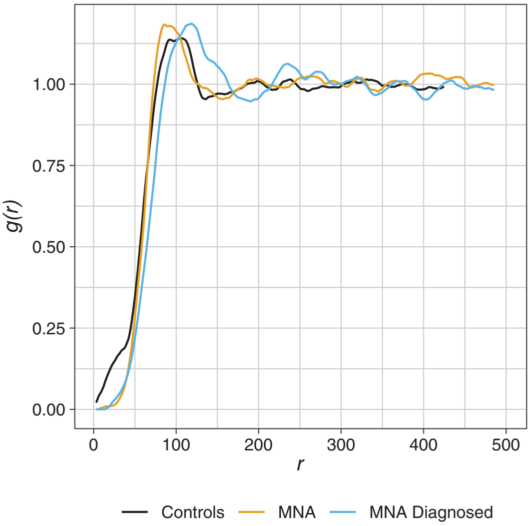
$$\hat{F}(r) = \frac{\text{the number of } d_i < r}{m}.$$

This estimator is hampered with edge effects since we cannot observe if a disk close to the boundary of the observation window would have any points of the process outside the window. Therefore, we used the Kaplan-Meier method to correct for the edge effects.[37]

2.4 | Descriptive statistics of the point patterns

We first estimated the pair correlation function for each of the sweat gland patterns shown in Figure 3 and thereafter, obtained the groupwise pair correlation functions (see Figure 4) by pooling the estimated pair correlation functions of all the subjects within the group by taking a weighted average of the values of the pair correlation fiunctions at each distance.[35] Since the number of points in the point patterns within each group varied quite a lot, the individual pair correlation functions were weighted by the squared number of points, not number of points, when pooling.[38-40] The pair

FIGURE 4 Pooled pair correlation functions for the three groups [Colour figure can be viewed at wileyonlinelibrary.com]



correlation functions show a clear sign of inhibition in all three groups ($g(r) < 1$ for small r). Further, the initial peak of the functions appears approximately at the same distance for the control and suspected neuropathy groups. However, the diagnosed neuropathy group has the initial peak at a slightly longer distance, indicating somewhat larger range of inhibition than in the other two groups.

At very short distances, especially the control subjects seem to have some unexpected close pairs of points. Upon a closer inspection of the point patterns and the videos, it was reasonable to assume that some sweat spots had been detected as two nearby spots, instead of having merged into one. An obvious, simple idea to remove such close pairs of spots would be to merge all small glands having a larger spot closer than at some specified distance with the larger spot. However, such erroneous pairs of glands may appear at various (small) distances from each other and thus, applying a global limiting distance is not reasonable. Instead of using an additional image analysis step, we include some of this inaccuracy in the modeling and/or parameter estimation.

3 | SEQUENTIAL POINT PROCESS MODEL

Since sweat glands activate at different times, we modeled the activation by using sequential point processes similar to those suggested for modeling eye movements.[28] The points, activated sweat glands in our case, are generated sequentially conditioning on the already existing points. Points are added until the observed number of points in the pattern has been reached and the main focus here is to make inference on the arrival density. Below, we first recall the general sequential model[28] (Section 3.1) and specify it in our case without (Section 3.2) and with noise (3.4). Further, we discuss efficient inference for the sequential models (Section 3.3) and, finally, fit the sequential model with noise to the sweat gland data (Section 3.5).

3.1 | General sequential point process model

Denote by W the observation window and by n the fixed number of points in the pattern. The first point x_1 is assumed to be uniformly distributed in W and the kth point, $k = 2, \dots, n$, is assumed to follow the density $y \mapsto f(y; \vec{\mathbf{x}}_{k-1})$, where $\vec{\mathbf{x}}_{k-1} = (x_1, x_2, \dots, x_{k-1})$. The density function for the whole point pattern $(x_1, \dots, x_N)$ is then

$$\vec{\mathbf{x}}_n \mapsto \frac{1}{|W|}\prod_{k=2}^{n} f(x_k; \vec{\mathbf{x}}_{k-1}),$$

where $1/|W|$ is the contribution of the first point. A nice feature of the sequential point process models is that they have a tractable likelihood even though it can be costly to compute.[28]

3.2 | Soft-core model

The function f above should be chosen based on the phenomenon we would like to model. The activated sweat gland location patterns are repulsive. Our first attempt was to use a hard-core model, where sweat glands cannot be closer together than some minimum hard-core distance, but such a model turned out not to be flexible enough. Therefore, we suggest to use a soft-core model with the density

$$f_{SC}(y; \vec{\mathbf{x}}_k, R, \kappa) \propto \exp\left(-\sum_{i=1}^{k} \left(\frac{R}{d(y, x_i)} \right)^{2/\kappa} \right)$$

for adding the point y in the realization. Above, $R > 0$ is an inhibition range parameter and $0 < \kappa < 1$ in the exponent describes how "soft-core" the model is. In the limit as $\kappa \to 0$, we obtain a hard-core process with hard-core distance R. Some soft-core Gibbs point process models have been introduced in the literature,[41,42] including models with the particular interaction function that we use here.[43]

The log likelihood of the model becomes

$$l(R, \kappa; \vec{\mathbf{x}}_n) = -\log|W| - \sum_{k=2}^{n} \sum_{i=1}^{k-1} \left(\frac{R}{d(x_k, x_i)} \right)^{2/\kappa} - \sum_{k=2}^{n} \log Z(R, \kappa, \vec{\mathbf{x}}_k), \tag{1}$$

where

$$Z(R, \kappa, \vec{\mathbf{x}}_k)^{-1} = \int_W \exp\left(-\sum_{i=1}^{k-1} \left(\frac{R}{d(y, x_i)} \right)^{2/\kappa} \right) dy$$

is a normalizing constant.

3.3 | Efficient likelihood inference for the sequential models

Even though the likelihood of a sequential point process can be costly to compute, the particular sum structure in (1) allows faster computations. Using an integration scheme with J integration points $y_1, y_2, \dots, y_J$ with weights $w_1, w_2, \dots, w_J$, the last term in (1) can be written as

$$\begin{aligned} \sum_{k=2}^{n} \log Z(R, \kappa, \vec{\mathbf{x}}_k) &= \sum_{k=2}^{n} \log \left(\int_W \exp\left(-\sum_{i=1}^{k-1} \left(\frac{R}{d(y, x_i)} \right)^{2/\kappa} \right) dy \right)^{-1} \\ &= -\sum_{k=2}^{n} \log \sum_{j=1}^{J} w_j \exp\left(-\sum_{i=1}^{k-1} \left(\frac{R}{d(y_j, x_i)} \right)^{2/\kappa} \right). \end{aligned}$$

In total, there are $Jn(n-1)/2$ summands, among which only Jn are distinct. Therefore, the integrals are efficiently calculated by evaluating the terms in the innermost sum only once.

3.4 | Soft-core model with noise

To account for the incorrectly identified close pairs in the extracted point patterns, we used a mixture model where one of the components is a uniformly distributed error component. Such an error component can be added to any point process

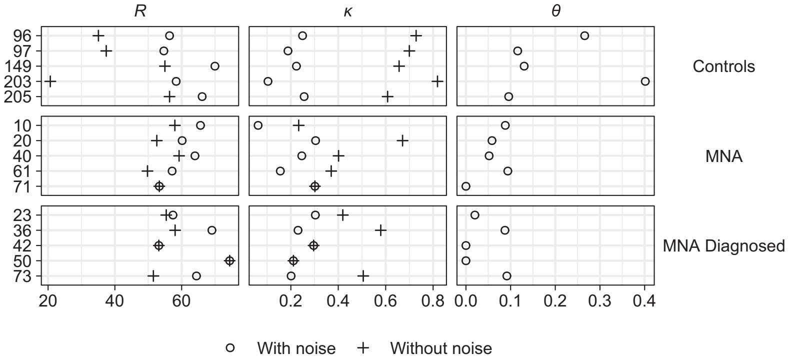


FIGURE 5 Parameter estimates of the soft-core model without and with noise fitted separately to each subject of the three groups (subject numbers shown on the left)

model and here, we add it in the sequential soft-core model. The arrival density of a point y (after the uniformly distributed first point) is then

$$
\begin{aligned}
f_M(y; \vec{\mathbf{x}}_k, R, \kappa, \theta) &= (1-\theta) f_{SC}(y; \vec{\mathbf{x}}_k, R, \kappa) + \frac{\theta}{|W|} \\
&= (1-\theta) Z(R, \kappa, \vec{x}_k) \exp\left(-\sum_{i=1}^{k} \left(\frac{R}{d(y, x_i)} \right)^{2/\kappa} \right) + \frac{\theta}{|W|}.
\end{aligned}
$$

Therefore, the point at y comes from the soft-core process with probability $1-\theta$ (the first term on the right-hand side of the formula) and from the uniformly distributed error process with probability θ. Even though this model allows extra points everywhere, not only near the real activated glands, it can improve estimation of the parameters as shown below. However, the parameter θ cannot be interpreted directly as the probability of incorrectly identified glands since some of the points without close neighbors regarded as noise could as well be true glands.

The log-likelihood of the soft-core model with uniformly distributed error is given by

$$
l_M(R, \kappa, \theta; \vec{\mathbf{x}}_n) = -\log|W| + \sum_{k=2}^{n} \log f_M(x_k; \vec{\mathbf{x}}_{k-1}, R, \kappa, \theta). \tag{2}
$$

3.5 | Application to the sweat gland data

The soft-core model was fitted without and with noise to each sweat gland point pattern independently. First, we compared the maximum likelihood estimates of the soft-core parameters obtained without or with added noise. Then, we fitted the model with noise to the data in a Bayesian framework to be able to better compare the goodness-of fit of the sequential soft-core model and the generative model presented in Section 4. We used regular grid based integration with 10 800 integration points to evaluate the likelihood in all cases.

3.5.1 | Parameter estimates without and with added noise

The parameter estimates obtained by maximizing the log likelihood (1) or (2) with respect to the parameters can be seen in Figure 5, where circles belong to the sequential soft-core model without noise and the pluses to the model with noise. The estimates obtained without noise for the range parameter R are on average smaller and the "softness" parameter κ larger in the control group than in the neuropathy groups. However, for the model fitted with noise (circles on Figure 5), only the mixture parameter θ, which is estimated larger for the control group than for the neuropathy groups, differs between the groups. Note that the κ and R parameters coincide for the two models if the mixture parameter θ is estimated to be zero, see subjects 42, 50, and 71 in Figure 5.

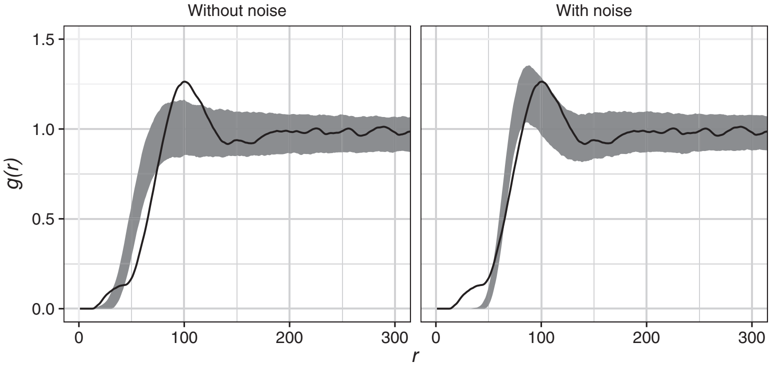


FIGURE 6 Empirical pair correlation functions (black lines) for subject 205 in the end of the video recording together with 95% global envelopes (grey areas) constructed from 25 000 simulations from the soft-core model estimated without (left) and with (right) noise

We investigated the goodness-of-fit of the fitted soft-core models by using the pair-correlation function. We generated samples from the sequential soft-core models with parameters R and κ estimated with and without noise. The uniform noise was not simulated. Figure 6 shows the empirical pair-correlation functions for subject 205 for the soft-core model estimated with and without noise together with 95% global envelopes[44,45] calculated from 25 000 samples of each model. It can be seen that for this subject, the range parameter is clearly underestimated if estimation is done without accounting for noise. For the other subjects, the goodness-of-fit of the model with noise was also as good or better than the goodness-of-fit of the model without noise. The difference between the observed and model (fitted with noise) based pair correlation function at short distances is explained by the incorrectly recorded close pairs of points that are present in the data but not in the simulations.

3.5.2 | Bayesian inference of the model with noise

We fitted the soft-core model with noise to the sweat gland data also by using standard likelihood-based Bayesian approach with robust adaptive Metropolis algorithm.[46] We ran the MCMC for 120 000 iterations and discarded the first 20 000 iterations as burn-in. As the prior distribution for the range parameter R, we used the Gamma distribution with shape parameter 3 and scale parameter 70/3 and the priors for κ and θ were both the uniform distribution on $[0, 1]$. The posterior histograms in Figure 7 show some variation within the groups but no clear differences between the groups: The arrival density parameters R and κ were estimated to be rather similar in all groups. The θ parameter related to the errors appears to be somewhat larger in the control group than in the other two groups.

Figure 8 shows the empirical pair-correlation functions for each subject together with the global envelopes[44,47] calculated from 25 000 simulations from the posterior predictive distribution of the fitted soft-core models with noise. In most cases, the envelopes cover the empirical curves. For some subjects, especially for the controls, the empirical pair-correlation function is not covered by the envelopes at very short distances. This is expected, as mentioned earlier, since according to the model used, this behavior is caused mainly by noise, which was not simulated. Thus, the pair correlation function reveals that, ignoring the noise at short distances, the quite clear inhibition and the gradual increase of interpoint distances around 50-75 pixels in the empirical patterns is captured by the soft-core model. However, the characteristic peak in the pair correlation functions around 100 pixels is not always captured by the model, particularly for the patients who have smaller number of activated sweat glands, as the envelopes are quite wide close to the peak. All the empirical pair correlation functions have the same characteristic behavior.

4 | GENERATIVE POINT PROCESS MODEL

In our second approach, we first model the underlying unobserved sweat glands and then, model the activated sweat glands as an independent thinning of the underlying gland pattern. Modeling the glands and the activation of them separately allows one to answer questions regarding specifically the activation process. One possible hypothesis is that the

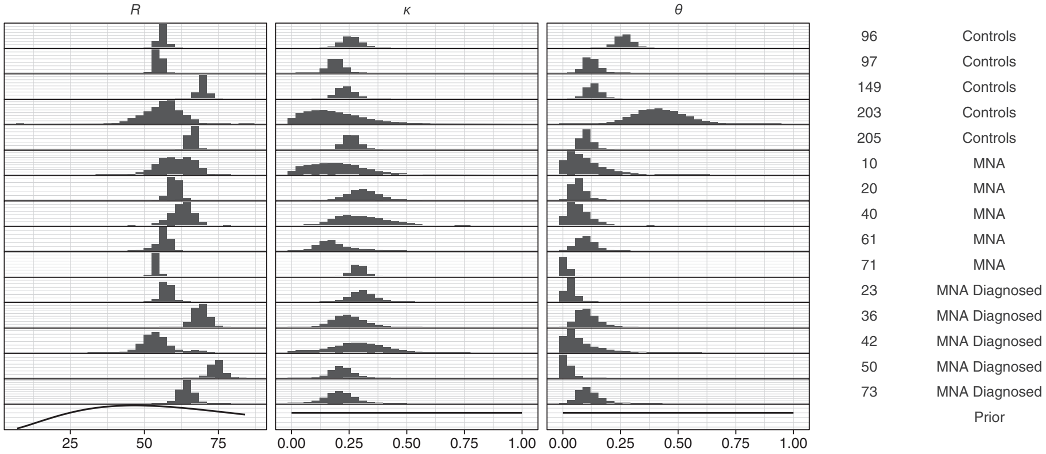


FIGURE 7 Posterior marginals for each subject (subject number and group given on the right) and each parameter (R, κ, and θ) for the soft-core model estimated with noise. The prior densities are given on the last row

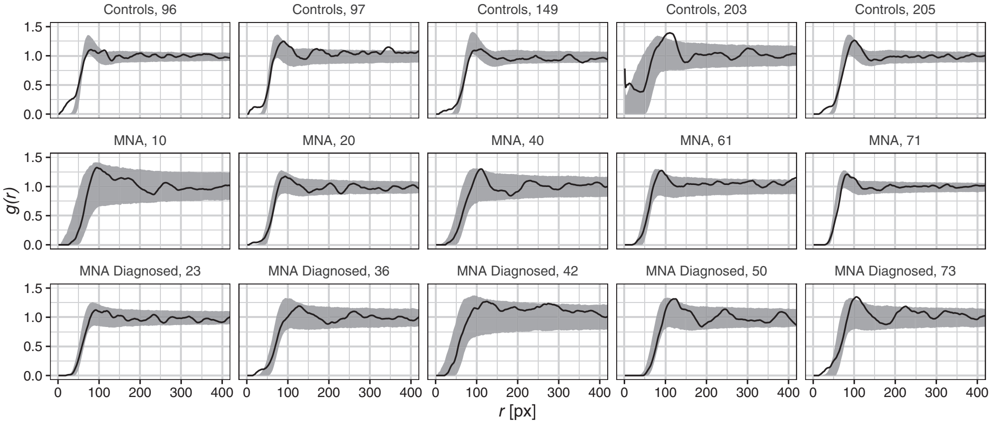


FIGURE 8 Empirical pair correlation functions (black lines) for each subject (subject number shown in the label above each plot) in the end of the video recording together with 95% global envelopes (gray areas) constructed from 25 000 simulations from the posterior predictive distribution of the soft-core model estimated with noise

underlying gland pattern itself is not different between controls and subjects with neuropathy, but the activation process is different. More specifically, almost all glands should activate on healthy subjects while the glands of the subjects with neuropathy could have a tendency to leave larger holes in the activation process.[16]

4.1 | Model specification

It seems reasonable to assume that the underlying (unobserved) sweat gland pattern is a rather densely packed regular point pattern covering the whole skin. To obtain such a structure, some type of soft-core sequential inhibition process, where points are added as long as it is possible (we do not know the actual number of glands), would be appropriate. However, it is not straightforward to decide when to stop adding points since theoretically, soft-core type of interaction always allows new points. Instead, we start by generating a simple sequential inhibition (SSI) model,[35] which is then disturbed to obtain a soft-core structure. A sample from the SSI model is generated sequentially by proposing points from the uniform distribution and accepting them if the pattern satisfies the hard-core condition with hard-core distance R, that is, the new proposed point does not lie within distance R from any earlier point. This is continued until there is no space left for new points. The disturbed SSI model is obtained from the "pure" SSI model by displacing the location of each point with an independent zero mean isotropic Gaussian random variable with covariance $\sigma^2 I$.

We assume independent gland activation, that is, that the final pattern is a result of an independent thinning of the underlying disturbed SSI process. Therefore, the model has three parameters: inhibition range R, hardness of inhibition σ, and probability of activation p.

4.2 | Parameter estimation using approximate Bayesian computation

For the generative model, we cannot write down the likelihood. However, sampling from the model is easy. We used the method proposed by Wang[48] to generate samples from the SSI process. Approximate Bayesian computation (ABC) is a method for Bayesian inference in situations where the likelihood of the model is intractable,[49,50] but it is possible to simulate the model. It is based on sampling from the (pseudo-) posterior distribution

$$\pi_\epsilon(\theta) = \pi(\theta)\mathbf{P}(\|s(Y_\theta) - s(y)\| < \epsilon),$$

where Y_θ follows the model with parameter vector θ, y is the data, $\pi(\cdot)$ is the prior distribution for the parameters, s is an appropriately chosen summary statistic, and ϵ is a tolerance level.

4.2.1 | ABC-MCMC

A simple ABC rejection sampler is expressed in Algorithm 1. This basic algorithm can be rather inefficient, but fortunately, there are several more efficient algorithms for performing ABC. We used an adaptive ABC-MCMC algorithm.[51] In our data study below, the MCMC was run for 10 000 000 iterations and the 250 000 simulated parameter values with the smallest distances $\|s(z) - s(y)\|$ were taken as the posterior sample.

Algorithm 1. A simple ABC rejection sampler

```
for $i \leftarrow 1, M$ do
    repeat
        Generate parameter vector $\theta'$ from the prior distribution $\pi$
        Generate a realization $z$ from the model with parameter vector $\theta'$
    until $\|s(z) - s(y)\| \leq \epsilon$
    $\theta_i \leftarrow \theta'$
end for
```

4.2.2 | Summary statistics

The choice of summary statistics is crucial for the ABC method to work. For a regular point process model, it is natural to use summary statistics based on the pair correlation function g. Instead of using the full pair correlation function g, we tried to find a specific part of it that would be sufficient for our purpose following the rule of thumb[52] that the number of summary statistics in the ABC approach should approximately match with the number of parameters to be estimated. The location of the first peak of the pair correlation function is intuitively connected to the inhibition range R. However, the

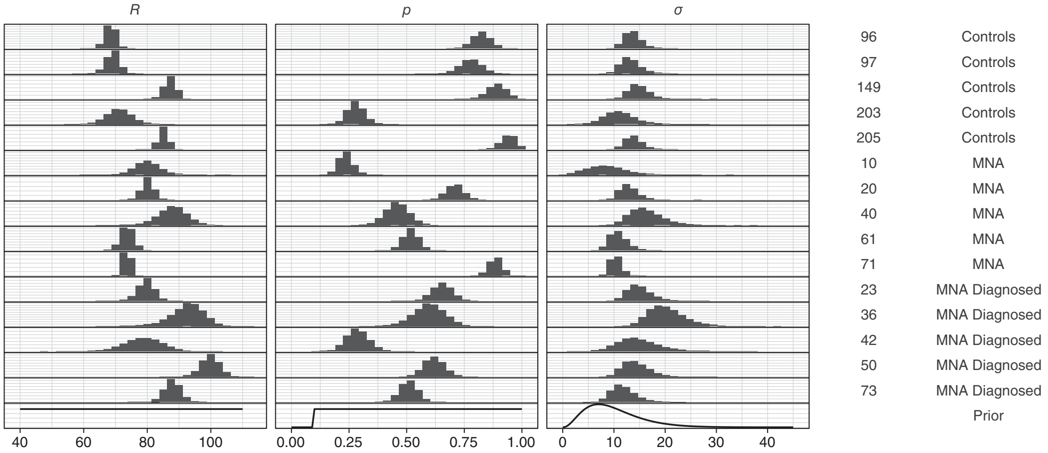


FIGURE 9 Posterior marginals for each subject (subject number and group given on the right) and each parameter (R, p, and σ) of the generative model

location of the first peak can be difficult to estimate exactly and thus, we used the smallest distance $r_1 > 10$ pixels where $g(r_1) = 0.75$ as the location of the uphill before the first peak. Furthermore, the slope of the uphill provides information on the "softness" parameter σ and we chose the smallest distance $r_2 > 10$ pixels where $g(r_2) = 1$ as the second summary statistic. Finally, the smallest distance r_3 in the empty space function F where $F(r_3) = 0.5$ was taken as the third summary statistic to represent the activation probability p. The empty space function was chosen because it gives information on the number of points but is not greatly affected by erroneous nearby points. Since all the chosen summary statistics, r_1, r_2, and r_3, have a similar order of magnitude, we did not have to add any weights in the ABC algorithm. The specific values 0.75 and 1 were chosen to be somewhat separated and not too small to account for possible errors caused by splitting of spots into multiple glands that would cause the pair-correlation function not to start from zero. In addition, we only considered distances greater than 10 pixels since at very short distances the kernel estimator of the pair correlation function is not very reliable. These choices worked well for the sweat gland data, as demonstrated below.

4.3 | Application to the sweat gland data

The generative model was fitted to the sweat gland data using the ABC approach described above. In addition to the above specifications, we needed to set the priors. For R we used an improper, uniform prior on $[40, \infty)$ restricting that R could not be unreasonably small, while in addition to being unrealistic, small R values result in a large number of points in the SSI process which is computationally challenging. The prior of p was uniform on $[0.1, 1]$, stating that at least 10% of the glands (modeled by the underlying disturbed SSI process) needed to activate and thus be observed. Furthermore, for σ, we used the gamma distribution with the shape parameter equal to 10/3 and scale parameter equal to 3. While the priors R and p can be considered rather non-informative, the prior for σ was somewhat informative suggesting positive, but not too large σ. Note that if σ was very large in comparison to R, it would break all the structure of the SSI process, which is unreasonable.

The posterior marginal histograms for the parameters can be seen in Figure 9 and 95% global envelopes for the pair correlation function constructed from 25 000 simulations from the posterior predictive distribution in Figure 10. As can be seen in Figure 9, the parameter estimates vary somewhat between the subjects and groups. Differences in the softness of the model, that is, in the values of the parameter σ, are small. However, there seems to be a slight tendency for the inhibition range R to be a little smaller in the control group than in the MNA groups, but the difference is not clear based on the limited amount of data we have. The range was always between 60 pixels and 100 pixels. Furthermore, the control subjects tend to have a larger activation probability than the MNA patients, but the within group variation is large. This is

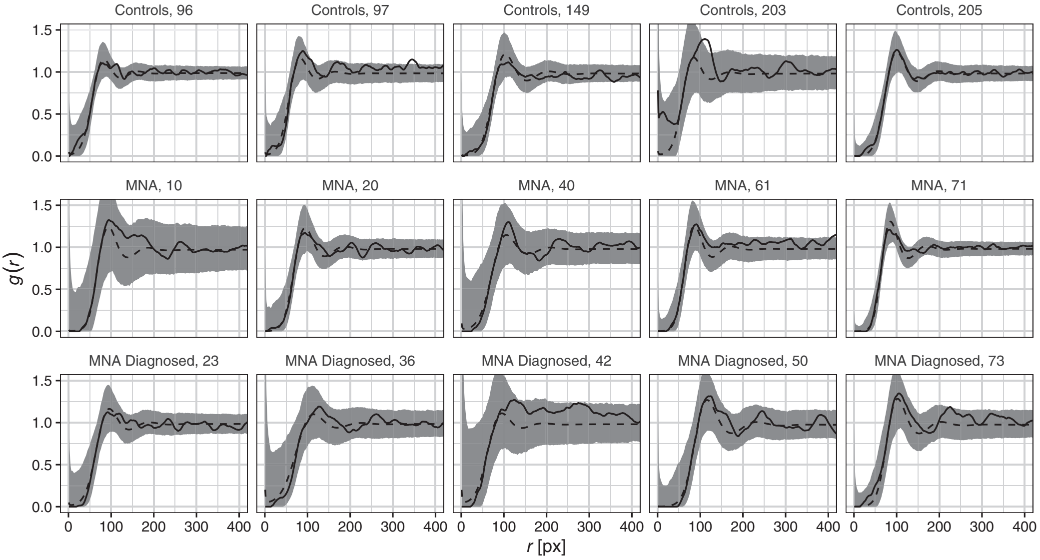


FIGURE 10 Empirical pair correlation functions (lines) for each subject (subject number shown in the label above each plot) in the end of the video recording together with 95% global envelopes (gray areas) and means (dashed lines) constructed from 25 000 simulations from the posterior predictive distribution of the generative model

in agreement with earlier studies, which indicate that a larger number of sweat glands of controls than of MNA patients activate.[4]

According to the visual evaluation of the global envelopes of the pair-correlation function (see Figure 10) and empty space function (see Figure 11), the model seems to fit quite well to the data. It captures the behavior of the pair correlation function both at small distances and around the initial peak. It should also be mentioned that the envelopes for the pair correlation function are rather wide at small distances covering the observed functions almost in all cases, even though the model did not include any error term. The wide envelopes are due to the relatively wide posterior distribution of σ. Namely, large σ can lead to some close pairs in the patterns and consequently also positive values of the pair correlation function at small distances. Another reason for the relatively wide envelopes may be that the summary statistics used in the ABC approach were chosen such that they do not use any information at very short distances.

We explored a few other priors for σ, namely, improper uniform and exponential distributions with means 1, 2, and 4. The posterior distributions of the other parameters were not affected by the choice of the prior for σ, but the posterior of σ itself was somewhat sensitive to the choice and also the goodness-of-fit of the model measured by the pair correlation was affected. Namely, the improper uniform prior led to wider posterior distribution of σ and large σ caused the variation of the pair correlation function to be even higher at small distances. On the other hand, the strict exponential priors shrank the posterior distribution toward zero, and very small σ caused the peak of the pair correlation function to be too sharp. Thus, the disturbance parameter σ needed a somewhat informative prior to lead to a good fit of the model.

We simulated patterns from the posterior predictive distribution and the simulated patterns mimic the data patterns rather well, see Figure 12. Note, in particular, that the independent thinning seems to produce empty regions similar to those observed in the point pattern data, as also indicated by the empty space function (Figure 11).

5 DISCUSSION

Videos produced by dynamic sweat tests provide not only the total and per gland sweat volume and the number and density of activated sweat glands but also a time series of spatial patterns of the activated sweat glands. Up to now, only

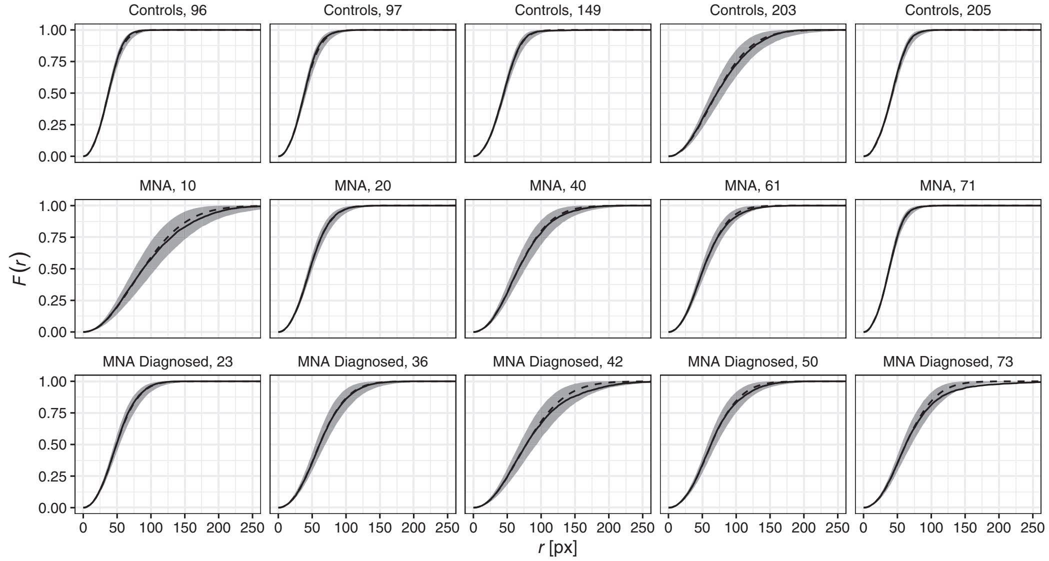


FIGURE 11 Empirical empty space functions (black lines) for each subject (subject number shown in label) in the end of the video recording together with 95% global envelopes (gray areas) and means (dashed lines) constructed from 25 000 simulations from the posterior predictive distribution of the generative model

the sweat volume and the number/density of activated sweat glands have been used in the analysis and comparison between healthy subjects and subjects with neuropathy. However, visual inspection has indicated that the spatial pattern of activated sweat glands of subjects with neuropathy seem to have more empty areas than the patterns of healthy controls. This may indicate that the neuropathy does not result in random deactivation of sweat glands but the glands that activate are arranged in a different pattern than in control patterns. To quantify this observation, spatial analysis is needed.

Here, we analyzed videos of sweat gland activation recorded from five controls, five subjects with suspected neuropathy, and five subjects with diagnosed neuropathy. The initial non-parametric spatial analysis by using pair correlation functions showed clear indication of inhibition in all three groups. In addition, it suggested some differences between the spatial patterns of activated sweat glands in subjects diagnosed for neuropathy and the non-diagnosed and control groups. To further investigate and compare the spatial patterns and the activation processes, we suggested two point process models for the activation of sweat glands, a sequential soft-core model describing the appearance of the activated sweat glands and a thinned disturbed SSI process, that we call a generative model, where we start by modeling the underlying unobserved sweat gland pattern.

Maximizing the log-likelihood function of a sequential point process has been considered computationally costly due to the integrals in the normalizing constants.[28] However, for the sequential soft-core model, these integrals have a particular sum form which allows efficient computation of the log-likelihood and in turn Bayesian inference. The same efficient computation scheme is applicable for any sequential point process having an arrival density with a similar sum structure. To estimate the parameters of the generative model, we employed an ABC algorithm since the likelihood function was not easily available.

Even though our proposed image analysis approach worked well, there were some incorrectly identified close pairs of glands in the extracted point patterns. To take into account such errors, we added an error term in the sequential soft-core model resulting in a mixture model having a soft-core component and a uniform noise component. For the generative model, on the other hand, the summary statistics in the ABC approach were chosen such that they were robust to close pairs of points. The good fit of the models shows that it is possible to account for some, but likely not many, errors in the pre-processing of videos. Such an analysis needs to be applied with care, but allows the researcher to focus on the analysis more than on the pre-processing.

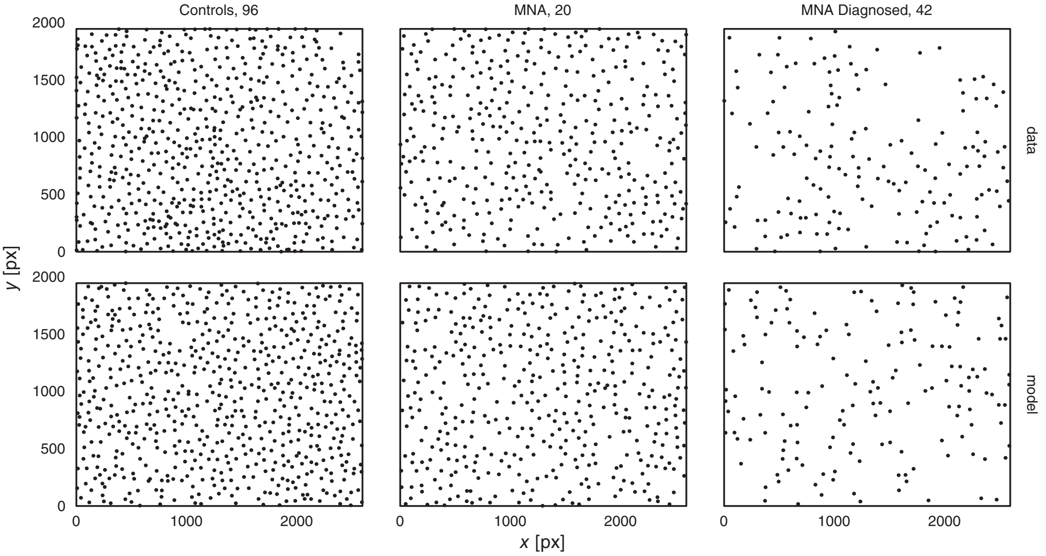


FIGURE 12 The original point patterns (top) and patterns generated from the corresponding posterior predictive distributions of the generative model (bottom) for one subject from each group (96, 20, and 42)

The proposed models were fitted to the data. The fit of the generative model was good regardless of the disease status, as the point patterns in Figure 12 and the spatial summary characteristics in Figures 10 and 11 show. It should be noted that this model based on independent activation of underlying sweat glands was also able to produce similar holes in the point patterns as observed earlier in the sweat gland patterns of diabetic patients;[16] see, for example, the bottom right plot in Figure 12. Apparently the SSI process with random displacements of points is a working model for the underlying sweat glands.

The model parameters estimated from the patterns of healthy subjects and of subjects suffering from neuropathy were compared. Based on the generative model, detailed information was obtained on the sweat patterns: while the hardness of inhibition parameter was rather similar for each subject, the inhibition range R and the activation probability p were characteristic for each subject. The former characterizes the pattern of all glands, while the latter is the probability of these glands to activate. There was some indication that the inhibition range R was slightly smaller and the activation probability p larger in the control group than in the neuropathy groups (Figure 9). Both of these observations are in agreement with an earlier study,[16] where the density of activated sweat glands was found to be lower for diabetic patients than controls. However, the spatial analysis provides more detailed information than the non-spatial analysis alone, and the model parameters might facilitate classification of subjects in the future. For example, the control subject 203 had rather small number of activated sweat glands (Figure 3), but simultaneously a small inhibition range parameter. Thus, the parameters may contain more valuable information jointly as either of them alone.

We believe that the models suggested here, especially the generative model, would give more impactful insight into the sweat patterns in a larger study including more subjects. Especially, combining the point pattern approach presented here with non-spatial covariates like age or body mass index would provide a better understanding of the sweat patterns and provide an explanation for the individually varying generative model parameters. In addition, incorporating the amount of sweat produced over time by individual sweat glands into the analysis would further improve our understanding and lead to further methodological development.

ACKNOWLEDGEMENTS

Mikko Kuronen and Mari Myllymäki have been financially supported by the Academy of Finland (Project Numbers 295100, 306875, and 327211) and Aila Särkkä by the Swedish Research Council (VR 2013-5212) and by the Swedish

Foundation for Strategic Research (SSF AM13-0066). The authors thank Matti Vihola for useful discussions. We are also grateful for the two anonymous reviewers for the valuable comments.

DATA AVAILABILITY STATEMENT

The original video data are not publicly available. The point pattern data and code used in the analysis is available as a github repository: https://github.com/mikkoku/SweatPaper.git.

ORCID

Mikko Kuronen https://orcid.org/0000-0002-8089-7895

REFERENCES

1. Coon JM, Rothman S. The sweat response to drugs with nicotine-like action. *J Pharmacol Exp Ther*. 1941;73(1):1-11.
2. Lader M, Montagu J. The psycho-galvanic reflex: a pharmacological study of the peripheral mechanism. *J Neurol Neurosurg Psychiatry*. 1962;25:126-133.
3. Minor V. Ein neues Verfahren zu der klinischen Untersuchung der Schweißabsonderung. *Deutsche Zeitschrift für Nervenheilkunde*. 1928;101:302-307.
4. Loavenbruck AJ, Hodges JS, Provitera V, Nolano M, Wendelshafer-Crabb G, Kennedy WR. A device to measure secretion of individual sweat glands for diagnosis of peripheral neuropathy. *J Peripher Nerv Syst*. 2017;22(2):139-148. https://doi.org/10.1111/jns.12212.
5. Sumner C, Sheth S, Griffin J, Cornblath D, Polydefkis M. The spectrum of neuropathy in diabetes and impaired glucose tolerance. *Neurology*. 2003;60(1):108-111. https://doi.org/10.1212/WNL.60.1.108.
6. Said G. Diabetic neuropathy–a review. *Nat Clin Pract Neurol*. 2007;3:331-340. https://doi.org/10.1038/ncpneuro0504.
7. Sharma S, Venkitaraman R, Vas PRJ, Rayman G. Assessment of chemotherapy-induced peripheral neuropathy using the LDIFLARE technique: a novel technique to detect neural small fiber dysfunction. *Brain Behav*. 2015;5(7):e00354. https://doi.org/10.1002/brb3.354.
8. Hilz MJ, Dütsch M. Quantitative studies of autonomic function. *Muscle Nerve Offic J Am Assoc Electrodiagnostic Med*. 2006;33(1):6-20. https://doi.org/10.1002/mus.20365.
9. Illigens BMW, Gibbons CH. Sweat testing to evaluate autonomic function. *Clin Auton Res*. 2009;19:79-87. https://doi.org/10.1007/s10286-008-0506-8.
10. Minota K, Coon EA, Benarroch EE. Neurologic aspects of sweating and its disorders. *Neurology*. 2019;92(21):999-1005. https://doi.org/10.1212/WNL.0000000000007540.
11. Fealey RD, Low PA, Thomas JE. Thermoregulatory sweating abnormalities in diabetes mellitus. *Mayo Clin Proc*. 1989;64(6):617-628. https://doi.org/10.1016/S0025-6196(12)65338-5.
12. Papanas N, Papatheodorou K, Christakidis D, et al. Evaluation of a new indicator test for sudomotor function (Neuropad®) in the diagnosis of peripheral neuropathy in type 2 diabetic patients. *Exp Clin Endocrinol Diabetes*. 2005;113:195-198. https://doi.org/10.1055/s-2005-837735.
13. Harris DR, Polk BF, Willis I. Evaluating sweat gland activity with imprint techniques. *J Investig Dermatol*. 1972;58(2):78-84. https://doi.org/10.1111/1523-1747.ep12551676.
14. Low PA, Caskey PE, Tuck RR, Fealey RD, Dyck PJ. Quantitative sudomotor axon reflex test in normal and neuropathic subjects. *Ann Neurol*. 1983;14:573-580.
15. Low VA, Sandroni P, Fealey RD, Low PA. Detection of small-fiber neuropathy by sudomotor testing. *Muscle Nerve*. 2006;34(1):57-61. https://doi.org/10.1002/mus.20551.
16. Provitera V, Nolano M, Caporaso G, Stancanelli A, Santoro L, Kennedy WR. Evaluation of sudomotor function in diabetes using the dynamic sweat test. *Neurology*. 2010;74(1):50-56. https://doi.org/10.1212/WNL.0b013e3181c7da4b.
17. Kennedy WR, Selim MM, Wendelschafer-Crabb G, et al. A device to quantify sweat in single sweat glands to diagnose neuropathy. *Journal of Medical Devices*. 2013;7(3):030941. https://doi.org/10.1115/1.4024527.
18. Loavenbruck A, Sit N, Provitera V, Kennedy W. High-resolution axon reflex sweat testing for diagnosis of neuropathy. *Clin Auton Res*. 2019;29:55-62. https://doi.org/10.1007/s10286-018-0546-7.
19. Byers S, Raftery AE. Nearest-neighbor clutter removal for estimating features in spatial point processes. *J Am Stat Assoc*. 1998;93(442):577-584. https://doi.org/10.1080/01621459.1998.10473711.
20. Cressie N, Lawson AB. Hierarchical probability models and Bayesian analysis of mine locations. *Adv Appl Probab*. 2000;32(02):315-330. https://doi.org/10.1017/s0001867800009940.
21. Walsh DCI, Raftery AE. Detecting mines in minefields with linear characteristics. *Technometrics*. 2002;44(1):34-44.
22. Redenbach C, Särkkä A, Sormani M. Classification of points in superpositions of Strauss and Poisson processes. *Spat Stat*. 2015;12:81-95. https://doi.org/10.1016/j.spasta.2015.03.003.
23. Rajala T, Redenbach C, Särkkä A, Sormani M. Variational Bayes approach for classification of points in superpositions of point processes. *Spat Stat*. 2016;15:85-99. https://doi.org/10.1016/j.spasta.2015.12.001.
24. Lund J, Rudemo M. Models for point processes observed with noise. *Biometrika*. 2000;87:235-249. https://doi.org/10.1093/biomet/87.2.235.

25. Lund J, Penttinen A, Rudemo M. Bayesian analysis of spatial point patterns from noisy observations. [Ph.D. thesis]. Statistical inference and perfect simulation for point processes observed with noise by J Lund. Denmark: Department of Mathematics and Physics, Royal Veterinary and Agricultural University; 1999.
26. Chakraborty A, Gelfand AE. Analyzing spatial point patterns subject to measurement error. *Bayesian Anal*. 2010;5(1):97-122. https://doi.org/10.1214/10-BA504.
27. Bar-Hen A, Chadœuf J, Dessard H, Monestiez P. Estimating second order characteristics of point processes with known independent noise. *Stat Comput*. 2013;23(3):297-309. https://doi.org/10.1007/s11222-011-9311-7.
28. Penttinen A, Ylitalo AK. Deducing self-interaction in eye movement data using sequential spatial point processes. *Spat Stat*. 2016;17:1-21. https://doi.org/10.1016/j.spasta.2016.03.005.
29. Bezanson J, Edelman A, Karpinski S, Shah VB. Julia: a fresh approach to numerical computing. *SIAM Rev*. 2017;59(1):65-98. https://doi.org/10.1137/141000671.
30. R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing; 2018.
31. Hinkley DV. Inference about the change-point in a sequence of random variables. *Biometrika*. 1970;57(1):1-17. https://doi.org/10.2307/2334932.
32. Yao YC, Davis RA. The asymptotic behavior of the likelihood ratio statistic for testing a shift in mean in a sequence of independent normal variates. *Sankhya Ind J Stat Ser A (1961-2002)*. 1986;48(3):339-353.
33. James B, James KL, Siegmund D. Tests for a change-point. *Biometrika*. 1987;74(1):71-83. https://doi.org/10.2307/2336022.
34. Pettitt AN. A non-parametric approach to the change-point problem. *J Royal Stat Soc Ser C (Appl Stat)*. 1979;28(2):126-135. https://doi.org/10.2307/2346729.
35. Illian J, Penttinen A, Stoyan H, Stoyan D. *Statistical Analysis and Modelling of Spatial Point Patterns*. Chichester, England: John Wiley & Sons, Ltd; 2008.
36. Stoyan D, Fractals SH. *Random Shapes and Point Fields: Methods of Geometrical Statistics*. Hoboken, NJ: Wiley; 1994.
37. Baddeley A, Gill RD. Kaplan-Meier estimators of distance distributions for spatial point processes. *Ann Stat*. 1997;25(1):263-292.
38. Diggle PJ, Mateu J, Clough HE. A comparison between parametric and non-parametric approaches to the analysis of replicated spatial point patterns. *Adv Appl Probab*. 2000;32:331-343.
39. Myllymäki M, Panoutsopoulou IG, Särkkä A. Analysis of spatial structure of epidermal nerve entry point patterns based on replicated data. *J Microsc*. 2012;247(3):228-239.
40. Schladitz K, Särkkä A, Pavenstädt I, Haferkamp O, Mattfeldt T. Statistical analysis of intramembranous particles using freeze fracture specimens. *J Microsc*. 2003;211:137-153.
41. Ogata Y, Tanemura M. Estimation of interaction potentials of spatial point patterns through the maximum likelihood procedure. *Ann Inst Stat Math*. 1981;33(2):315-338. https://doi.org/10.1007/BF02480944.
42. Ogata Y, Tanemura M. Likelihood analysis of spatial point patterns. *J Royal Stat Soc Ser B (Methodol)*. 1984;46(3):496-518. https://doi.org/10.1111/j.2517-6161.1984.tb01322.x.
43. Baddeley A, Rubak E, Turner R. *Spatial Point Patterns: Methodology and Applications with R*. London, UK: Chapman & Hall/CRC Press; 2015.
44. Myllymäki M, Mrkvička T, Seijo H, Grabarnik P, Hahn U. Global envelope tests for spatial processes. *J Royal Stat Soc Ser B (Stat Methodol)*. 2017;79:381-404. https://doi.org/10.1111/rssb.12172.
45. Myllymäki M, Mrkvička T. GET: global envelopes in R; 2019. arXiv:1911.06583 [stat.ME].
46. Vihola M. Robust adaptive Metropolis algorithm with coerced acceptance rate. *Stat Comput*. 2012;22:997-1008. https://doi.org/10.1007/s11222-011-9269-5.
47. Mrkvička T, Myllymäki M, Jílek M, Hahn U. A one-way ANOVA test for functional data with graphical interpretation; 2018. arXiv:1506.01646 [stat.ME].
48. Wang JS. A fast algorithm for random sequential adsorption of discs. *Int J Modern Phys C*. 1994;05(04):707-715. https://doi.org/10.1142/S0129183194000817.
49. Marin JM, Pudlo P, Robert CP, Ryder RJ. Approximate Bayesian computational methods. *Stat Comput*. 2012;22(6):1167-1180. https://doi.org/10.1007/s11222-011-9288-2.
50. Sunnåker M, Busetto AG, Numminen E, Corander J, Foll M, Dessimoz C. Approximate Bayesian computation. *PLoS Comput Biol*. 2013;9(1):1-10. https://doi.org/10.1371/journal.pcbi.1002803.
51. Vihola M, Franks J. On the use of approximate Bayesian computation Markov chain Monte Carlo with inflated tolerance and post-correction. *Biometrika*. 2020;107(2):381-395. https://doi.org/10.1093/biomet/asz078.
52. Li W, Fearnhead P. On the asymptotic efficiency of approximate Bayesian computation estimators. *Biometrika*. 2018;105(2):285-299. https://doi.org/10.1093/biomet/asx078.

How to cite this article: Kuronen M, Myllymäki M, Loavenbruck A, Särkkä A. Point process models for sweat gland activation observed with noise. *Statistics in Medicine*. 2021;40:2055–2072. https://doi.org/10.1002/sim.8891

III

HIERARCHICAL LOG GAUSSIAN COX PROCESS FOR REGENERATION IN UNEVEN-AGED FORESTS

by

Kuronen, M., Särkkä, A., Vihola, M., and Myllymäki, M.

Environmental and Ecological Statistics, 2021.

doi:10.1007/s10651-021-00514-3

Environmental and Ecological Statistics
https://doi.org/10.1007/s10651-021-00514-3

Hierarchical log Gaussian Cox process for regeneration in uneven-aged forests

Mikko Kuronen[1] · **Aila Särkkä**[2] · **Matti Vihola**[3] · **Mari Myllymäki**[1]

Received: 12 April 2021 / Revised: 6 July 2021 / Accepted: 26 July 2021

Abstract

We propose a hierarchical log Gaussian Cox process (LGCP) for point patterns, where a set of points $\boldsymbol{x}$ affects another set of points $\boldsymbol{y}$ but not vice versa. We use the model to investigate the effect of large trees on the locations of seedlings. In the model, every point in $\boldsymbol{x}$ has a parametric influence kernel or signal, which together form an influence field. Conditionally on the parameters, the influence field acts as a spatial covariate in the intensity of the model, and the intensity itself is a non-linear function of the parameters. Points outside the observation window may affect the influence field inside the window. We propose an edge correction to account for this missing data. The parameters of the model are estimated in a Bayesian framework using Markov chain Monte Carlo where a Laplace approximation is used for the Gaussian field of the LGCP model. The proposed model is used to analyze the effect of large trees on the success of regeneration in uneven-aged forest stands in Finland.

Keywords Bayesian inference · Competition kernel · Laplace approximation · MCMC · Spatial random effects · Tree regeneration

Mathematics Subject Classification 62M30 · 62F15 · 60G55

Handling Editor: Luiz Duczmal.

Mikko Kuronen, Mari Myllymäki and Matti Vihola were financially supported by the Academy of Finland (Project Numbers 306875, 327211, 295100 and 315619) and Aila Särkkä by the Swedish Research Council (VR 2018-03986).

✉ Mikko Kuronen
mikko.kuronen@luke.fi

[1] Natural Resources Institute Finland (Luke), Helsinki, Finland

[2] Mathematical Sciences, Chalmers University of Technology and the University of Gothenburg, Gothenburg, Sweden

[3] Department of Mathematics and Statistics, University of Jyväskylä, Jyvaskyla, Finland

1 Introduction

Hierarchical relationships or interactions, where a plant species affects the locations or intensity of another species but not vice versa, often occur in ecological communities (e.g. Dieckmann et al. 2000). An example of such a hierarchical relationship is that proximity of large trees affects the intensity of seedling either positively, e.g. by protecting against wind, or negatively by giving too much shade. Mathematically, we can describe such plant communities by two point processes, Y and X, where one (X) is affecting the other (Y) but not vice versa.

The hierarchical interaction assumption affects the inference for Y and X greatly since X can be modeled independently of Y and Y is modeled conditionally on X. A realization of the point process X acts then as a source of heterogeneity in the distribution of Y. Högmander and Särkkä (1999) modeled interaction between two territorial ant species using Gibbs point processes under such an assumption. A similar hierarchical Gibbs point process approach was used in Grabarnik and Särkkä (2009) and Genet et al. (2014). Furthermore, Illian et al. (2009) modeled the spatial pattern of resprouter species (Y) given the locations of seeders (X) in a hierarchical set-up having an inhomogeneous Poisson process as a model for the resprouters.

Here, we model the intensity of new seedlings in a spruce-dominated uneven-aged (boreal) forest given the locations and diameters at breast height (dbh) of large trees. Thus, our X process of large trees is a marked point process, where the mark of a tree is the dbh. The data consist of 14 sample plots from an experiment of continuous cover forestry involving single-tree selection in four nearby areas in Southern Finland (Fig. 1). The system relies on the natural emergence of new seedlings, and a continuous recruitment is necessary for long-term sustainability in a wide sense (e.g. Eerikäinen et al. 2014; Kuusinen et al. 2019). While a sufficient number of seedlings is necessary for the success of regeneration, our focus here is on the spatial distribution of the seedlings within the plots, and the effect of large trees on it.

Like in the resprouter and seeder case above, an inhomogeneous Poisson process would be a reasonable model since the effect of large trees could be added in the model as an explanatory variable. However, already visual inspection of the patterns of seedlings $\boldsymbol{y}$ indicates that the patterns tend to be rather clustered, beyond the clustering that may be explained by the patterns of large trees $\boldsymbol{x}$. Due to such unexplained clustering, a log Gaussian Cox process (LGCP) (Møller et al. 1998) is a more appropriate model for the conditional point process of seedlings given large trees.

To model the effect of the large trees X, we assume that each tree $x \in X$ emits a signal or impulse that describes the effect of the tree on its neighborhood. We assume that this effect decreases with the distance from the tree x. In general, the size of the effect as well as the range of the effect could depend on the size or other properties of the tree, e.g. its dbh. Because we do not have precise a priori information on the size and range of the effects, we use parametric signals similar to the ones found in the literature (Adler 1996; Pommerening et al. 2011; Häbel et al. 2019; Pommerening and Grabarnik 2019). The individual signals are then superimposed to form an influence field, which describes the overall influence of the points of X on any location s in the observation window W. These kinds of models have been used to model, for example, effect of neighboring individuals on the growth of a subject tree, survival of seedlings

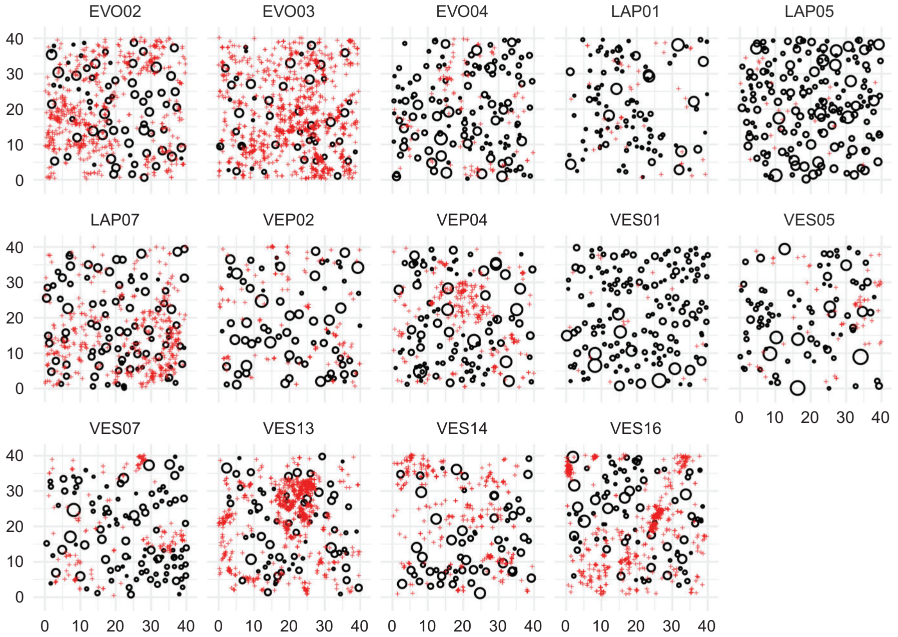


Fig. 1 Trees with dbh at least 7 cm (open circles with radii relative to the dbh of the tree) and new seedlings (red crosses) in areas of size 40 m × 40 m. The headings give abbreviations for the plot locations and numbers

and ground vegetation in different contexts (e.g. Wu et al. 1985; Miina and Pukkala 2002; Pommerening et al. 2011; Häbel et al. 2019; Kuuluvainen and Pukkala 1989; Kühlmann-Berenzon et al. 2005).

Our idea here is to include the superimposed individual signals in the log intensity function of the LGCP model. Using parametric models for the signals, the intensity of the LGCP is a non-linear function of the model parameters. According to Pommerening and Sánchez Meador (2018) the signals are aggregated additively or multiplicatively and there is no evidence to prefer either of these ways. We follow Pommerening et al. (2011) and Illian et al. (2009) and aggregate the signals additively.

Our Bayesian inference algorithm is based on Markov chain Monte Carlo (MCMC) sampling for parameters, and a Laplace approximation is used for the latent random field of the LGCP to avoid high-dimensional MCMC sampling. Laplace approximations are widely used for inference of latent Gaussian fields, for instance within the popular INLA method (Rue et al. 2009). However, in contrast to INLA, MCMC is more robust, and can cope with multimodal parameter posteriors.

The large tree process typically extends beyond the borders of the sample plot. However, we have observed the process in the same observation window as the seedlings. Thus, the influence field computed only from the observed trees is weaker near the borders than the field computed from the fully observed large tree process would be. In order to account for the unobserved trees outside the observation window, we compute the influence field using an edge correction method similar to that suggested in Kühlmann-Berenzon et al. (2005): the unobserved trees are imputed based on the assumption that the locations of large trees are distributed according to a Poisson process. This rather simple edge correction method can be efficiently implemented within the Bayesian inference, in contrast to alternatives where the locations (and sizes) of unobserved large trees would be included in the Bayesian inference as unknowns and simulated within the MCMC approach.

The rest of the paper is organized as follows. In Sect. 2, we give some examples of influence kernels and introduce the conditional LGCP model. The Bayesian estimation approach, including the edge correction, is described in Sect. 3. Section 4 presents the results of a simulation experiment that was conducted to explore the performance of the proposed estimation and edge correction methods. Finally, the forestry data are described in further detail and studied in Sect. 5. Section 6 is for discussion.

2 Conditional log Gaussian Cox process model

Let us have a bivariate point process in $\mathbb{R}^2$ consisting of an unmarked point process Y and an unmarked or a marked point process X. Let us further assume that we have observed a realization of process Y, namely $\boldsymbol{y} = \{y_i\}$, in a bounded window $W \subset \mathbb{R}^2$. Our primary interest is in the spatial pattern $\boldsymbol{y}$ which is affected by a realization $\boldsymbol{x}$ of the spatial point process X. The spatial pattern $\boldsymbol{x}$ can consist only of the point locations x_j or of the point locations and marks, $[x_j, m_j]$, if some characteristics (marks) m_j of the points x_j are available. In our forestry application, $\boldsymbol{y}$ consists of the locations of seedlings, while $\boldsymbol{x}$ is the pattern of locations and dbh's of large trees.

In our approach, the effect of $\boldsymbol{x}$ on $\boldsymbol{y}$ is modeled using the influence kernels around the points of $\boldsymbol{x}$ that are explained in Sect. 2.1. To account for the clustering in the pattern $\boldsymbol{y}$ not explained by $\boldsymbol{x}$, the LGCP model is proposed and defined in Sect. 2.2. Replicated point patterns are discussed in Sect. 2.3.

2.1 Influence kernels and influence field

We assume that each point $[x_j, m_j]$ of the process X introduces an influence kernel around its location. We focus on isotropic influence kernels of the form $c(h; m_j, \boldsymbol{\theta}_I)$, where $h = \|s - x_j\|$ is the distance between the location s of interest and x_j. Many kernels have been suggested in the literature for different applications (e.g. Adler 1996; Illian et al. 2008; Pommerening et al. 2011; Pommerening and Maleki 2014; Schneider et al. 2006). We used a mark independent Gaussian kernel

$$c(h; \theta) = \exp\left(-(h/\theta)^2\right), \tag{1}$$

where $\theta > 0$ is an unknown influence range parameter. Here the influence of a point gradually decreases with the distance from the point.

A mark dependent generalization of (1) is given by

$$c(h, m; \boldsymbol{\theta}_I) = m^{\alpha} \exp\left(-\left(\frac{h}{\theta m^{\delta}}\right)^2\right) \tag{2}$$

with $\boldsymbol{\theta}_I = (\theta, \delta, \alpha)$, where $\theta > 0$, $\delta > 0$, and $\alpha \geq 0$. If $\alpha = 0$, the mark affects only the range of influence and if $\alpha > 0$, it affects both the range and the strength (e.g. Pommerening et al. 2011).

The influence field of the process X can then be defined as a superposition of the individual influence kernels,

$$C(s; \boldsymbol{\theta}_I, X) = \sum_{[x_j, m_j] \in X} c(\|s - x_j\|, m_j; \boldsymbol{\theta}_I).$$

2.2 Conditional model

Since $\boldsymbol{y}$ is affected by $\boldsymbol{x}$, we introduce a conditional point process model for $\boldsymbol{y}$ given $X = \boldsymbol{x}$, where the intensity of Y is affected by the influence field of $\boldsymbol{x}$. This conditional model is a LGCP with the intensity

$$\Lambda(s; \boldsymbol{\beta}, \boldsymbol{\theta}_I, \boldsymbol{x}, Z) = \exp(\beta_0 + \beta_1 C(s; \boldsymbol{\theta}_I, \boldsymbol{x}) + Z(s)), \tag{3}$$

where $C(s; \boldsymbol{\theta}_I, \boldsymbol{x})$ is a parametric influence field, $\boldsymbol{\beta} = (\beta_0, \beta_1)$ and the unknown coefficients $\beta_0 \in \mathbb{R}$ and $\beta_1 \in \mathbb{R}$ are the intercept and the strength of the influence field, respectively. If $\beta_1 < 0$, $\boldsymbol{x}$ affects the intensity of Y negatively and the influence field $C(s; \boldsymbol{\theta}_I, \boldsymbol{x})$ can be interpreted as a thinning of the LGCP process with intensity $\Lambda(s) = \exp(\beta_0 + Z(s))$. If, however, $\beta_1 > 0$, $\boldsymbol{x}$ has a positive effect on the intensity of Y and there are more points of Y in areas with a high value of $C(s; \boldsymbol{\theta}_I, \boldsymbol{x})$. Furthermore, $Z := \{Z(s) : s \in \mathbb{R}^2\}$ is a zero-mean stationary Gaussian random field with a covariance function $C_Z(r; \boldsymbol{\theta}_Z)$ and independent of the influence field. In our application below, we use the Matérn covariance function

$$C_Z(r; \boldsymbol{\theta}_Z, \nu) = \sigma_Z^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\sqrt{2\nu} \frac{r}{\rho_Z} \right)^{\nu} K_\nu \left(\sqrt{2\nu} \frac{r}{\rho_Z} \right), \quad r > 0, \tag{4}$$

with the smoothness parameter $\nu = 2$ and $\boldsymbol{\theta}_Z = (\sigma_Z^2, \rho_Z)$, where σ_Z^2 and ρ_Z are the variance and range parameters, respectively, and K_ν is the modified Bessel function of the second kind (e.g. Cressie 1993; Chilés and Delfiner 1999; Banerjee et al. 2004). The choice $\nu = 2$ was made since we expect that the unobserved environmental conditions that affect the clustering of $\boldsymbol{y}$ in our application vary rather smoothly and since it is computationally convenient (Lindgren et al. 2011).

2.3 Replicates

Assume that we have several independent replicated point patterns $\boldsymbol{y}_k$, $k = 1, \ldots, N$, from the conditional distribution of the point process Y given $X = \boldsymbol{x}_k$, $k = 1, \ldots, N$. Conditionally on $X = \boldsymbol{x}_k$, the model for $\boldsymbol{y}_k$ is a LGCP with the intensity $\Lambda(s; \beta_0, \beta_1, \boldsymbol{\theta}_I, \boldsymbol{x}_k, Z_k)$ in (3), where Z_k, $k = 1, \ldots, N$, are independent replicates of the Gaussian random field with parameters $\boldsymbol{\theta}_Z$. For our data, it is not reasonable to assume that all replicates have the same β_0, which controls the number of points of Y, and we let each pattern $\boldsymbol{y}_k$ have its own intercept parameter β_0, i.e. β_{0k} for $\boldsymbol{y}_k$, $k = 1, \ldots, N$. Consequently, in our application below, the pattern $\boldsymbol{y}_k$ is assumed to be a realization of the LGCP model with the intensity $\Lambda(s; \beta_{0k}, \beta_1, \boldsymbol{\theta}_I, \boldsymbol{x}_k, Z_k)$.

3 Inference

The likelihood of the conditional LGCP model for a point pattern $\boldsymbol{y}$ with n points observed in W is

$$p(\boldsymbol{y}; \boldsymbol{\beta}, \boldsymbol{\theta}_I, \boldsymbol{\theta}_Z, \boldsymbol{x}) = \mathbf{E}_{\boldsymbol{\theta}_Z} \prod_{i=1}^{n} \Lambda(y_i; \boldsymbol{\beta}, \boldsymbol{\theta}_I, \boldsymbol{x}, Z) \exp\left(-\int_W \Lambda(u; \boldsymbol{\beta}, \boldsymbol{\theta}_I, \boldsymbol{x}, Z)\mathrm{d}u\right), \tag{5}$$

where $\boldsymbol{\beta}, \boldsymbol{\theta}_I, \boldsymbol{\theta}_Z$ are the model parameters, Z denotes the Gaussian random field and the expectation is over Z given $\boldsymbol{\theta}_Z$. As we use Bayesian inference we need to be able to evaluate the likelihood (5) efficiently. Below, we describe the approximations needed: discretization of the observation window (Sect. 3.1), an edge-corrected influence field (Sect. 3.2), and approximations related to the Gaussian field (Sect. 3.3), which include approximating the field by a Gaussian Markov random field and using the Laplace approximation to evaluate the likelihood. Finally, the approximated likelihood based on replicates is given in Sect. 3.4 and the MCMC algorithm is described in Sect. 3.5.

3.1 Discretization

To be able to make inference on LGCP models, the observation window W of the point pattern $\boldsymbol{y}$ is discretized using a regular grid in a similar manner as in Rue et al. (2009) and Møller et al. (1998). Namely, the observation window W is divided into G disjoint cells $\{w_g\}$ with center locations ξ_g and area A. Furthermore, we let n_g^y denote the number of observations $\boldsymbol{y}$ within w_g in W and $\mathbf{n}^y = (n_1^y, \ldots, n_G^y)$. A piecewise constant approximation is used for the Gaussian field Z and the competition field C and the locations of $\boldsymbol{y}$ are replaced by the counts n_g^y. The approximate likelihood for $\mathbf{n}^y$ is

$$p(\mathbf{n}^y; \boldsymbol{\beta}, \boldsymbol{\theta}_I, \boldsymbol{\theta}_Z, \boldsymbol{x}) = \mathbf{E}_{\boldsymbol{\theta}_Z} p(\mathbf{n}^y; \boldsymbol{\beta}, \boldsymbol{\theta}_I, \boldsymbol{x}, Z^D), \tag{6}$$

where

$$p(\mathbf{n}^y; \boldsymbol{\beta}, \boldsymbol{\theta}_I, \boldsymbol{x}, Z^D) = \prod_{g=1}^{G} \mathrm{Pois}(n_g^y; \Lambda_g),$$

$\Lambda_g = A \exp(\beta_0 + \beta_1 C^D(\xi_g; \boldsymbol{\theta}_I, \boldsymbol{x}) + Z^D(\xi_g; \boldsymbol{\theta}_Z))$, and C^D and Z^D are the piecewise constant approximations of C and Z.

3.2 Edge correction

The large tree process X is only partially observed, and generating the influence field only based on the observed large trees would result in too weak influence near the borders. Therefore, we propose an imputation type approach, similar to the one proposed by Kühlmann-Berenzon et al. (2005), to correct for the unobserved points of X. Specifically, we propose to replace the influence generated by the unobserved trees with the expected influence generated assuming that the whole process X is an independently marked homogeneous Poisson process. In the unmarked case, X is assumed to be a homogeneous Poisson process. In general, the point pattern outside the window would depend on the pattern inside the window, but this is not the case for the Poisson process.

Let λ and F be the intensity and mark distribution of X, and X_{W^c} the restriction of X to W^c, the complement of W. Using the Campbell theorem (e.g. Chiu et al. 2013) we can write

$$\begin{aligned}\mathbf{E}C(s;\boldsymbol{\theta}_I, X_{W^c}) &= \mathbf{E}\sum_{[x_j,m_j]\in X_{W^c}} c(\|s-x_j\|, m_j;\boldsymbol{\theta}_I)\\ &= \int_{R_+}\int_{R^2} c(s-x, m;\boldsymbol{\theta}_I)\mathbf{1}_{W^c}(x)\lambda\,\mathrm{d}x\,\mathrm{d}F(m),\end{aligned}$$

where $\mathbf{1}_{W^c}$ is the indicator function of the set W^c, i.e. $\mathbf{1}_{W^c}(x) = 1$ if $x \in W^c$, and 0 otherwise. By changing the order of the integrals we find that

$$\begin{aligned}\mathbf{E}C(s;\boldsymbol{\theta}_I, X_{W^c}) &= \int_{R^2} f(s-y)\mathbf{1}_{W^c}(y)\lambda\,\mathrm{d}y\\ &= \int_{R^2} f(s-y)\lambda\,\mathrm{d}x - \int_{R^2} f(s-x)\mathbf{1}_W(x)\lambda\,\mathrm{d}x,\end{aligned}$$

where $f(x) = \int_{R_+} c(\|x\|, m;\boldsymbol{\theta}_I)\,\mathrm{d}F(m)$. By changing to polar coordinates and with a slight abuse of notation

$$\int_{R^2} f(s-x)\lambda\,\mathrm{d}x = \lambda 2\pi\int_0^\infty rf(r)\,\mathrm{d}r,$$

which can be computed using numerical integration. Since we are only interested in locations $s \in W$, we can replace the function f with $f\mathbf{1}_{W^S}$, the restriction of f to the set $W^S = \{s-x : s \in W, x \in W\}$, and

$$\int_{R^2} f(s-x)\mathbf{1}_W(x)\,\mathrm{d}x = \int_{R^2}(f\mathbf{1}_{W^S})(s-x)\mathbf{1}_W(x)\lambda\,\mathrm{d}x = (f\mathbf{1}_{W^S} * \mathbf{1}_W)(s).$$

The discrete convolution of the piecewise constant approximations of $f\mathbf{1}_{W^S}$, and $\mathbf{1}_W$ can be efficiently computed using discrete Fourier transforms (Oppenheim et al. 1999; Frigo and Johnson 2005). For F, we use the empirical distribution of marks in the sample plot under study.

The edge-corrected influence field value at any location $s \in W$ is then obtained as the sum of the influence field calculated from the observed $\boldsymbol{x}_W$, $C(s; \boldsymbol{\theta}_I, \boldsymbol{x}_W)$, and the expected influence load of the unobserved X_{W^c}. In general, we use the numerical approximation explained above, but for the special case of the Gaussian influence kernel (1) and a rectangular observation window, it is easy to compute the edge correction by hand.

3.3 Approximations related to the Gaussian field

We use Laplace approximation (Tierney and Kadane 1986; Rue et al. 2009) to approximate the likelihood (6) and obtain

$$\mathbf{E}_{\boldsymbol{\theta}_Z} p(\mathbf{n}^y; \boldsymbol{\beta}, \boldsymbol{\theta}_I, \boldsymbol{x}, Z^D) \approx \sqrt{\frac{(2\pi)^d}{\det(-\mathbf{H}(\hat{\mathbf{z}}))}}\, p(\mathbf{n}^y; \boldsymbol{\beta}, \boldsymbol{\theta}_I, \boldsymbol{x}, \hat{\mathbf{z}}) p(\hat{\mathbf{z}}; \boldsymbol{\theta}_Z), \qquad (7)$$

where $\mathbf{H}$ and $\hat{\mathbf{z}}$ are the Hessian and maximizer of $\log p(\mathbf{n}^y; \boldsymbol{\beta}, \boldsymbol{\theta}_I, \boldsymbol{x}, \mathbf{z}) p(\mathbf{z}; \boldsymbol{\theta}_Z)$, respectively, and $p(\mathbf{z}; \boldsymbol{\theta}_Z)$ is the probability density of the vector $\mathbf{Z}^D$ which contains the values of Z^D at grid cells.

Since the Gaussian random field Z is assumed to have mean zero and the Matérn covariance function (4) with $\nu = 2$, we can utilize the explicit link between Gaussian fields and Markov random fields (Lindgren et al. 2011), which tells us that the distribution of $\mathbf{Z}^D$ should be approximated with a Gaussian distribution with a precision matrix given by Lindgren et al. (2011).

3.4 Replicates

Since the point patterns are assumed to be conditionally independent, the likelihoods (5) for each replicate $\boldsymbol{y}_k$ can be multiplied to yield the final likelihood

$$p(\boldsymbol{y}_1, \ldots, \boldsymbol{y}_N; \boldsymbol{\beta}, \boldsymbol{\theta}_I, \boldsymbol{\theta}_Z, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_N) = \prod_{k=1}^{N} p(\boldsymbol{y}_k; \beta_{0k}, \beta_1, \boldsymbol{\theta}_I, \boldsymbol{x}_k), \qquad (8)$$

where now $\boldsymbol{\beta}$ contains all the regression coefficients, i.e. $\boldsymbol{\beta} = (\beta_{01}, \ldots, \beta_{0N}, \beta_1)$. To obtain an approximation of (8), the approximations (6) and (7) are applied to each pattern separately.

3.5 MCMC

Combining the likelihood (8) with the prior $p(\boldsymbol{\beta}, \boldsymbol{\theta}_I, \boldsymbol{\theta}_Z)$ yields the approximate posterior distribution. To sample from this distribution, we use Robust Adaptive Metropolis algorithm (Vihola 2012, 2020), which uses a Gaussian random-walk proposal distribution, whose covariance is updated adaptively. The limiting proposal covariance matches the shape of the posterior, such that an average acceptance rate of 0.234 is attained, following the theoretical findings presented e.g. in Roberts et al. (1997).

4 Simulation experiment

We made a simulation experiment to study the performance of the inference approach and the edge correction method suggested above. The point pattern $\boldsymbol{x}$ was a realization of either a Poisson process or a regular Strauss process. The Strauss process (e.g. Illian et al. 2008) was included to see whether the edge correction based on the Poisson assumption of X would work even in a more regular case. We did not include any cluster process since in our application, the large tree patterns $\boldsymbol{x}$ are regular. Also, based on a small simulation study (results not shown here), it is unlikely that the Poisson correction would work well when the $\boldsymbol{x}$ pattern is strongly clustered. We did not include marks in the simulation experiment.

4.1 Set-up

The intensity parameters of the Poisson and Strauss processes were chosen such that they result in approximately 60 points in the observation window $W = [0, 40] \times [0, 40]$. In the Strauss process (parametrized as in Baddeley et al. 2015), the intensity related parameter was 0.06, the interaction strength 0.1, and the interaction radius 2, making the resulting patterns rather regular. The $\boldsymbol{y}$ patterns were generated on W and the $\boldsymbol{x}$ patterns on the extended window $W_{\text{ext}} = [-20, 60] \times [-20, 60]$ to be able to use plus sampling which represents the ideal situation where no imputation is needed as the complete pattern is known. The Gaussian kernel (1) was used as the influence kernel. Initially, the parameters of the competition field and of the Gaussian field were set to the estimates found in Sect. 5 and the intercept β_0 was chosen such that the resulting LGCP model would have 600 points on average. First we used the estimated values $\beta_1 = -0.7$ and $\theta = 2.1$, called "estimated" in Fig. 2. In addition, we used either the values $\beta_1 = -3$, and $\theta = 2.1$ corresponding to a much stronger effect of the influence kernel (β_1) ("strong" in Fig. 2) or the values $\beta_1 = -0.7$, and $\theta = 6$ corresponding to a much larger range of influence θ ("wide" in Fig. 2) than in the data. In all cases, $\sigma_Z = 1.6$ and $\rho_Z = 2.6$. We generated 100 replicates of each X process and one $\boldsymbol{y}$ pattern for each $\boldsymbol{x}$. The random intensity of the Cox process was approximated by a piecewise constant function using 0.1 m × 0.1 m cells.

We fitted the conditional LGCP model to the simulated point patterns. We discretized the observation windows to pixels of size 1 m × 1 m and set weakly informative independent priors for all model parameters as follows: For the parameters in $\boldsymbol{\beta}$, we used Gaussian distributions with mean zero and standard deviation 10. For the range parameters ρ_Z and θ, very small and very large values do not make sense based on the discretization and window used. Thus, we set the prior to be the Gamma distribution with shape parameter 2.4 and scale parameter 1.8, implying that approximately 90% of the prior probability is between 1 m and 10 m. Furthermore, the prior for the standard deviation of the Gaussian field σ_Z was the exponential distribution with expectation 10, slightly favoring small values.

For each point pattern, we then ran the MCMC scheme using (a) no edge correction, (b) the Poisson edge correction and (c) plus sampling edge correction with 100,000 updates using the true parameter values as the starting values. For each chain we

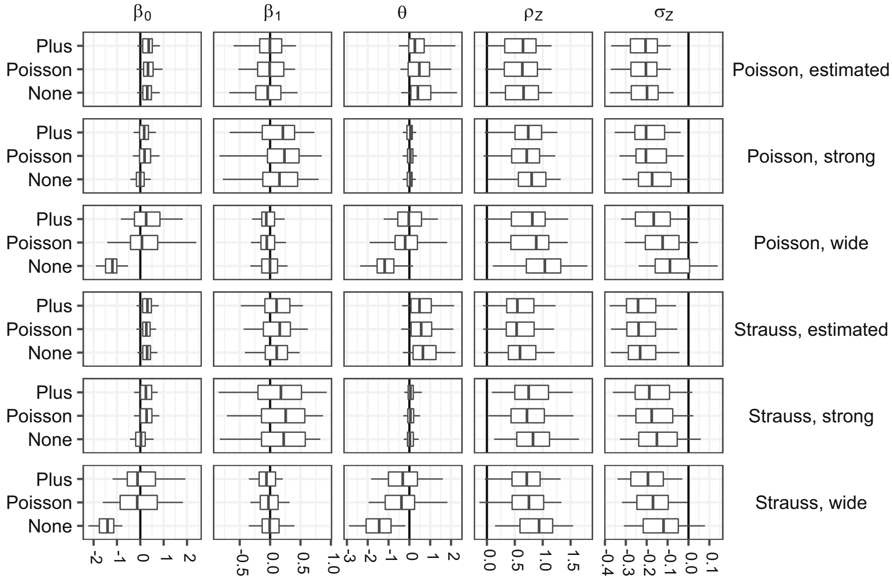


Fig. 2 Quantiles (0.05, 0.25, 0.5, 0.75, 0.95) of differences between posterior means and reference values. For each row of the figure, we display the X process and the competition effect on the right and within each subfigure, we label the three different edge corrections (left). The quantiles are based on 100 replicates

discarded 20,000 first samples as burn-in and saved every 10th sample. When influence was strong, most chains converged and mixed well. However, there were problems with mixing if the influence was not so strong. In this case the effective sample size was estimated to be less than 1000 in half of the chains. Upon closer inspection multi-modality was often the cause. We used posterior means of each chain in the comparisons. Using posterior modes led to identical conclusions.

4.2 Results

First, we investigated the performance of the Bayesian inference approach. To avoid edge effects, we estimated the parameters using plus sampling, utilizing the true pattern $\boldsymbol{x}$ in the extended window. Based on the distributions of the posterior means for the plus sampling method (see Plus in Fig. 2), we can see that the Bayesian MCMC approach with the approximations used performed reasonably well for the main parameters β_1 and θ. However, the less interesting random field parameters were clearly biased. As expected, the distribution of the X pattern did not affect the performance of the inference.

Second, we investigated the performance of the Poisson edge correction. An example of the expected intensity field with and without edge correction for the conditional LGCP model with the parameters estimated from the EVO02 pattern and Strauss pattern $\boldsymbol{x}$ is shown in top row of Fig. 3. It can be seen that the Poisson corrected and the plus sampling corrected intensities are quite similar to each other. The bottom row of Fig. 3 further shows the components of the influence field for the Poisson correction,

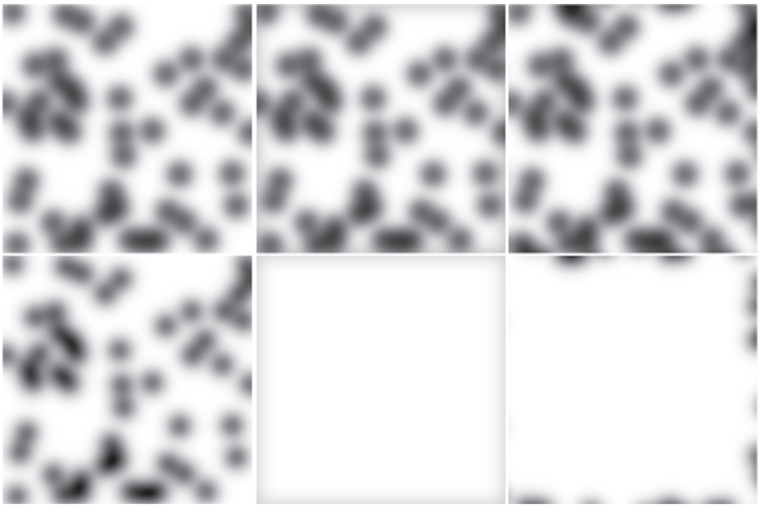


Fig. 3 Top row: expected intensity of the conditional LGCP with parameters estimated from the EVO02 pattern using no edge correction (left), Poisson correction (middle) and plus sampling (right). Bottom row: influence field induced by the observed points (left), expected influence field caused by the unobserved points under the Poisson assumption (middle) and influence field caused by the unobserved points (right). The $\boldsymbol{x}$ pattern is a realization of a Strauss process with interaction parameter 0.1, interaction range 2, and with on average 60 points. Dark color means low intensity/high influence

namely the contribution of the observed points (left) and the expected contribution of the unobserved points under the Poisson assumption (middle). The contribution of unobserved points is shown for comparison (right). The Poisson correction simply approximates the contribution of the unobserved points.

To assess the performance of the proposed edge correction method, we compared the posterior means of the model parameters β_0, β_1 and θ, obtained by using plus sampling to the estimates obtained by using the Poisson correction and those obtained by using no edge correction. The distribution of the posterior means is shown in Fig. 2. It can be seen that the estimates of the different methods are very similar when the influence of the large trees was not too wide, for both X processes. However, when the influence was wide, the proposed Poisson correction produced estimates that were closer to the plus sampling based estimates than the uncorrected estimates were. The results were altogether very similar for the Poisson and Strauss processes. Thus, the edge correction plays a role if the range of influence of the $\boldsymbol{x}$ points on the intensity of Y is wide.

5 Application

The data shown in Fig. 1 have been collected on 40 m × 40 m squares in southern Finland. They are part of a larger data set collected for studies on tree and stand development in managed, uneven-aged Norway spruce forests conducted under the ERIKA research project at the Natural Resources Institute Finland (Eerikäinen et al.

2007; Eerikäinen et al. 2014; Saksa and Valkonen 2011). Using the conditional LGCP model, we studied the effect of large trees $\boldsymbol{x}_i$ (black circles) to the seedling patterns $\boldsymbol{y}_i$ (red crosses). The patterns $\boldsymbol{x}_i$ consist of trees which had a vital crown with no damages and with a dbh at least 7 cm in 1991. Most trees (78% of trees, 70% basal area) were Norway spruces and the remaining ones either Scots pines or broadleaves. The seedlings were naturally generated with height at least 10 cm in 1996 and had reached this height after the data collection in 1991. The seedlings were mostly Norway spruces (98%).

We fitted the conditional LGCP model using different mark dependent and mark independent Gaussian influence fields: the full mark dependent model (2), the two reduced models where either of the mark specific parameters, namely δ or α were set to zero, the mark independent model (1), and a model without an influence field. The mark was always the dbh. We used the same discretization of the observation window (1 m × 1 m pixel size) and the same priors as in the simulation experiment (Sect. 4.1). The pixel size 1 m × 1 m was chosen because variations in smaller units are practically unimportant in forests. The priors for α and δ were both the exponential distribution with expectation 10. We then ran the MCMC scheme using the Poisson edge correction with 120,000 updates, leaving out the first 20,000 observations of the chains as the burn-in.

To compare the models, we used the posterior predictive model assessment based on various summary characteristics, namely the L-function (variance stabilizing version of Ripley's K), the empty space function F, and the nearest neighbour distribution function G summarizing the spatial pattern $\boldsymbol{y}$ and, to investigate the relationship between the large trees and seedlings, the cross L-function, L_{12} (e.g. Illian et al. 2008; Diggle 2013). We used the standard estimators of these functions with translational (L, L_{12}) and Kaplan-Meier edge correction methods (F, G) (Baddeley and Gill 1997). For each plot, we generated 10,000 patterns of seedlings from the posterior predictive distributions of the conditional LGCP models given the observed $\boldsymbol{x}$ and calculated the summary functions for the data and for each of the generated patterns. The posterior predictive simulations were made using a discretization with 0.2 m × 0.2 m cell size.

Figure 4 shows the empirical L_{12} functions together with the 95% global extreme rank length envelopes (Myllymäki et al. 2017; Myllymäki and Mrkvička 2020) constructed from the L_{12} summary functions of the simulations of the fitted model with mark independent influence kernel (1) (shaded region), mark dependent influence kernel (2) (dotted lines), and no influence kernel (dashed line) separately for each plot. The observed L_{12} function is distinctly better covered by the envelopes based on the models with influence field than without. While the envelopes of the model without an influence field are centred around zero, i.e., no interaction between trees and seedlings, the empirical L_{12} functions have the tendency to go below zero in most plots, indicating repulsion or inhibition of trees and seedlings, and the envelopes of the models with influence kernels are shifted downwards as well. The difference between the two models with influence kernels is, however, minor. Other summary functions (L, F, G) produced very similar envelopes regardless of the type or lack of influence field, see figures in Appendix 1. The empirical functions were inside the envelopes, except the nearest neighbor distance distribution functions of four sample plots VES07, VES13,

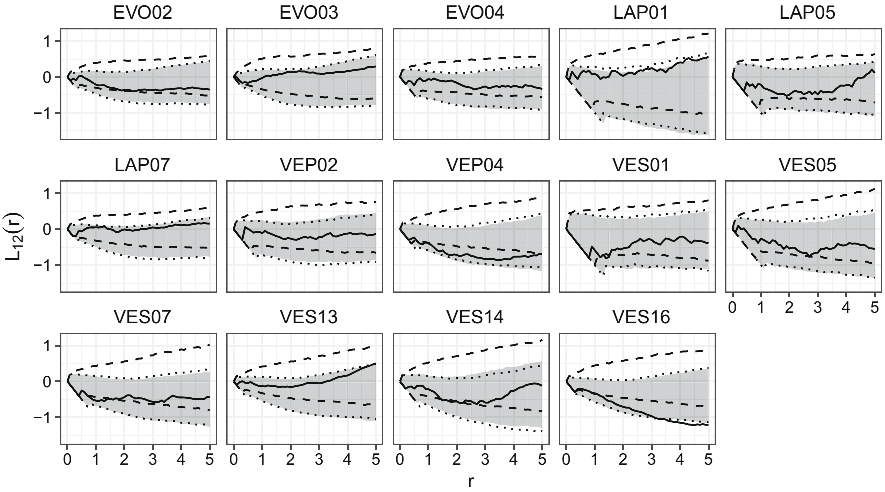


Fig. 4 Empirical L_{12} functions (solid line) together with the 95% global envelopes constructed from 10,000 simulations from the posterior predictive distribution of the fitted conditional LGCP models for the 14 plots in Fig. 1 with mark independent (1) (grey shade), mark dependent (2) (dotted lines), and no (dashed lines) influence

VES14 and VES16, which were slightly outside the envelopes at distances less than 1 m, i.e. less than the pixel size used in the discretization. This may suggest that the spatial distribution of the seedlings is not Poisson at a very small scale, but we did not investigate this further.

The envelopes for the models with mark dependent kernels with either δ or α set to zero are omitted because they were very similar to the envelopes of the other two influence kernels.

Based on the analysis above, it is clear that an influence kernel is needed. However, since all the models with an influence kernel fitted the data equally well, we report the results of the simplest model (1). The marginal posterior distributions of the model parameters of this model are shown in Fig. 6. The influence of the large trees on the seedlings (β_1) is clearly negative meaning that the seedlings avoid locations in the close vicinity of the large trees. The range of influence θ of the large trees was estimated to be around 2.1 m, indicating that the influence of a large tree decreases from its maximum influence (at the tree location) to 37% of it at distance 2.1 m from the tree, or to 5% of it at distance 3.6 m. However, there is a lot of unexplained variability, as the quite wide envelopes in Fig. 4 and Appendix 1 show.

Figure 5 shows for each plot one realization drawn from the posterior predictive distribution of the model with mark independent influence kernel. It is difficult to detect the relationship between trees and seedlings by eye, but one can compare the clustering of the seedling patterns to the observed patterns (Fig. 1). The patterns in Figs. 1 and 5 look rather similar, and according to the envelope tests (see Fig. 4 and Appendix 1) the model captures small scale structures up to 5 m distances.

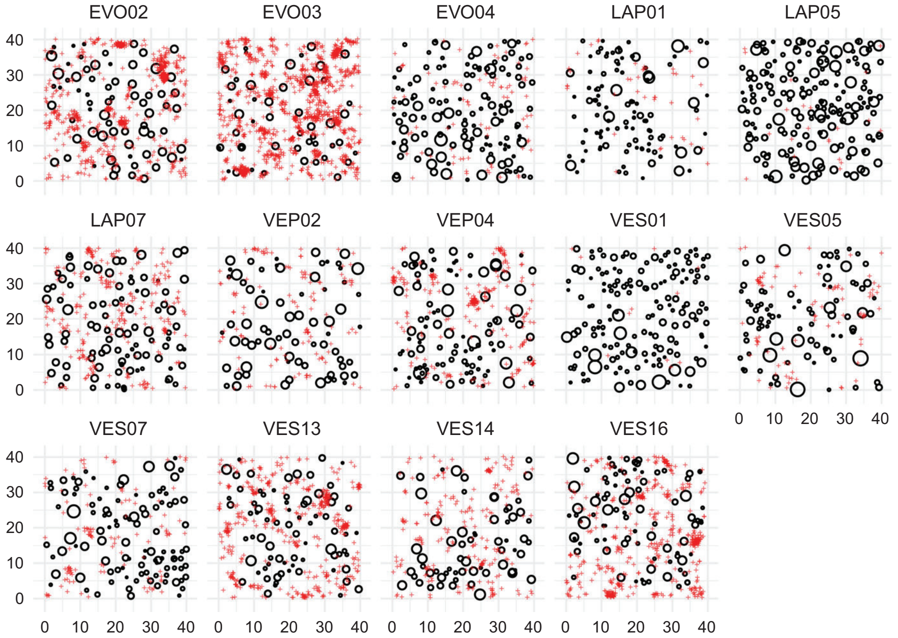


Fig. 5 Simulated seedling pattern (red crosses) and observed large trees (black circles, radius relative to dbh). The simulation was done using the posterior predictive distribution of the fitted conditional LGCP models for the 14 plots in Fig. 1 with mark independent influence of large trees

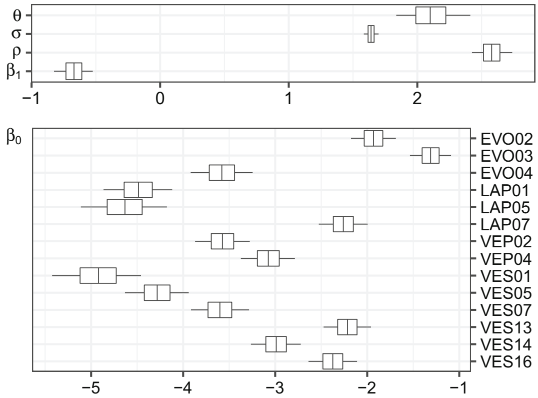


Fig. 6 Posterior quantiles (0.05, 0.25, 0.5, 0.75, 0.95) of the common parameters (top) and the sample plot specific intercepts β_0 (bottom)

6 Discussion

We proposed a LGCP model to investigate the effect of large trees on the intensity of seedlings under the presence of unexplained clustering. The influence of large trees was modeled by using parametric influence kernels around them. Our analysis suggests that tree regeneration is affected by the pattern of large trees in the studied data. Namely, the large trees were found to have negative effect on the seedling density in the vicinity of large trees. Further, the LGCP model could capture much of the unexplained clustering. For parameter estimation, we constructed a Bayesian approach using MCMC and Laplace approximation. All computations were implemented in Julia language (Bezanson et al. 2017), while graphics were done using ggplot2 (Wickham 2016).

Estimation of the influence field parameters worked well in our simulation experiment and we did not observe any problems due to possible confounding between the influence field and the spatial random effect as reported in the literature (e.g. Dupont et al. 2020). However, the random field parameter estimates were biased. We suspect that this is caused by weak identifiability (cf. Anderes 2010; Zhang 2004) or discretization bias coupled with the Laplace approximation. We used replicates to help with the weak identifiability which was necessary for the plots with very few seedlings. In our further experiments with finer discretizations (results not shown), we observed issues with the approximation. In particular, when the number of points per cell was small, the approximate posterior appeared to degenerate. We are unaware of exact inference methods that would be feasible in our setting, but we are currently investigating new methods that could allow for more detailed investigation of this issue.

There are many alternative approaches to inference with log Gaussian Cox processes. For example, the R package INLA (Rue et al. 2009) uses Laplace approximation in a similar fashion as we did, but is somewhat restricted to linear models. Indeed, INLA can in principle accommodate our model using the `rgeneric` class (personal communication with Håvard Rue). However, we faced some computational difficulties in estimation. The R package lgcp (Taylor et al. 2015) uses MALA algorithm for efficient Bayesian inference for the full model including the latent field. The use of full MCMC might lead to better estimation of the random field parameters. However, the lgcp package is also restricted to linear models, whereby we were not able to apply it directly to our model. For Stan (Stan Development Team 2018) our random field model appears to be too complicated, however there are some recent advances see e.g. Margossian et al. (2020). Also, inlabru (Bachl et al. 2019) could be further investigated.

Since the large trees outside the sampling window may affect the intensity of the seedlings within the window, an edge correction assuming that the large trees were from a Poisson process was included in the estimation procedure. We demonstrated by a small simulation study that this edge correction can work well even when the large trees are from a regular process. Compared to no edge correction, it improved the parameter estimates when the range of influence was rather wide.

An obvious alternative strategy would be to include the locations of the unobserved trees outside the observation window to the MCMC estimation in a similar manner as considered in the inference for Neyman-Scott point processes (Møller and Waagepetersen 2004). This approach would allow incorporating prior information on the large tree process in the edge correction at the cost of increased complexity. Since we did not have important prior information and the effect of edge correction appeared minor, we did not explore this approach further. Ideas from Geyer (1999) or Gabriel et al. (2017) could be used to find further alternative edge correction methods.

Our proposed edge correction method can be efficiently implemented when the influence field is constructed as the sum of individual signals. In principle, a similar edge correction could be applied with different combination rules, such as product (e.g. Wu et al. 1985; Miina and Pukkala 2002) or max-fields (e.g. Penttinen and Niemi

2007). If also the influence kernel is binary, e.g. $c(h;\theta) = \mathbf{1}(h \leq \theta)$, then the max-field is split into two phases as well, namely influence and influence-free zones. However, we note that our proposed calculation of the expected influence of trees outside the observation window W does not generalize directly to other combination rules.

The models introduced in this paper could be useful even for natural (e.g. Abellanas and Pérez-Moreno 2018) or urban forests (Hauru et al. 2012). Furthermore, they could be used in an experimental setup, where realizations of seedlings would be generated for different large tree patterns and the success of regeneration evaluated by some spatial summary functions such as the empty space function. In a similar manner, the effect of different thinning strategies on the regeneration of trees could be evaluated.

It could be argued that, since the management was the same for all plots and the geographical differences minor, the plots should have had a common intercept parameter. However, this was clearly not the case due to large variation in numbers of seedlings from plot to plot. Since we used plot specific intercepts, it could be argued that all other parameters should be plot specific too. This was not possible in practice due to the problems with the random field parameters. We did not explore the alternative where the intercept and the influence field parameters would be plot specific but the random field parameters shared since the envelope tests already suggested adequate fit of the model.

The observed and simulated seedling patterns in the Figures 1 and 5, respectively, are very similar in several aspects while quite different in others. For example, the clusters seemed to be clustered in the VES13 plot. Although the envelope tests suggest that the model was able to capture the variability in the data, it depends on the specific application if the model is adequate. To best of our knowledge, this is the first point process model accounting for clustering of the seedlings in these uneven-aged forests.

There are many other factors than the vicinity of large trees that may affect the intensity of seedlings (Valkonen and Maguire 2005; Kuusinen et al. 2019). Therefore, the model could be further improved and unexplained variability decreased if some covariate information on local conditions within plots would be available to be included in the model. Further, plot level covariate effects could be added to the model in order to explain the numbers of seedlings in different plots. Finally, we modeled the influence of large trees as a function of the dbh, whose effect on the influence was, however, minor in our data. Other possibly useful marks could be the height, crown ratio or crown width of the tree, for example.

Acknowledgements The authors thank Helena Henttonen, Jari Hynynen and Sauli Valkonen (Luke) for discussions on the application, and Antti Penttinen for his comments on an earlier version of the manuscript. Further, the authors are grateful to Hilkka Ollikainen and Juhani Korhonen, who mastered the maintenance and measurements on the plots of the ERIKA data set. The authors wish to acknowledge CSC – IT Center for Science, Finland, for computational resources. We are grateful to the two anonymous reviewers for their constructive comments.

Funding Open access funding provided by Natural Resources Institute Finland (LUKE). MK, MM and MV were financially supported by the Academy of Finland (Project Numbers 306875, 327211, 295100 and 315619) and AS by the Swedish Research Council (VR 2018-03986).

Data availability The datasets analysed during the current study are available from the corresponding author on reasonable request.

Declarations

Conflict of interest The authors have no relevant financial or non-financial interests to disclose.

Code availability The custom code is available from authors on request.

Appendix: Envelopes

See Appendix Figures 7, 8 and 9.

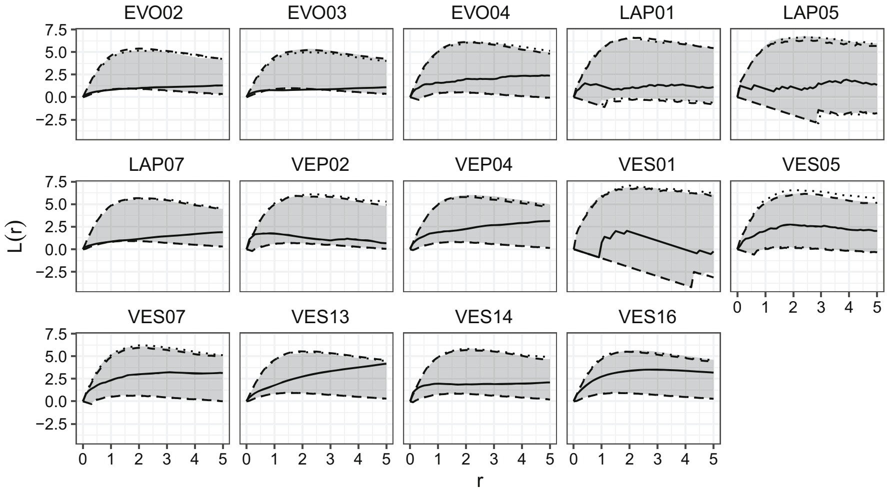


Fig. 7 Empirical *L* functions (solid line) together with the 95% global envelopes constructed from 10,000 simulations from the posterior predictive distribution of the fitted conditional LGCP models for the 14 plots in Fig. 1 with mark independent (grey shade), mark dependent (dashed lines), and no (dotted lines) influence

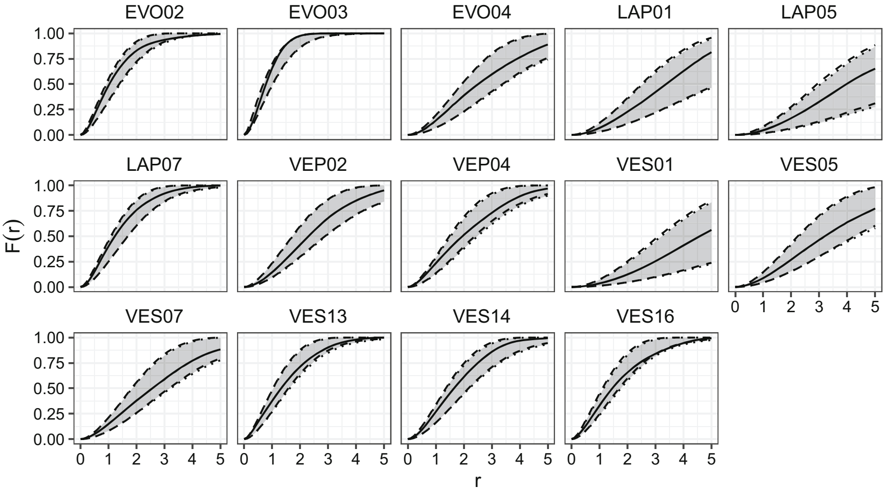


Fig. 8 Empirical empty space functions (solid line) together with the 95% global envelopes constructed from 10,000 simulations from the posterior predictive distribution of the fitted conditional LGCP models for the 14 plots in Fig. 1 with mark independent (grey shade), mark dependent (dashed lines), and no (dotted lines) influence

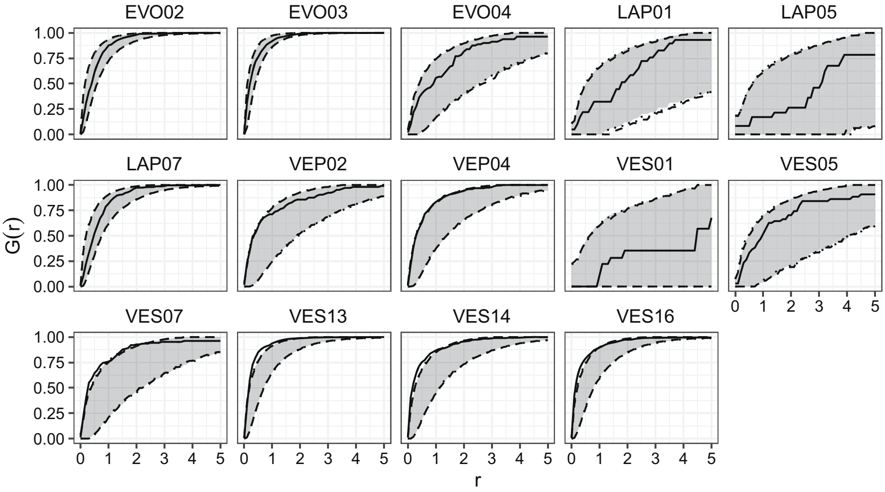


Fig. 9 Empirical nearest neighbor distance distribution functions (solid line) together with the 95% global envelopes constructed from 10,000 simulations from the posterior predictive distribution of the fitted conditional LGCP models for the 14 plots in Fig. 1 with mark independent (grey shade), mark dependent (dashed lines), and no (dotted lines) influence

References

Abellanas B, Pérez-Moreno P (2018) Assessing spatial dynamics of a Pinus nigra subsp. salzmannii natural stand combining point and polygon patterns analysis. For Ecol Manag 424:136–153. https://doi.org/10.1016/j.foreco.2018.04.050

Adler F (1996) A model of self-thinning through local competition. Proc Natl Acad Sci USA 93:9980–9984

Anderes E (2010) On the consistent separation of scale and variance for Gaussian random fields. Ann Stat 38(2):870–893. https://doi.org/10.1214/09-AOS725

Bachl FE, Lindgren F, Borchers DL, Illian JB (2019) inlabru: an R package for Bayesian spatial modelling from ecological survey data. Methods Ecol Evol 10:760–766. https://doi.org/10.1111/2041-210X.13168

Baddeley A, Gill RD (1997) Kaplan-Meier estimators of distance distributions for spatial point processes. Ann Stat 25(1):263–292

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

Banerjee S, Carlin BP, Gelfand AE (2004) Hierarchical modeling and analysis for spatial data, 1st edn. Chapman & Hall/CRC, Boca Raton

Bezanson J, Edelman A, Karpinski S, Shah VB (2017) Julia: a fresh approach to numerical computing. SIAM Rev 59(1):65–98. https://doi.org/10.1137/141000671

Chilés JP, Delfiner P (1999) Geostatistics: modeling spatial uncertainty. Wiley, New York

Chiu SN, Stoyan D, Kendall WS, Mecke J (2013) Stochastic geometry and its applications, 3rd edn. Wiley, Chichester

Cressie NAC (1993) Statistics for spatial data, revised edn. Wiley series in probability and mathematical statistics, Wiley, New York

Dieckmann U, Law R, Metz J (2000) The geometry of ecological interactions: simplifying spatial complexity. Cambridge studies in adaptive dynamics. Cambridge University Press, Cambridge. https://doi.org/10.1017/CBO9780511525537

Diggle PJ (2013) Statistical analysis of spatial and spatio-temporal point patterns, 3rd edn. CRC Press, Boca Raton

Dupont E, Wood SN, Augustin N (2020) Spatial+: a novel approach to spatial confounding. arXiv:2009.09420 [stat.ME]

Eerikäinen K, Valkonen S, Saksa T (2014) Ingrowth, survival and height growth of small trees in uneven-aged picea abies stands in southern Finland. For Ecosyst 1:5. https://doi.org/10.1186/2197-5620-1-5

Eerikäinen K, Miina J, Valkonen S (2007) Models for the regeneration establishment and the development of established seedlings in uneven-aged, Norway spruce dominated forest stands of southern Finland. For Ecol Manag 242(2):444–461. https://doi.org/10.1016/j.foreco.2007.01.078

Frigo M, Johnson SG (2005) The design and implementation of FFTW3. Proc IEEE 93(2):216–231. (special issue on “Program Generation, Optimization, and Platform Adaptation”)

Gabriel E, Coville J, Chadœuf J (2017) Estimating the intensity function of spatial point processes outside the observation window. Spat Stat 22:225–239. https://doi.org/10.1016/j.spasta.2017.07.008

Genet A, Grabarnik P, Sekretenko O, Pothier D (2014) Incorporating the mechanisms underlying inter-tree competition into a random point process model to improve spatial tree pattern analysis in forestry. Ecol Model 288:143–154. https://doi.org/10.1016/j.ecolmodel.2014.06.002

Geyer C (1999) Likelihood inference for spatial point processes: Likelihood and computation. In: Kendall W, Barndroff-Nielsen O, van Lieshout M (eds) Stochastic geometry. Chapman and Hall/CRC, Boca Raton, pp 141–172

Grabarnik P, Särkkä A (2009) Modelling the spatial structure of forest stands by multivariate point processes with hierarchical interactions. Ecol Model 220(9–10):1232–1240. https://doi.org/10.1016/j.ecolmodel.2009.02.021

Häbel H, Myllymäki M, Pommerening A (2019) New insights on the behaviour of alternative types of individual-based tree models for natural forests. Ecol Model 406:23–32. https://doi.org/10.1016/j.ecolmodel.2019.02.013

Hauru K, Niemi A, Lehvävirta S (2012) Spatial distribution of saplings in heavily worn urban forests: implications for regeneration and management. Urban For Urban Green 11(3):279–289. https://doi.org/10.1016/j.ufug.2012.03.004

Högmander H, Särkkä A (1999) Multitype spatial point patterns with hierarchical interactions. Biometrics 55(4):1051–1058

Illian J, Penttinen A, Stoyan H, Stoyan D (2008) Statistical analysis and modelling of spatial point patterns, 1st edn. Wiley, Chichester

Illian JB, Møller J, Waagepetersen RP (2009) Hierarchical spatial point process analysis for a plant community with high biodiversity. Environ Ecol Stat 16(3):389–405. https://doi.org/10.1007/s10651-007-0070-8

Kühlmann-Berenzon S, Heikkinen J, Särkkä A (2005) An additive edge correction for the influence potential of trees. Biometr J 47:517–526. https://doi.org/10.1002/bimj.200310157

Kuuluvainen T, Pukkala T (1989) Effect of scots pine seed trees on the density of ground vegetation and tree seedlings. Silva Fennica. https://doi.org/10.14214/sf.a15536

Kuusinen N, Valkonen S, Berninger F, Mäkelä A (2019) Seedling emergence in uneven-aged norway spruce stands in Finland. Scand J Forest Res 34(3):200–207. https://doi.org/10.1080/02827581.2019.1575976

Lindgren F, Rue H, Lindström J (2011) An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. J R Stat Soc 73(4):423–498. https://doi.org/10.1111/j.1467-9868.2011.00777.x

Margossian C, Vehtari A, Simpson D, Agrawal R (2020) Hamiltonian Monte Carlo using an adjoint-differentiated laplace approximation: Bayesian inference for latent Gaussian models and beyond. In: Thirty-fourth conference on neural information processing systems, Morgan Kaufmann Publishers, advances in neural information processing systems, conference on neural information processing systems, NeurIPS. Conference date: 06-12-2020 through 12-12-2020

Miina J, Pukkala T (2002) Application of ecological field theory in distance-dependent growth modelling. For Ecol Manag 161(1):101–107. https://doi.org/10.1016/S0378-1127(01)00489-3

Møller J, Waagepetersen RP (2004) Statistical inference and simulation for spatial point processes, 1st edn. Chapman & Hall/CRC, Boca Raton

Møller J, Syversveen AR, Waagepetersen RP (1998) Log Gaussian Cox processes. Scand J Stat 25(3):451–482. https://doi.org/10.1111/1467-9469.00115

Myllymäki M, Mrkvička T (2020) GET: Global envelopes in R. arXiv:1911.06583 [stat.ME]

Myllymäki M, Mrkvička T, Seijo H, Grabarnik P, Hahn U (2017) Global envelope tests for spatial processes. J R Stat Soc 79:381–404. https://doi.org/10.1111/rssb.12172

Oppenheim AV, Schafer RW, Buck JR (1999) Discrete-time signal processing, 2nd edn. Prentice-Hall Inc, USA

Penttinen A, Niemi A (2007) On statistical inference for the random set generated cox process with set-marking. Biometr J 49(2):197–213. https://doi.org/10.1002/bimj.200610272

Pommerening A, Grabarnik P (2019) Individual-based methods in forest ecology and management, 1st edn. Springer, New York. https://doi.org/10.1007/978-3-030-24528-3

Pommerening A, Maleki K (2014) Differences between competition kernels and traditional size-ratio based competition indices used in forest ecology. For Ecol Manag 331:135–143. https://doi.org/10.1016/j.foreco.2014.07.028

Pommerening A, Sánchez Meador AJ (2018) Tamm review: tree interactions between myth and reality. For Ecol Manag 424:164–176. https://doi.org/10.1016/j.foreco.2018.04.051

Pommerening A, LeMay V, Stoyan D (2011) Model-based analysis of the influence of ecological processes on forest point pattern formation–a case study. Ecol Model 222(3):666–678. https://doi.org/10.1016/j.ecolmodel.2010.10.019

Roberts GO, Gelman A, Gilks WR (1997) Weak convergence and optimal scaling of random walk metropolis algorithms. Ann Appl Probab 7(1):110–120. https://doi.org/10.1214/aoap/1034625254

Rue H, Martino S, Chopin N (2009) Approximate Bayesian inference for latent Gaussian models using integrated nested Laplace approximations (with discussion). J R Stat Soc 71:319–392

Saksa T, Valkonen S (2011) Dynamics of seedling establishment and survival in uneven-aged boreal forests. For Ecol Manag 261(8):1409–1414. https://doi.org/10.1016/j.foreco.2011.01.026

Schneider MK, Law R, Illian JB (2006) Quantification of neighbourhood-dependent plant growth by Bayesian hierarchical modelling. J Ecol 94(2):310–321. https://doi.org/10.1111/j.1365-2745.2005.01079.x

Stan Development Team (2018) Stan modeling language users guide and reference manual, version 2.18.0. http://mc-stan.org/

Taylor BM, Davies TM, Rowlingson BS, Diggle PJ (2015) Bayesian inference and data augmentation schemes for spatial, spatiotemporal and multivariate log-Gaussian Cox processes in R. J Stat Softw 63(7):1–48. https://doi.org/10.18637/jss.v063.i07

Tierney L, Kadane JB (1986) Accurate approximations for posterior moments and marginal densities. J Am Stat Assoc 81(393):82–86. https://doi.org/10.2307/2287970

Valkonen S, Maguire DA (2005) Relationship between seedbed properties and the emergence of spruce germinants in recently cut Norway spruce selection stands in southern Finland. For Ecol Manag 210(1):255–266. https://doi.org/10.1016/j.foreco.2005.02.039

Vihola M (2012) Robust adaptive metropolis algorithm with coerced acceptance rate. Stat Comput 22:997–1008. https://doi.org/10.1007/s11222-011-9269-5

Vihola M (2020) Ergonomic and reliable Bayesian inference with adaptive Markov chain Monte Carlo. In: Piegrorsch WW, Levine R, Zhang HH, Lee TCM (eds) Handbook of computational statistics and data science. Wiley, New York

Wickham H (2016) ggplot2: Elegant graphics for data analysis. Springer-Verlag, New York. https://ggplot2.tidyverse.org

Wu HI, Sharpe PJ, Walker J, Penridge LK (1985) Ecological field theory: a spatial analysis of resource interference among plants. Ecol Model 29(1):215–243. https://doi.org/10.1016/0304-3800(85)90054-7

Zhang H (2004) Inconsistent estimation and asymptotically equal interpolations in model-based geostatistics. J Am Stat Assoc 99(465):250–261. https://doi.org/10.1198/016214504000000241

Mikko Kuronen is a doctoral student. His field of research is spatial statistics and computational methods.

Aila Särkkä is Professor of Mathematical Statistics. She is working on spatial point process statistics, spatio–temporal modeling, and applications in forestry, materials science and neurology.

Matti Vihola is Associate Professor of Statistics. His research interests include computational statistics, in particular Monte Carlo methods for Bayesian inference.

Mari Myllymäki is Academy Research Fellow. Her research topics include spatial and spatio–temporal statistics and simulation-based methods with applications to forestry and life sciences.

IV

TESTING GLOBAL AND LOCAL DEPENDENCE OF POINT PATTERNS ON COVARIATES IN PARAMETRIC MODELS

by

Myllymäki, M., Kuronen, M., and Mrkvička, T.

Spatial Statistics, 42, 100436, 2021.

doi:10.1016/j.spasta.2020.100436

Spatial Statistics 42 (2021) 100436

Contents lists available at ScienceDirect

Spatial Statistics

journal homepage: www.elsevier.com/locate/spasta

Testing global and local dependence of point patterns on covariates in parametric models

Mari Myllymäki [a,*], Mikko Kuronen [a], Tomáš Mrkvička [b]

[a] Natural Resources Institute Finland (Luke), Latokartanonkaari 9, FI-00790 Helsinki, Finland
[b] Department of Applied Mathematics and Informatics, Faculty of Economics, University of South Bohemia, Studentská 13, 37005 České Budějovice, Czech Republic

ARTICLE INFO

Article history:
Received 14 November 2019
Received in revised form 26 February 2020
Accepted 26 February 2020
Available online 3 March 2020

Keywords:
F-statistic
General linear model
Global envelope test
Local dependence
Monte Carlo test
Spatial point pattern

ABSTRACT

Testing for a covariate effect in a parametric point process model is usually done through the Wald test, which relies on an asymptotic null distribution of the test statistic. We propose a Monte Carlo version of the test that also allows local investigation of the covariate effect in the globally fitted model. Two different test statistics are suggested for this purpose: the first, a spatial statistic computed at every location of the observation window, resembles the classical F-statistic that is usually used in general linear models (GLMs) to express the distance between a model and its sub model. This statistic allows one to detect locations where the smoothed point process residuals are reduced by adding the interesting covariates into the model. The second spatial statistic tries to capture local improvements in the shape of the predicted intensity caused by an interesting, continuous covariate. A simulation scheme resembling the permutation inference for GLMs is used to obtain the null distribution of the statistics. Thereafter, a Monte Carlo test with graphical interpretation (a global envelope test) is applied to the empirical and simulated statistic fields to determine the global significance of the covariate and the spatially significant areas. We study the empirical significance level and power of the test in different scenarios and, by applying the test to simulated and real point pattern data, show that the proposed statistics can be valuable for model construction.

---

[*] Corresponding author.
E-mail address: mari.myllymaki@luke.fi (M. Myllymäki).

https://doi.org/10.1016/j.spasta.2020.100436

1. Introduction

A wide range of spatial point process models have been developed to model real data in applications. Furthermore, a wealth of diagnostic plots (see e.g. Baddeley et al., 2005, 2015) and goodness-of-fit tests (see e.g. Illian et al., 2008; Diggle, 2013; Myllymäki et al., 2017) are available to evaluate the models. In any case, building up a useful model in real life applications is a challenging task. One important step in model construction is testing covariate effects in a parametric model.

For inhomogeneous Poisson point processes, the likelihood ratio test can be used to examine whether a point pattern depends on covariates in the presence of confounding covariates (Baddeley et al., 2015, p. 372). Another classical test for the null hypothesis '$H_0 : \beta = 0$' for a single parameter β of a point process model under the presence of effect of other nuisance covariates is the Wald test. It has also been adjusted for Gibbs processes (Coeurjolly and Rubak, 2013) and cluster processes (Waagepetersen and Guan, 2009) through asymptotic behavior of the estimators. In the Bayesian context, the posterior distribution of β is used to draw conclusions about the effect of the covariate. The Bayesian analysis is easily available for the log Gaussian Cox point processes (Møller et al., 1998) by the R package R-INLA (Rue et al., 2009, www.r-inla.org), while the Wald test is implemented in the R library spatstat (Baddeley et al., 2015). We are interested in the same hypothesis either for a scalar or vector β for a general point process model that may include second or higher order structures. Thus our focus is also on processes and estimation methods, for which the Wald test is not yet available. Moreover, our focus is on the local assessment of dependence between the point pattern and the covariate, whereby we use test statistics for which the asymptotic distributions are unknown. To this end, we present tests based on the Monte Carlo simulation and test statistics based on spatial residuals that can be applied to any model for which spatial residuals can be computed.

The basic idea in Monte Carlo testing is to estimate the variability of a chosen test statistic under H_0 from the simulations in order to determine whether the test statistic calculated from the data deviates from the H_0 case (Barnard, 1963; Besag and Diggle, 1977; Illian et al., 2008; Myllymäki et al., 2017). In principle, the simulation scheme allows the use of any point process model and test statistic, as long as suitable simulations can be obtained under H_0. Our simulation based inference mimics the permutation inference in general linear models (GLMs) proposed by Freedman and Lane (1983); see also Winkler et al. (2014) and Mrkvička et al. (2019a,b). In neuroimaging applications, the same linear model is typically fitted at each spatial pixel or voxel of an image (Winkler et al., 2014) and the aim is to find the spatial pixels where the data contradicts the null hypothesis. The first step of such a test is to choose a test statistic to express the distance between a model and its sub model, which is the null model; typically the F-statistic is used here. Secondly, permutations are used to obtain the distribution of the test statistic under H_0. One strategy is to permute the residuals from the null model (Freedman and Lane, 1983); although this is an approximative method, according to Anderson and Ter Braak (2003) (see also Winkler et al., 2014) it is the most precise permutation method in the case of nuisance effects. The last step is to apply a multiple testing correction. These steps are followed in our tests of covariate effects in point process models. In the second step, instead of permutations, the Monte Carlo simulation is used.

Our proposed test statistics are based on point process residuals (see e.g. Stoyan and Grabarnik, 1991; Baddeley et al., 2005; Waagepetersen, 2005; Baddeley et al., 2015). For point process models for which the conditional intensity is known in closed form, we use smoothed raw residuals defined by Baddeley et al. (2005). For Cox processes we instead apply smoothed raw residuals based on the intensity, as suggested by Waagepetersen (2005); because our tests rely on Monte Carlo simulation, there is no problem in using the intensity instead of the conditional intensity, if the statistic is informative of the null hypothesis. Our first test statistic is motivated by the classical F-statistic. The difference between the model and the sub model is summarized as the difference of squared smoothed residuals at each spatial location u in the observation window W. In the case where only the effect of a continuous covariate is investigated, we also propose a test statistic to detect where the covariate improves the match of the shape of the predicted and empirical intensities, roughly speaking, where the covariate locally influences the intensity of the point process. This local characteristic is particularly useful for assessing the quality of the covariate in the point process

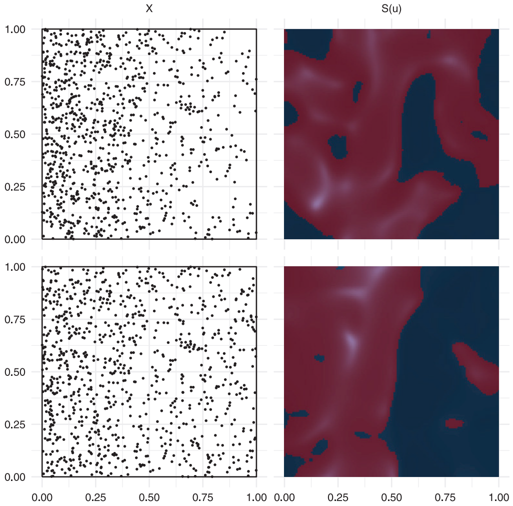


Fig. 1. On the left: Two point patterns with a trend in the whole observation window (top), and in the left part of the observation window (bottom). On the right: the significant area (red) detected by a global envelope test using the proposed local characteristic and significance level $\alpha = 0.05$. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

model. Consider the example of Fig. 1: the points represent locations of objects (e.g. infected trees or saplings) whose intensity depends on the distance from the left edge of the observation window W (e.g. a road). An interesting question that arises is whether there is a trend in the x-direction, which can be inspected by the inhomogeneous Poisson process with x-coordinate as the only covariate. In the first pattern (first row) there is indeed a decreasing trend in the x-direction, while in the second pattern (second row) there is a change in dependence in the middle of the image. Namely, on the left the intensity of the second pattern decreases as the distance from the edge increases, while on the right the intensity is flat. While the first statistic detects a similar overall improvement of the model fit as the Wald test for both cases (figure not shown), the proposed local test correctly identifies that in the first pattern the linear trend leads to an improved local fit in nearly the whole observation window; in the second pattern the fit is improved only on the left part of the window (Fig. 1, right). Therefore, a researcher can think of alternative models to better capture the variation in the intensity. A more detailed description will be given below (Section 4).

The last step of our Monte Carlo test is to apply a global envelope test (Myllymäki et al., 2017) to the set of empirical and simulated spatial statistics. The global envelope test solves the multiple testing problem that arises from considering the spatial statistics in the observation window W. This multiple testing method allows one to also detect the spatial area of significance of the test, as shown in the example above (Fig. 1).

The rest of the article is organized as follows. Point processes and residuals considered in this article are described in Section 2. Thereafter, the new global and local statistics and the simulation scheme for testing the effect of spatial covariates are presented in detail in Section 3. Section 4 illustrates the local test with further examples. A simulation study on the empirical significance levels and power of the proposed test is presented in Section 5, and Section 6 shows an application of the method to real data from the literature. Further discussion is included in Section 7.

The proposed tests are implemented in the R library GET (Myllymäki et al., 2017; Myllymäki and Mrkvička, 2019).

2. Preliminaries on point processes and their residuals

A spatial point pattern $\mathbf{x} = \{x_1, \ldots, x_n\}$ consists of the (unordered) locations $x_1, \ldots, x_n$ that are observed in a bounded window $W \subset \mathbb{R}^2$. In this paper, we investigate the effect of covariates in a general parametric spatial point process model which has been fitted to $\mathbf{x}$ through some arbitrary method. The covariates are assumed to affect only the *first-order trend* $b(u)$, $u \in W$, of the model, and in the following sections we first describe what we mean by the first-order trend $b(u)$ for different point process models X. Section 2.1 describes $b(u)$ for finite point processes; for Cox processes we use $b(u)$, as defined in Section 2.2. Section 2.3 defines the residuals considered in the new statistics thereafter.

2.1. First-order trend of finite point processes

Assume that the model has a probability density $f_\theta(\mathbf{x})$ with respect to the unit rate Poisson process on W satisfying the positivity condition, i.e. if $f_\theta(\mathbf{x}) > 0$ and $\mathbf{y} \subset \mathbf{x}$ then $f_\theta(\mathbf{y}) > 0$ for any finite point patterns $\mathbf{x}, \mathbf{y} \subset W$. Examples of such processes are homogeneous and inhomogeneous Poisson processes and pairwise interaction point processes. For all these processes, the density can be expressed in the 'Gibbs' form $f(\mathbf{x}) = \exp\{V_0 + V_1(\mathbf{x}) + V_2(\mathbf{x}) + \cdots + V_n(\mathbf{x})\}$ for unique functions V_k, called the potentials of order k, where $n = n(\mathbf{x})$ is the random number of points (Ripley and Kelly, 1977; Baddeley et al., 2005). Here V_0 determines the normalizing constant and, formally, $V_k(\mathbf{x}) = \sum_{\mathbf{y}} v_k(\mathbf{y})$ where $v_k(\mathbf{y}) \in [-\infty, \infty)$ and the sum is over all subsets $\mathbf{y} \subseteq \mathbf{x}$ with $n(\mathbf{y}) = k$, $k = 1, \ldots, n$. We define the first-order trend for these processes as

$$b(u) = \exp(v_1(\{u\})).$$

If the potentials of higher order $V_{\geq 2} = V_2(\mathbf{x}) + \cdots + V_n(\mathbf{x})$, that determine interpoint interaction, are identically zero, then the process is a Poisson process with intensity function $b(u)$, $u \in W$.

Baddeley et al. (2005) defined the residuals for these type of finite point processes X in W through their Papangelou conditional intensity $\lambda(u, \mathbf{x})$ that uniquely determines the probability density f, and vice versa. Loosely speaking, $\lambda(u, \mathbf{x})\mathrm{d}u$ is the conditional probability that there is a point of X in an infinitesimal region of area $\mathrm{d}u$ containing u, given that the rest of the point process coincides with $\mathbf{x}$. Formally, for $u \in W$, $u \notin \mathbf{x}$, the conditional intensity is defined by $\lambda(u, \mathbf{x}) = f(\mathbf{x} \cup \{u\})/f(\mathbf{x})$ if $f(\mathbf{x}) > 0$, and $\lambda(u, \mathbf{x}) = 0$ otherwise. For $u \in \mathbf{x}$, it is $\lambda(u, \mathbf{x}) = \lambda(u, \mathbf{x}/\{u\})$. In general, it holds that

$$\log\{\lambda(u, \mathbf{x})\} = v_1(u) + \sum_i v_2(\{u, x_i\}) + \sum_{i<j} v_3(\{u, x_i, x_j\}) + \cdots, \quad u \notin \mathbf{x}.$$

For example, for the Poisson process, the conditional intensity coincides with the intensity function and also with the first-order trend, i.e.

$$\lambda(u, \mathbf{x}) = b(u), \quad u \in W.$$

For the general pairwise interaction process, it is

$$\lambda(u, \mathbf{x}) = b(u) \prod_{i=1}^{n} c(u, x_i),$$

where the 'activity' (or first-order trend) $b(u) \geq 0$, $u \in W$, is typically used to model spatial variation in the abundance of points, and $c(u, v) = c(v, u) \geq 0$, $u, v \in W$, is the 'interaction' that can be used to model the association between points.

2.2. First-order trend of Cox processes

The Cox process is a 'doubly stochastic Poisson process' with a random intensity $\Lambda(u)$; given the realization $l(u)$ of the non-negative random intensity, the points follow the inhomogeneous Poisson process with intensity function $l(u)$. Examples of Cox processes are the log Gaussian Cox process (LGCP) (Møller et al., 1998) and the Neyman–Scott point process (see e.g. Illian et al., 2008; Mrkvička et al., 2014). In the LGCP, which we use in our examples, the random intensity is

$$\Lambda(u) = b(u) \exp(Z(u)),$$

where $b(u)$ is the first-order trend and $Z(u)$ is a zero-mean Gaussian random field, with variance σ^2, that creates additional clustering. The intensity of the LGCP is

$$\lambda(u) = \mathbf{E}\Lambda(u) = b(u)\exp(\sigma^2/2).$$

2.3. Point process residuals

Different types of residuals have been defined for finite point processes (see Section 2.1) based on the Georgii–Nguyen–Zessin (GNZ) formula (Georgii, 1976; Nguyen and Zessin, 1979; Stoyan and Grabarnik, 1991; Baddeley et al., 2005). We consider only the raw residuals that correspond to those in a linear model: Given data $\mathbf{x}$ and a general parameter estimate $\hat{\theta} = \hat{\theta}(\mathbf{x})$, Baddeley et al. (2005) defined the raw residuals

$$R_{\hat{\theta}}(B) = n(\mathbf{x} \cap B) - \int_B \hat{\lambda}(u, \mathbf{x})\mathrm{d}u \tag{1}$$

with the fitted conditional intensity $\hat{\lambda}(u, \mathbf{x}) = \lambda_{\hat{\theta}}(u, \mathbf{x})$ for any $B \subseteq W$. The raw residuals $R_{\hat{\theta}}(B)$ are a signed measure on W, with atoms of mass 1 at the data points, and a negative density $-\hat{\lambda}(u, \mathbf{x})$ at all locations $u \in W$ (Baddeley et al., 2005). For diagnostic plots to investigate spatial trends, Baddeley et al. (2005, p. 634) defined the smoothed raw residuals as

$$\begin{aligned} s(u) &= e(u)\left(\sum_{x_i \in \mathbf{x}} k(u - x_i) - \int_W k(u - v)\hat{\lambda}(v, \mathbf{x})\mathrm{d}v\right) \\ &= \tilde{\lambda}(u) - \lambda^{\dagger}(u) \end{aligned} \tag{2}$$

for $u \in W$, where e is an edge correction in the window W, $e(u)^{-1} = \int_W k(u - v)\mathrm{d}v$, and k is a smoothing kernel. Thus, the smoothed raw residuals are the difference between a non-parametric kernel smoothing estimator of the point process intensity function, $\tilde{\lambda}(u)$, and a kernel-smoothed version of the parametric estimator of the conditional intensity, $\lambda^{\dagger}(u)$. According to Baddeley et al. (2005), these two intensities should be approximately equal if the fitted model is correct, and it is hoped that then $s(u) \approx 0$. Positive and negative values of $s(u)$ relate to the underestimation and overestimation of the intensity, respectively.

For some point processes such as Cox and cluster processes, including inhomogeneous log Gaussian Cox processes and certain Neyman–Scott processes, the intensity function is known in a closed form, whereas the conditional intensity is not. For such processes, in the analogy of the raw residuals (1), Waagepetersen (2005) proposed the residuals

$$R'_{\hat{\theta}}(B) = n(\mathbf{x} \cap B) - \int_B \hat{\lambda}(u)\mathrm{d}u$$

where $\hat{\lambda}(u) = \lambda_{\hat{\theta}}(u)$ is the fitted intensity. Baddeley et al. (2005) suspected that these residuals could indeed be valuable for assessing spatial trends, but they might be insensitive to the lack of fit if the model parameters are estimated using the minimum contrast estimation (as in Baddeley et al., 2015). In the analogy of (2), the corresponding smoothed residuals can be defined as

$$s(u) = e(u)\left(\sum_{x_i \in \mathbf{x}} k(u - x_i) - \int_W k(u - v)\hat{\lambda}(v)dv\right).$$

We employ these smoothed residuals for Cox processes.

3. Testing for the covariate effect

Let $\mathbf{x}$ be a point pattern and X be the point process model under consideration for $\mathbf{x}$ in W. Assume that the first-order trend of the process X takes the form

$$b(u) = \exp\left\{\sum_{k=1}^{K} \beta_k C_k(u) + \sum_{l=1}^{L} \gamma_l U_l(u)\right\}, \tag{3}$$

where $C_k(u)$, $k = 1, \dots, K$, are spatial covariates of interest, $U_l(u)$, $l = 1, \dots, L$, are (spatial) nuisance covariates, and $\boldsymbol{\beta} = (\beta_1, \dots, \beta_K)$ and $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_L)$ are the regression coefficient vectors. We are interested in the null hypothesis

$$H_0 : \beta_1 = \cdots = \beta_K = 0. \tag{4}$$

To test the hypothesis (4) for inhomogeneous Poisson processes, it is possible to use the likelihood ratio test (see e.g. Baddeley et al., 2015, p. 372). Furthermore, for $K = 1$, the Wald test can be used with the test statistic

$$V = \frac{\hat{\beta}}{\mathrm{se}(\hat{\beta})} \tag{5}$$

where $\hat{\beta}$ is the estimate of β under the alternative hypothesis H_1, and $\mathrm{se}(\hat{\beta})$ is the estimate of the standard error of $\hat{\beta}$. For inhomogeneous Poisson processes and the maximum likelihood estimator $\hat{\beta}$, the asymptotic null distribution of this statistic is standard normal (see e.g. Baddeley et al., 2015, Section 10.3.3). The test can also be used for inhomogeneous Cox processes (Waagepetersen and Guan, 2009) and Gibbs processes (Coeurjolly and Rubak, 2013). These tests are implemented in the R package spatstat (Baddeley et al., 2015); we compared the performance of our spatial statistics to these in Section 5.

In Section 3.1, we first present two spatial statistics for testing covariate effects. The first statistic inspects the hypothesis (4) through spatial residuals. In addition to the test decision, the test allows the detection of spatial locations where the residuals are significantly reduced. The second statistic allows the inspection of the local dependence between an interesting continuous covariate and the point density under the presence of nuisance variables. In Section 3.2, we describe how these new statistics can be used in a Monte Carlo test.

3.1. Spatial statistics

To test the hypothesis H_0, we first defined a *spatial* statistic motivated by the classical F-statistic, namely

$$F(u) = s_0(u)^2 - s_1(u)^2, \quad u \in W, \tag{6}$$

where $s_1(u)$ are the smoothed raw residuals of the model X with the first-order trend (3), and $s_0(u)$ are the smoothed raw residuals of the *reduced* model that has the first-order trend

$$b_{H_0}(u) = \exp\left\{\sum_{l=1}^{L} \gamma_l U_l(u)\right\}, \tag{7}$$

and that is otherwise equal to the full model X. While we do not know the asymptotic distribution of the statistic (6), a Monte Carlo test can be constructed based on it, leading to a simulation-based alternative to the Wald test. This test can detect where the residuals are reduced due to the covariates $C_k(u)$, $k = 1, \dots, K$. However, interpretation of the local $F(u)$ can sometimes be challenging due to the nature of the residuals: the residuals $s_0(u)$ of the reduced model are always close to zero in some parts of the observation window W, and in these locations the full model cannot be any better than the reduced model in terms of smaller residuals. Thus, the detected areas where the residuals are decreased can sometimes be surprising at first and difficult to interpret. We recommend always inspecting the residuals $s_0(u)$ and $s_1(u)$ alongside $F(u)$.

In the case that there is only one interesting covariate (i.e. β is a scalar and $K = 1$) we can also address local dependence between the point pattern and the covariate in the presence of nuisance covariates. Local dependence of X on $C(u)$ at $u \in W$ could be assessed by fitting the model (3), and testing H_0 in the surrounding area of the spatial location u. However, a computationally less demanding alternative is to investigate the dependence from the fitted global model, which, as we show below, can provide similar information and avoid possible problems with local fits when the data are sparse.

We define our second spatial statistic to investigate the local dependence of the point pattern and the spatial covariate $C(u)$ as

$$S(u) = \frac{a_0(u)}{a_1(u)} = \frac{\int_{B_u \cap W} (s_0(x) - \bar{s}_0(u))^2 dx}{\int_{B_u \cap W} (s_1(x) - \bar{s}_1(u))^2 dx}, \quad u \in W, \tag{8}$$

where

$$\bar{s}_i(u) = \frac{1}{|B_u \cap W|} \int_{B_u \cap W} s_i(x) dx, \quad u \in W, i = 0, 1,$$

and B_u is a ball of radius R centered at u. The numerator $a_0(u)$ and denominator $a_1(u)$ measure variation in the smoothed residuals of the reduced and full models in the sub windows B_u for $u \in W$, respectively. The quantity $a_i(u)$, $i = 0, 1$, is smallest when the smoothed residuals $s_i(u)$, $i = 0, 1$, are constant in B_u. This happens when the deviations of the smoothed predicted (conditional) intensity from the kernel smoothed intensity $\tilde{\lambda}(u)$ are zero (perfect fit) or constant (perfect fit for slope) in B_u. Therefore, the statistic $S(u)$ obtains values larger than one at $u \in W$ if the full model captures the shape of the point density $\tilde{\lambda}(u)$ better than the reduced model.

For the statistics $F(u)$ and $S(u)$, we need to choose the bandwidth for smoothing the raw residuals. For $S(u)$, also the radius R needs to be chosen. As the kernel for smoothing in (2), we used the Gaussian kernel and chose the bandwidth according to the rule of thumb of Scott (1992, p. 152). This rule is typically useful when a gradual trend is expected. Furthermore, we chose R relative to Scott's bandwidths given in the x and y directions; namely we took R to be the mean of the two bandwidths. In general, the choices should be made depending on the application, but relatively large bandwidths are typically recommended to investigate trends in the patterns.

3.2. Simulation based inference with graphical interpretation

According to Anderson and Ter Braak (2003) and Winkler et al. (2014), the approximative permutation method proposed by Freedman and Lane (1983) is one of the most precise permutation methods in the presence of nuisance effects in GLMs. The algorithm is based on permuting the residuals from the null model. It does not satisfy the exchangeability property, but the significance level of such permutation schemes is close to the nominal level. We mimic this algorithm to test the hypothesis (4). In the case of point processes, we cannot permute the residuals, because our residuals are a spatial field in W, and we assume here to have only observed a point pattern in W. Instead we can simulate point patterns from the reduced model. More precisely, we propose the following algorithm to generate simulations under H_0:

1. Fit the full model (3) and the reduced model (7) to the data pattern X, and calculate the test statistic (6) (and/or (8)). Denote the estimated statistics by $F_0(u)$ (and $S_0(u)$) at $u \in W$.
2. Generate s simulated patterns X_i^*, $i = 1, \dots, s$, from the fitted reduced model.
3. For each $i = 1, \dots, s$, fit the full and reduced models to the simulated data X_i^* and calculate the test statistic (6) (and/or (8)). Denote the estimated statistics by $F_i(u)$ (and $S_i(u)$) at $u \in W$.

Consequently, we have $s + 1$ spatial fields $F_i(u)$, $i = 0, 1, \dots, s$ (and/or $S_i(u)$, $i = 0, 1, \dots, s$). Under the null hypothesis (4), the statistics $F_i(u)$ (or $S_i(u)$) come from the same distribution. A Monte Carlo test can be done based on the statistics $F_i(u)$ (or $S_i(u)$) to test the null hypothesis (4).

As we are not testing a simple hypothesis, since parameters of the null model are estimated, the two-stage Monte Carlo test (Dao and Genton, 2014; Myllymäki et al., 2017; Baddeley et al., 2017) should be applied. On the other hand, the residuals are not strongly related to estimation procedures, and thus the one-stage Monte Carlo test is not assumed to be seriously liberal or conservative. Because the two-step test is computationally rather demanding, we describe below the one-stage version and investigate its applicability in Section 5.

To make a Monte Carlo test based on $F_i(u)$, $i = 0, 1, \dots, s$ (or $S_i(u)$, $i = 0, 1, \dots, s$), we utilize the completely non-parametric global envelope test based on the extreme rank length (ERL) measure (Myllymäki et al., 2017; Narisetty and Nair, 2016; Mrkvička et al., 2018). The ERL measure solves the multiple testing problem which arises by comparing the functions $F_i(u)$ (Mrkvička et al., 2017). The test also provides a graphical interpretation of the test results by identifying the locations where the data contradicts the null hypothesis.

In practice, the functions are evaluated on a finite number of locations $u_1, \ldots, u_d \in W$, and we denote the statistics at these locations by $F_i = F_i(u_1), \ldots, F_i(u_d)$, $i = 0, 1, \ldots, s$. We consider the one-sided test, where large values of $F(u)$ are considered extreme, i.e. we want to find if and where the covariates $C_k(u)$, $k = 1, \ldots, K$, improve the model.

The ERL measure proposed by Myllymäki et al. (2017) and Narisetty and Nair (2016) is based on the ranks of the F_i among each other. Let $R_{0j}, R_{1j}, \ldots, R_{sj}$ be the raw ranks of $F_0(u_j), F_1(u_j), \ldots, F_s(u_j)$, such that the largest $F_i(u_j)$ has rank 1. The vector of pointwise ranks $(R_{i1}, R_{i2}, \ldots, R_{id})$ at $u_1, \ldots, u_d$ is attached to each F_i. Let $\mathbf{R}_i = (R_{i[1]}, R_{i[2]}, \ldots, R_{i[d]})$ be these pointwise ranks ordered from smallest to largest, i.e. $R_{i[j]} \leq R_{i[j']}$ whenever $j \leq j'$. Formally, the ERL measure is defined by the lexicographic ordering of the $\mathbf{R}_i$, $i = 0, 1, \ldots, s$:

$$E_i = \frac{1}{s+1} \sum_{i'=0}^{s} \mathbf{1}(\mathbf{R}_{i'} \prec \mathbf{R}_i) \tag{9}$$

where

$$\mathbf{R}_{i'} \prec \mathbf{R}_i \quad \iff \quad \exists\, n \leq d : R_{i'[l]} = R_{i[l]} \,\forall\, l < n,\ R_{i'[n]} < R_{i[n]}.$$

The p-value of a Monte Carlo test based on the ERL measure is $p_{\text{erl}} = \sum_{i=0}^{s} \mathbf{1}(E_i \leq E_i)\big/(s+1)$ and the $100(1-\alpha)\%$ global ERL envelope induced by the ERL measure is defined as a hull of the vectors that are considered non-extreme by the measure at the given significance level α, see details in Mrkvička et al. (2018) and Narisetty and Nair (2016). Throughout this paper, we use the significance level $\alpha = 0.05$. If the data function F_0 goes outside the constructed 95% global envelope, then also $p_{\text{erl}} \leq 0.05$. In the simulation experiment below, we explore the empirical significance level of this test in a few specific cases.

We utilized the R package GET (Myllymäki et al., 2017; Myllymäki and Mrkvička, 2019) for the global envelope tests.

4. Simulated examples

Let us consider a few simulated examples to understand the behavior of the spatial test statistics (6) and (8). Let $W = [0, 1]^2$ and $u = (x, y) \in W$. Each of the example patterns shown in Fig. 2 (left) were generated from the inhomogeneous Poisson process with intensity

$$b(u) = \exp(\gamma + \beta Z(u)), \tag{10}$$

where $Z(u)$ is a covariate and β is its coefficient (both specified in each example below), and we used $\gamma = \log(1000/\int_W \exp(\beta Z(u))\mathrm{d}u)$ leading approximately to 1000 points in the patterns. We tested the significance of a covariate $C(u)$ in the inhomogeneous Poisson process specified below. Although the inhomogeneous Poisson process may not be the most suitable model for patterns of trees, we describe the simulated examples below as they were patterns of (certain) trees (or saplings or other plants) in a particular neighborhood, in order to resemble real life situations.

Example A (*Change in Dependence*)**.** The covariate in the simulated model (10) was

$$Z(u) = (0.5 - x)\mathbf{1}(x \leq 0.5),$$

where $\mathbf{1}(\cdot)$ denotes the indicator function, and $\beta = 2$. The point pattern (Fig. 2, top left) can be imagined to represent the locations of infected trees that are affected by the distance from the road (on the left of the window) up to a certain distance.

We considered the test for a linear trend (i.e., our interesting covariate was $C(u) = x$) and the reduced model was simply the homogeneous Poisson process. The significant regions for both spatial statistics (6) and (8) are shown in Fig. 2 (first row). The spatial F-statistic (6) shows that

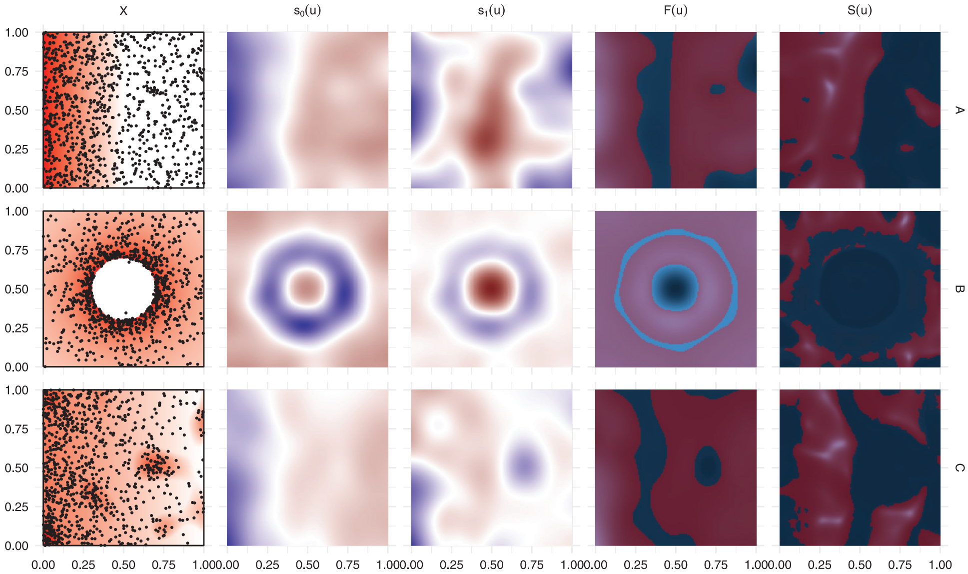


Fig. 2. Results for testing the significance of the (A) x-coordinate, (B) distance from the center, and (C) x-coordinate for the three different point patterns generated from the inhomogeneous Poisson process with intensity (10), see details of Examples A–C in the text. The columns from left to right show: the simulated point pattern (points) with the covariate $Z(u)$ in (10) on the background (red hues), the residuals $s_0(u)$ and $s_1(u)$ (red overestimation, blue underestimation), and the empirical $F(u)$ and $S(u)$ (blue hues) overlaid by the significant regions of the tests (red). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

the residuals of the model were decreased almost everywhere by adding the x-coordinate to the model. However, in the vertical line in the middle of the window no reduction was obtained; this is because the residuals of the reduced model, $s_0(u)$, are close to zero there (see Fig. 2, top row). The S-statistic (8) shows us that the changes in the point pattern intensity were better captured with the full model in the left part of the window. However, as expected, it shows no improvements in comparison to the constant model on the right part of the window. Thus, the S-statistic correctly suggests that the linear trend may be appropriate on the left hand side of W, but not on the right hand side.

Example B (*Spatial Break in Dependence*)**.** The point pattern (Fig. 2, middle left) was simulated from the model (10) with

$$Z(u) = 0.5/d(u) \cdot \mathbf{1}(d(u) > 0.2)$$

where $d(u) = \sqrt{(x-0.5)^2 + (y-0.5)^2}$ is the distance from the center of the window W and $\beta = 2$. This example can be understood as mimicking a landscape with a swamp in the middle of a valley: generally the trees prefer the valley, but they do not grow in the swamp.

We tested the effect of the covariate $C(u) = d(u)$. The reduced model was the homogeneous Poisson process. In this case, the spatial F-statistic shows reduction in the residuals next to the swamp where the tree density is highest, as well as in the outskirts of the valley (Fig. 2, middle row). The residuals were not reduced in the swamp, where the new model is obviously flawed and also not in a circle where the $s_0(u)$ residuals were around zero (see again Fig. 2, middle row). Thus, the swamp is detected by the spatial $F(u)$ as a location where the full model is not any better than the reduced model (in fact, in terms of residuals it is worse). The local S-statistic shows that the changes in intensity are better captured in the outskirts. The shape is not captured very close to the border of the swamp due to smoothing. When we divided the smoothing parameters by two, the circle in the middle became smaller, but at the same time the significant region in the outskirts became less uniform.

Example C (*Nonadditive Dependence*)**.** The point pattern (Fig. 2, bottom left) was simulated from the model (10) with $\beta = 2$ and

$$Z(u) = \max(1.5(1 - C(u)), U(u)),$$

where $C(u) = x$, and $U(u)$ is a realization of the zero-mean Gaussian random field with the Gaussian covariance function that has unit variance and the scale parameter 0.2. The field $1.5(1-C(u))$ creates a decreasing linear trend in the point density from left to right, which is confounded by $U(u)$. This pattern can be understood again as a pattern of infected trees, which is affected by two factors, an environmental covariate such as elevation ($U(u)$) and the distance from a road (on the left of the window). Infected trees are observed either in high elevation or close to the road.

Our nuisance covariates were the constant and the elevation $U(u)$, and we investigated the effect of the covariate $C(u) = x$. The spatial F-statistic shows a reduction in residuals almost in the whole window, except in the high elevation (bump on the left), and in the region where the $s_0(u)$ residuals are close to zero (Fig. 2, bottom row). This suggests that even though the linear trend improves the model fit, it is not entirely satisfactory. The spatial S-statistic further detects that the changes in intensity are better captured close to the road (left) and at the bottom and top right hand corners with relatively low intensities. On the other hand, no significant improvements were obtained in the high elevation area. Observing significance only at specific locations, rather than everywhere, suggests that the additive trend in the model contradicts the data.

5. Simulation study

The previous examples were designed to show how the new spatial statistics can be helpful to find improvements and misfits (no improvement) of the fitted model due to the interesting covariate $C(u)$. Here, we edit these models for the purpose of studying the empirical significance levels of the proposed tests and the Wald type tests, as well as their power in different cases where the fitted model is equal to the true model. In this scenario, we expect that the Wald test and the $F(u)$ test (the global envelope test based on (6)) have more power to reject the null hypothesis (4) than the $S(u)$ test (the global envelope test based on (8)); this is due to the nature of $S(u)$, that is designed to detect local fits rather than global fits. Therefore, to compare the Wald and $F(u)$ tests, and to determine how much less power the test based on $S(u)$ would have, we simulated data in the unit square $[0, 1]^2$ from the full (log linear) models with

M1: $b(u) = \exp(\gamma_1 + \beta x)$
M2: $b(u) = \exp(\gamma_1 + \beta Z(u))$
M3: $b(u) = \exp(\gamma_1 + \beta x + \gamma_2 Z(u))$
M4: $b(u) = \exp(\gamma_1 + \beta Z(u) + \gamma_2 x)$

where x is the x-coordinate, $u = (x, y)$, and $Z(u)$ is the realization of the zero-mean Gaussian random field $U(u)$ from Example C (Gaussian covariance function with variance 1 and scale 0.2). In each model, the factor with coefficient β is the interesting factor, and the others are considered nuisance factors in the simulations. Together with each of the first-order trends M1–M4, we considered the following second order structures

S0: none, i.e. the process was the inhomogeneous Poisson process,
S1: hard-core process with interaction radius 0.01 (small scale repulsion),
S2: LGCP with small scale clustering produced by the zero-mean Gaussian random field with exponential covariance function with interaction range parameter 0.02.

In all our simulations, the value of γ_1 was chosen to lead to 1000 points in the patterns on average, and the coefficient γ_2 of the nuisance covariate was fixed to 1, resulting in a rather strong (visually clear) effect of the nuisance covariate on the point pattern. First, to explore the empirical

Table 1
Empirical significance levels for testing the null hypothesis (4) at the significance level 0.05 for the three tests, the Wald test and the global ERL envelope test based on $F(u)$ and $S(u)$. For details about the models M1–M4 and S0–S2 see the text.

	Tests			
		Wald	$F(u)$	$S(u)$
M1	S0	0.058	0.048	0.047
M2		0.058	0.062	0.051
M3		0.058	0.059	0.040
M4		0.053	0.046	0.045
M1	S1	0.075	0.072	0.059
M2		0.057	0.056	0.033
M3		0.064	0.077	0.043
M4		0.082	0.080	0.064
M1	S2	0.059	0.033	0.050
M2		0.081	0.052	0.052
M3		0.048	0.048	0.044
M4		0.072	0.043	0.053

significance levels, we considered the models M1–M4 with S0, S1 and S2 and with $\beta = 0$ (i.e. the interesting factor had no effect on the intensity). Second, to explore the power of the tests, we simulated data from the models M1–M4 with S0, S1 and S2 and three different values for the parameter β (see Table 2). The smallest values of β were chosen so that the V-values of the Wald test were close to 2 on average, which was explored in a small side study. The other two values of β were taken to be a bit larger, to obtain higher powers. We simulated 1000 realizations of each model, performed the tests for each of these realizations, and calculated the empirical rejection rates of H_0 among the 1000 repetitions. We used the significance level $\alpha = 0.05$, and in each global envelope test we performed $s = 499$ simulations in the algorithm presented in Section 3.2. To estimate $F(u)$ and $S(u)$, we used the choices of the bandwidths given in Section 3.1. The Wald test was provided by the R library spatstat (Baddeley et al., 2015; Coeurjolly and Rubak, 2013; Waagepetersen and Guan, 2009).

The empirical significance levels are shown in Table 1. Note that the levels should be within the interval $(0.037, 0.064)$ with a probability of 0.95 (given by the 2.5% and 97.5% quantiles of the binomial distribution with parameters 1000 and 0.05), if the level of the test was correct. First, we can see that for the inhomogeneous Poisson processes (i.e. M1–M4 with S0), all the empirical significance levels are close to the nominal level 0.05. For the hard-core process (the case S1), the empirical significance levels of the Wald and $F(u)$ tests tended to be slightly liberal, while for $S(u)$ the levels were either close to 0.05 (M1, M3, M4) or slightly conservative (M2). The empirical levels of the $S(u)$ test were also good for the LGCP processes (the case S2). However, the $F(u)$ test was slightly conservative for M1; the Wald test, on the other hand, was liberal for both M2 and M4. Thus, to conclude, the significance levels of the $S(u)$ test were close to the nominal level 0.05 in all cases, while the $F(u)$ and Wald tests tended to be slightly liberal for S1 and Wald also for S2. It might be that $S(u)$ is even less related to the estimation procedure than the other two tests.

Next we compared the empirical powers, see Table 2. First, for the inhomogeneous Poisson processes (the case with the second-order structure S0), all the tests had rather similar powers, but the order from the highest to lowest power was always: Wald, $F(u)$, $S(u)$. For the hard-core processes (the case S1), the Wald and $F(u)$ tests had rather similar power, while the power of the $S(u)$ test was a bit lower. Note, however, that the empirical significance levels of the Wald and $F(u)$ tests were a bit above the nominal level, which may explain why they were a bit higher in power than the $S(u)$ test. For the LGCP model (the case S2), the Wald test had higher power than the $F(u)$ and $S(u)$ tests. Again, this result may be at least partly due to differences in the empirical significance levels of the tests. In addition, in this case the $S(u)$ test always had higher power than the $F(u)$ test.

Table 2
Rejection rates for testing the null hypothesis (4) at the significance level 0.05 for the three tests, the Wald test and the global ERL envelope test based on $F(u)$ and $S(u)$. The data were simulated from the models with first-order trends M1–M4 and second-order structures S0–S2, and with different parameter values of β specified in the first column of the table. See description of the models in the text.

		S0			S1			S2		
	β	Wald	$F(u)$	$S(u)$	Wald	$F(u)$	$S(u)$	Wald	$F(u)$	$S(u)$
M1	0.25	0.616	0.541	0.533	0.432	0.425	0.333	0.264	0.151	0.186
	0.50	0.998	0.991	0.984	0.957	0.945	0.905	0.683	0.409	0.554
	0.75	1.000	1.000	1.000	1.000	0.999	0.999	0.946	0.671	0.871
M2	0.10	0.616	0.544	0.541	0.354	0.360	0.291	0.248	0.156	0.207
	0.20	0.995	0.988	0.986	0.911	0.904	0.833	0.700	0.523	0.620
	0.30	1.000	1.000	1.000	0.999	0.999	0.994	0.946	0.839	0.909
M3	0.25	0.557	0.530	0.461	0.485	0.476	0.406	0.181	0.150	0.147
	0.50	0.989	0.980	0.959	0.967	0.954	0.920	0.513	0.377	0.407
	0.75	1.000	1.000	1.000	0.999	0.999	0.999	0.822	0.616	0.719
M4	0.10	0.561	0.499	0.483	0.415	0.406	0.316	0.210	0.138	0.152
	0.20	0.988	0.971	0.966	0.940	0.928	0.868	0.642	0.479	0.553
	0.30	1.000	1.000	1.000	1.000	1.000	0.998	0.915	0.777	0.872

In fact, we had expected the $S(u)$ test to have less power than the other tests in this simulation experiment where the tested model was a truly global model, as it has an inherently different nature. However, its power was only slightly lower than the power of the Wald and $F(u)$ tests for the cases of inhomogeneous Poisson (S0) and hard-core (S1) processes. For the clustered case (S2), it led to a higher power than the $F(u)$ test.

6. Data examples

We illustrate the proposed tests on two point patterns from the literature. Both point patterns are available in the R library spatstat (Baddeley et al., 2015).

6.1. Rainforest data

Fig. 3 shows the point pattern of the locations of 3605 trees in a tropical rain forest, accompanied by the elevation and gradient (norm of the elevation gradient) in the study region (Condit, 1998; Condit et al., 1996; Hubbell and Foster, 1983; Møller and Waagepetersen, 2007). As the first model for the pattern, we considered the inhomogeneous Poisson process with intensity $b(u) = \exp(\gamma + \beta \text{elevation}(u))$. The Wald test indicated that the elevation has a slightly significant effect ($\hat{\beta} = 0.005$, $V = 2.32$). The outputs of the global envelope tests based on $F(u)$ and $S(u)$ with $s = 999$ simulations and $R = 59$ m are given in Fig. 4(a). While the $F(u)$ shows a reduction of residuals in quite large areas, according to the $S(u)$ test there were no improvements in the local fits. By looking at the point pattern in Fig. 3, we can see, for example, that in the area on the right where the elevation changes greatly, the point intensity does not vary much.

Secondly, we considered the model $b(u) = \exp(\gamma + \beta \text{gradient}(u))$. The Wald test indicated a significant effect of the gradient ($\hat{\beta} = 5.03$, $V = 20.5$). The $F(u)$ test shows that the residuals were particularly reduced in the middle of the observation window with low gradient, and in the four spots with an intermediate gradient (see Fig. 4(b)). On the other hand, no reduction in residuals was obtained in the top left part of the window, where the gradient is rather low but the number of trees is high. The $S(u)$ test shows improvements in somewhat different areas than $F(u)$, namely in the areas where the point density changes with the gradient. No improvements in the local fit were obtained in the top left hand corner or bottom parts of the window.

We then followed Baddeley et al. (2015, p. 309) and further considered the quadratic model

$$b(u) = \exp(\gamma_1 + \gamma_2 \text{gradient}(u) + \beta \text{gradient}^2(u)),$$

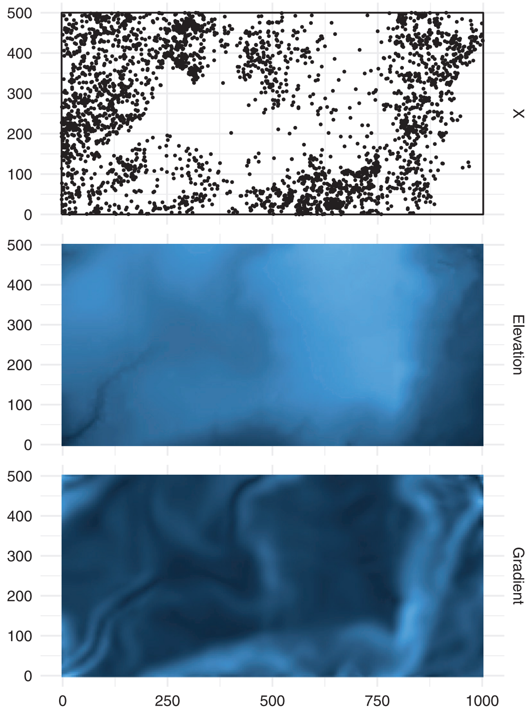


Fig. 3. The point pattern of the tree species *Beilschmiedia pendula*, elevation and slope (norm of elevation gradient) in a 1000 m × 500 m area in the tropical rainforest of Barro Colorado Island.

where the gradient was regarded as a nuisance covariate and the gradient^2 was the covariate of interest. According to the Wald test, the gradient squared had a significant effect ($\hat{\beta} = -55.5$, $v = -13.0$) and also the $F(u)$ test shows a reduction in the residuals at several locations, particularly in the area at the top discussed above (Fig. 4(c)). However, the $S(u)$ test shows that locally the gradient^2 did not improve the fit. Therefore, it appears that these covariates were not sufficient to explain the pattern, and some further covariates should be considered.

As a summary, we showed first that the elevation could not explain much of the variation in the point density. Secondly, the gradient was better, but it could not explain the high intensity in the upper left corner of W. Thus, the gradient was also insufficient, suggesting that dependence may be non-linear. However, the gradient squared could not improve the local fits further, suggesting that some other phenomenon plays a role.

6.2. Forest fires

Fig. 5 shows the locations of the forest fires that were caused by lightning in the Castilla-La Mancha region of Spain between 2004 and 2007. The data is a part of the data set *clmfires* available at the R library spatstat (Baddeley et al., 2015). Fig. 5 shows further elevation and a land use area that was classified as different types of forests (namely conifer, dense forest or mixed forests). The rest of the area is comprised of urban, farm, meadow, grassland, bush and scrub, and artificial green land use classes. We first studied the inhomogeneous Poisson process with intensity

$$b(u) = \exp(\gamma_1 + \gamma_2 1(u \in \text{forest}) + \beta \text{elevation}(u)),$$

i.e. the effect of elevation accounting for the effect of the forest class variable (forest or non-forest) as the nuisance covariate of the model. We first used Scott's rule for the smoothed residuals (with

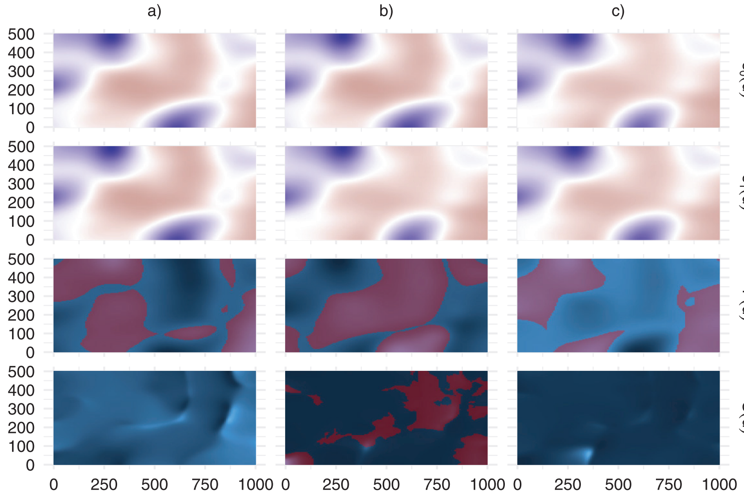


Fig. 4. Investigating the effect of a covariate in the inhomogeneous Poisson process model for the point pattern of Fig. 3 in three different models, where the interesting covariate was: (a) elevation, (b) gradient, and (c) gradient^2. In (a) and (b), the constant was the only nuisance factor, while in (c) the nuisance factors were the constant and gradient. The residuals $s_0(u)$ and $s_1(u)$ of the reduced and full model (red overestimation, blue underestimation) and the $F(u)$ and $S(u)$ statistics for the data (blue hue on the background) overlaid by the significant regions of the global envelope test (red). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

corresponding $R = 29.5$), as we did earlier, but as these residuals were very smooth we divided the bandwidths by two to look at the phenomena on a finer scale ($R = 14.8$). The results shown below are for the latter case. The effect of elevation was again shown to be significant by the Wald test ($\hat{\beta} = 0.0015$, $V = 11.48$) and the $F(u)$ test also detected large areas with a reduction in spatial residuals (Fig. 6). However, in the $S(u)$ test we observed that the linear trend was significant only in a small region of W, indicating that the linear dependence on the elevation did not hold locally. On account of the possibility that the forest fires may not be frequent in areas of very high elevation, we next tested the global and local significance of the quadratic term of elevation in the model

$$b(u) = \exp(\gamma_1 + \gamma_2 1(u \in \text{forest}) + \gamma_3 \text{elevation}(u) + \beta(\text{elevation}(u))^2).$$

For all the tests, the null hypothesis was again rejected ($\hat{\beta} = -2.3 \cdot 10^{-6}$, $V = -5.624$, Fig. 6). The $S(u)$ test shows that the quadratic term only helped capture the shape of the point pattern intensity in some small intermediate elevation areas, and in high elevation areas in the southeast and northwest. On the other hand, in the high elevation area in the northeast, no improvements were obtained and $F(u)$ suggests that the residuals did not decrease there either; in this high elevation area the amount of fires is rather high, while the quadratic term suggested a decreased number of points in comparison to the null model. This suggests that other covariates are needed to explain the variation in the intensity in high elevation areas.

7. Discussion

We proposed new Monte Carlo tests for the dependence of a point pattern on covariates in a general point process model and general estimation strategy. In particular, the proposed test statistic $S(u)$ in (8) allows us to explore local dependence. In our approach, we did not fit local models, but instead the proposed test statistic used the fitted global model. It would be possible to fit local models in a similar manner as in Viladomat et al. (2014) for random fields and bind the local tests together in a global envelope test. Even though the R library spatstat.local (Baddeley, 2019) supports the fitting of local models, this approach would be more computational than the one used here.

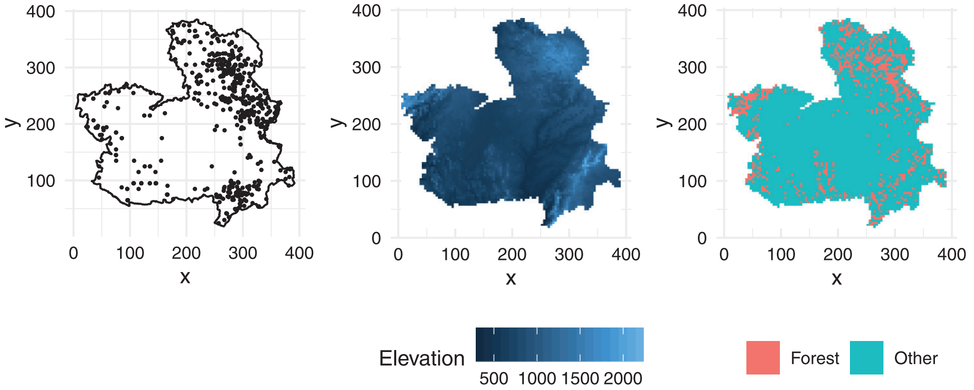


Fig. 5. The locations of forest fires that were caused by lightning in the Castilla-La Mancha region of Spain between 2004 and 2007 as well as the elevation and forest area (land use classes: forest, dense forest and mixed forest).

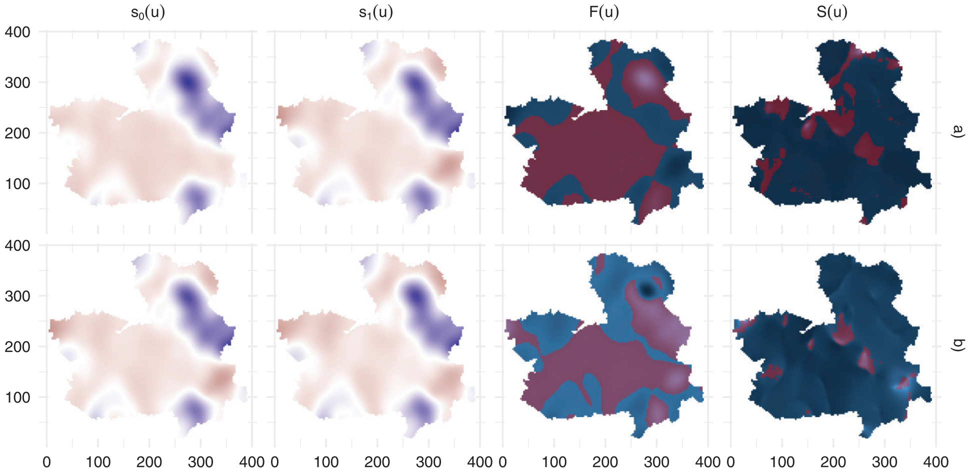


Fig. 6. Investigating the effect of a covariate in the inhomogeneous Poisson process model for the point pattern of Fig. 5 in two different models, where the interesting covariate was: (a) elevation and (b) elevation^2. In (a), the nuisance factors were the constant as well as class variable of forest/non-forest, while in (b) the nuisance factors were the constant, forest/non-forest, and elevation. The residuals $s_0(u)$ and $s_1(u)$ of the reduced and full model (red overestimation, blue underestimation) and the $F(u)$ and $S(u)$ statistics for the data (blue hue on the background) overlaid with the significant regions of the global envelope test (red). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Baddeley et al. (2005) recommended a diagnostic tool to plot residuals with respect to a given covariate. This tool helps to identify non-linear dependence of the intensity and the covariate by eye. In contrast, our test is able to identify the spatial locations where the covariate leads to a significantly better local fit of the model and where it does not. Of course, insignificance may also be present due to the lack of gradient or noise in the data, but generally the test can be used in the following way: if there are large areas without any significant local dependence, the dependence in that area is spoiled, the reasons for which are to be determined. Different reasons were shown in data studies and simulated examples, including nonlinearity, lack of dependence in a certain area, opposite dependence in a certain area, presence of another unobserved covariate in the model, non-additivity of the covariates effects. The Monte Carlo tests can also be used in cases where the Wald test is not applicable, such as in cluster processes with a composite likelihood estimation strategy. The Monte Carlo tests also reduced the slight liberality of the Wald test for Gibbs and cluster processes in our simulation study. In particular, the test based on the statistic $S(u)$ performed surprisingly well in testing the global hypothesis given that its focus is on local dependencies. Therefore, we believe that the new test can be a useful tool in model construction, complementing the toolbox for the analysis of point patterns to understand the significance of spatial covariates in point process models.

The hypothesis tested by the Monte Carlo method in this work was not simple. However, in our limited simulation study, the one-step Monte Carlo tests based on $F(u)$ and $S(u)$ achieved empirical significance levels close to the nominal level 0.05 for inhomogeneous Poisson processes, and for inhomogeneous models with small scale regularity and clustering; therefore, the one-stage tests are applicable. In any case, it is also possible to employ the two-stage tests (Dao and Genton, 2014; Baddeley et al., 2017; Myllymäki et al., 2017; Myllymäki and Mrkvička, 2019), which are exact, albeit more computational.

In principle, other forms of dependence on the covariates than (3) could be considered by the Monte Carlo tests. The only requirement is that the full and reduced models can be fitted both to the data and to the simulations of the reduced model.

Global envelope tests (Myllymäki et al., 2017) have so far been used for spatial point patterns mainly in goodness-of-fit testing. Here we have presented a new application where the global envelope tests with graphical interpretation can be useful. The tests are directly applicable to testing other hypotheses when the test statistic is a function, an image, or a general multivariate vector and simulations under the null hypothesis are available.

Acknowledgments

M.M. and M.K. were financially supported by the Academy of Finland (Project Numbers 295100 and 306875) and T.M. by the Grant Agency of Czech Republic (Project No. 19-04412S). The authors wish to acknowledge CSC — IT Center for Science, Finland, for computational resources.

References

Anderson, M., Ter Braak, C., 2003. Permutation tests for multi-factorial analysis of variance. J. Stat. Comput. Simul. 73, 85–113.

Baddeley, A., 2019. spatstat.local: Extension to 'spatstat' for local composite likelihood. R package version 3.6-0. URL: https://CRAN.R-project.org/package=spatstat.local.

Baddeley, A., Hardegen, A., Lawrence, T., Milne, R.K., Nair, G., Rakshit, S., 2017. On two-stage Monte Carlo tests of composite hypotheses. Comput. Statist. Data Anal. 114, 75–87.

Baddeley, A., Rubak, E., Turner, R., 2015. Spatial Point Patterns: Methodology and Applications with R. Chapman and Hall/CRC Press, London.

Baddeley, A., Turner, R., Møller, J., Hazelton, M., 2005. Residual analysis for spatial point processes (with discussion). J. R. Stat. Soc. Ser. B Stat. Methodol. 67, 617–666.

Barnard, G.A., 1963. Discussion of professor Bartlett's paper. J. R. Stat. Soc. Ser. B Stat. Methodol. 25, 294.

Besag, J., Diggle, P.J., 1977. Simple Monte Carlo tests for spatial pattern. J. R. Stat. Soc. Ser. C. Appl. Stat. 26, 327–333.

Coeurjolly, J.F., Rubak, E., 2013. Fast covariance estimation for innovations computed from a spatial Gibbs point process. Scand. J. Stat. 40, 669–684.

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

Condit, R., Hubbell, S.P., Foster, R.B., 1996. Changes in tree species abundance in a neotropical forest: Impact of climate change. J. Trop. Ecol. 12, 231–256.

Dao, N.A., Genton, M.G., 2014. A Monte Carlo adjusted goodness-of-fit test for parametric models describing spatial point patterns. J. Comput. Graph. Statist. 23, 497–517.

Diggle, P.J., 2013. Statistical Analysis of Spatial and Spatio-Temporal Point Patterns, third ed. CRC Press, Boca Raton.

Freedman, D., Lane, D., 1983. A nonstochastic interpretation of reported significance levels. J. Bus. Econom. Statist. 1, 292–298.

Georgii, H.O., 1976. Canonical and grand canonical Gibbs states for continuum systems. Comm. Math. Phys. 48, 31–51.

Hubbell, S., Foster, R., 1983. Diversity of canopy trees in a neotropical forest and implications for conservation. In: Sutton, S., Whitmore, T., Chadwick, A. (Eds.), Tropical Rain Forest: Ecology and Management. Blackwell Scientific Publications, Oxford, pp. 25–41.

Illian, J., Penttinen, A., Stoyan, H., Stoyan, D., 2008. Statistical Analysis and Modelling of Spatial Point Patterns, first ed. In: Statistics in Practice, John Wiley & Sons, Ltd, Chichester.

Møller, J., Syversveen, A.R., Waagepetersen, R.P., 1998. Log Gaussian Cox processes. Scand. J. Stat. 25, 451–482.

Møller, J., Waagepetersen, R.P., 2007. Modern statistics for spatial point processes. Scand. J. Stat. 34, 643–684.
Mrkvička, T., Muška, M., Kubečka, J., 2014. Two step estimation for Neyman–Scott point process with inhomogeneous cluster centers. Stat. Comput. 24, 91–100.
Mrkvička, T., Myllymäki, M., Hahn, U., 2017. Multiple Monte Carlo testing, with applications in spatial point processes. Stat. Comput. 27, 1239–1255.
Mrkvička, T., Myllymäki, M., Jílek, M., Hahn, U., 2018. A one-way ANOVA test for functional data with graphical interpretation. arXiv:1506.01646 [stat.ME].
Mrkvička, T., Myllymäki, M., Narisetty, N.N., 2019a. New methods for multiple testing in permutation inference for the general linear model. arXiv:1906.09004 [stat.ME].
Mrkvička, T., Roskovec, T., Rost, M., 2019b. A nonparametric graphical tests of significance in functional GLM. Methodol. Comput. Appl. Probab. http://dx.doi.org/10.1007/s11009-019-09756-y.
Myllymäki, M., Mrkvička, T., 2019. GET: Global envelopes in R. arXiv:1911.06583 [stat.ME].
Myllymäki, M., Mrkvička, T., Seijo, H., Grabarnik, P., Hahn, U., 2017. Global envelope tests for spatial processes. J. R. Stat. Soc. Ser. B Stat. Methodol. 79, 381–404.
Narisetty, N.N., Nair, V.J., 2016. Extremal depth for functional data and applications. J. Amer. Statist. Assoc. 111, 1705–1714.
Nguyen, X.X., Zessin, H., 1979. Integral and differential characterizations of the Gibbs process. Math. Nachr. 88, 105–115.
Ripley, B.D., Kelly, F.P., 1977. Markov point processes. J. Lond. Math. Soc. s2-15, 188–192.
Rue, H., Martino, S., Chopin, N., 2009. Approximate Bayesian inference for latent Gaussian models using integrated nested Laplace approximations (with discussion). J. R. Stat. Soc. Ser. B Stat. Methodol. 71, 319–392.
Scott, D.W., 1992. Multivariate Density Estimation. Theory, Practice and Visualization. Wiley, New York.
Stoyan, D., Grabarnik, P., 1991. Second-order characteristics for stochastic structures connected with Gibbs point processes. Math. Nachr. 151, 95–100.
Viladomat, J., Mazumder, R., McInturff, A., McCauley, D.J., Hastie, T., 2014. Assessing the significance of global and local correlations under spatial autocorrelation: A nonparametric approach. Biometrics 70, 409–418.
Waagepetersen, R., 2005. Comment on 'residual analysis for spatial point processes (with discussion)' by A. Baddeley et al.. J. R. Stat. Soc. Ser. B Stat. Methodol. 67, 617–666.
Waagepetersen, R., Guan, Y., 2009. Two-step estimation for inhomogeneous spatial point processes. J. R. Stat. Soc. Ser. B Stat. Methodol. 71, 685–702.
Winkler, A.M., Ridgway, G.R., Webster, M.A., Smith, S.M., Nichols, T.E., 2014. Permutation inference for the general linear model. NeuroImage 92, 381–397.