CHALMERS | UNIVERSITY OF GOTHENBURG

PREPRINT 2010:9

Some edge correction methods for marked spatio-temporal point process models

OTTMAR CRONIE

Department of Mathematical Sciences
Division of Mathematical Statistics
CHALMERS UNIVERSITY OF TECHNOLOGY
UNIVERSITY OF GOTHENBURG
Gothenburg Sweden 2010

Preprint 2010:9

Some edge correction methods for marked spatio-temporal point process models

Ottmar Cronie

Department of Mathematical Sciences
Division of Mathematical Statistics
Chalmers University of Technology and University of Gothenburg
SE-412 96 Gothenburg, Sweden
Gothenburg, February 2010

Preprint 2010:9
ISSN 1652-9715

Matematiska vetenskaper
Göteborg 2010

Some edge correction methods for marked spatio-temporal point process models

Ottmar Cronie[1]

[1]Mathematical Sciences, Chalmers University of Technology and University of Gothenburg, 412 96 Göteborg, Sweden
E-mail address: ottmar@chalmers.se

Abstract:

We propose three edge correction methods for (marked) spatio-temporal point processes. They are all based on the idea of placing an approximated expected behaviour of the process at hand (simulated realizations) outside the study region which interacts with the data during the estimation. These methods are applied to the Renshaw-Särkkä growth-interaction model (RS-model) presented in [16]. The specific choices of growth function and interaction function made are purely motivated by the forestry application considered here. A new estimator has been derived for the death rate (since the distribution of the life-time of an individual is allowed to depend on its current size) and, furthermore, we propose a new estimator for the (Poisson process) arrival intensity which compensates for the (unobserved) individuals arriving and dying between two sample time points without having been observed. The parameters related to the development of the marks are estimated using the same least-squares approach as proposed in [16]. Finally, the edge corrected estimation methods, in the context of fitting the RS-model, are applied to a data set of Swedish Scots pines.

Key words: Edge correction, Spatio-temporal marked point process, Least squares estimation, Maximum likelihood estimation

1 Introduction

Many of the spatial point structures, with appurtenant marks, which are encountered in nature and in our surrounding environments, are in fact results of evolutionary processes which have been developing over time. One example of such a process is a forest stand which, from once being an empty piece of land, grows and changes over time to become the full stand observed at a later time point. Often these marked spatial structures are measured only at one specific time point, thus containing no information regarding the temporal aspects of the evolutionary process responsible for the generation of the data. Hence, in situations such as these, tree stands and other marked patterns are treated as realizations of marked point processes (see e.g. [18] and [6]).

However, if one wants a more thorough understanding of the development process and its inherent interaction mechanisms one cannot ignore the collective development of the locations and the marks (sizes) through time. This new scenario makes us to take on a somewhat different approach where one treats recorded time series of marked patterns as outcomes of the development of spatio-temporal marked point processes. This second approach has been less studied, however. As the aspect of time enters the model the level of complexity quickly increases and formulating involved models, which try to cover every aspect of the development, usually has the drawback of creating decrease in tractability, applicability and interpretability (see e.g. [5]). It is therefore necessary to formulate models which are tractable and easily interpreted but yet manage to cover the relevant aspects of spatio-temporal modelling. One such model is what here will be referred to as the Renshaw-Särkkä growth-interaction model (RS-model) which has been studied in a series of papers, most recently in [15], [13], [16] and [14].

When measurements are made in some bounded study region, the structure of the spatial dependences and interactions existing between individuals outside and inside the study region remains unobserved. This phenomenon, which in particular concerns those individuals inside the study region who are close to its boundary, is generally referred to as edge effects. In the context of estimation, if the study region contains a large number of individuals the edge effects may not have a large impact on the estimates. However, it may be the case that we deal with a small study region which contains only a small amount of data, which often is the case with tree data. In such cases

there is a substantial risk that the edge effects generate quite severe biases and we therefore need some type of edge correction method when estimating the model parameters and summary statistics of interest. In the case of non-temporal analyses a number of methods for edge correction have been devised (see e.g. [7]) but these are not so easily generalized to the spatio-temporal setting. Hence, our main objective here is to develop methods which correct for these edge effects in the spatio-temporal setting.

We consider three edge correction methods which all, more or less, are based on the same idea. Initially one makes a first estimation (without edge correction) of the parameters of interest, Θ, thereby generating a set of biased parameter estimates, $\hat{\Theta}_*$. Once these estimates have been found one re-estimates the parameters, although, this time placing the "expected behaviour" of our spatio-temporal process, under the regime of $\hat{\Theta}_*$, in a buffer zone which surrounds the study region. During the re-estimation the individuals in this buffer zone have the purpose of interacting with the individuals (trees) at the boundary of the observation window, hence affecting the new estimates. These new edge corrected estimates will now replace $\hat{\Theta}_*$ and will then in turn be used to generate a new expected behaviour of the process. By letting this new expected behaviour take the place of the previous one we re-run the whole procedure, hence producing a new set of estimates. We iteratively continue in this fashion until we see convergence in the estimates. Now the question still remains regarding what is meant by and how to find this so called "expected behaviour" of the spatio-temporal process. The three edge corrections presented in this paper are basically three ways of estimating this expected behaviour and they are all based on successive simulations of an interacting process living outside our study region.

All three edge corrections presented in this paper will be applied to a slightly modified version of the RS-model (see [16]). The model considered here differs from its predecessor in that it allows the (exponential) distribution of each individual's life-time to vary with its size. This slight change of the process has had the consequence that a new maximum likelihood (ML) estimator for the death rate parameter has been derived, which takes into account that the size of an individual influences its viability. Furthermore, a new ML-estimator has been derived for the arrival intensity of the immigration process (Poisson process) governing the arrivals in time of new individuals. This new arrival intensity estimator tries to compensate for the unobserved births and deaths occurring between the time points at which the

process is sampled. The parameters related to the growth of the marks and the interaction between the marks, just as in [16], will be estimated separately from the arrival intensity and the death rate. [16] presents an approach where these mark related parameters are estimated using the least-squares method and we here choose to follow the exact same approach.

The paper is set up as follows. In Section 2 we will present the slightly modified version of the RS-model, in which the distributions of the life-times are allowed to vary with the sizes of the individuals. The least squares approach used in the estimation will be presented in Section 3 together with the new death rate estimator and the new arrival intensity estimator mentioned above. Further, in Section 4, we present the data set of Swedish Scots pines considered. In section 5 we describe in detail the three previously mentioned edge correction methods developed for spatio-temporal point processes (with interacting marks). In Section 5 we will also present the results obtained in the evaluation of the methods and once these methods have been presented and evaluated (in the context of the RS-model), they are applied to our Scots pine data set.

2 The model

The spatio-temporal growth-interaction model has recently been studied by Renshaw and Särkkä in [15] and [16] and by Renshaw et al. in [14]. We here investigate the model given in [16], with the modification that the distribution of an individual's lifetime is allowed to depend on its size. The process is defined as follows.

The base of the process can be described as an immigration-death process where the immigration part governs arrivals of new individuals to a region of interest, $W \subseteq \mathbb{R}^2$, and a death part handling the number of 'natural deaths' occurring. Additionally, upon arrival, individuals are assigned locations and appurtenant marks (sizes) which change deterministically over time.

More precisely, individuals enter W randomly in time according to a homogeneous Poisson process with intensity $\alpha\nu(W)$, $\alpha > 0$, where $\nu(W)$ denotes the area of W. As individual i arrives at time t_i^0 it is assigned a location $\mathbf{x}_i \sim Uni(W)$. Together with its location each individual is also given

an initial mark (size) $m_i(t_i^0) = m_i^0$ which can be taken as some fixed positive value (suitable when individuals are not observed until they have reached a certain size). Alternatively, one could draw m_i^0 from some distribution, for example the $Uni(0, \epsilon)$-distribution, $\epsilon > 0$, as in [16]. Note that at this stage, at each fixed t, the point process generated by the $\mathbf{x}_i$'s corresponds to a homogeneous spatial Poisson process with intensity αt, observed on W.

Once an individual arrives at W it instantly starts changing its size deterministically according to $m_i(t) = m_i^0 + \int_{t_i^0}^t dm_i(s)$, $t \geq t_i^0$, where

$$dm_i(t) \quad = \quad f\left(m_i(t); \theta\right) dt + \sum_{\substack{j \in \Omega_t \\ j \neq i}} h\left(m_i(t), m_j(t), \mathbf{x}_i, \mathbf{x}_j; \theta\right) dt. \qquad (1)$$

Here Ω_t is the index set comprising the individuals alive at time t, $f\left(m_i(t); \theta\right)$ is a function determining the individual growth of mark i in absence of competition with other (neighbouring) individuals and $h\left(m_i(t), m_j(t), \mathbf{x}_i, \mathbf{x}_j; \theta\right)$ is a function handling the individual's spatial interaction with other individuals. Note that it may happen that $m_i(t) \leq 0$ and once this happens we consider an individual to have died 'competitively', just as in [16].

As previously mentioned the so called natural deaths are governed by the death process which is defined as a simple death process having intensity function $\mu\rho(\cdot)$, $\mu > 0$, where $\rho(\cdot)$ is a function of the marks. This means that as time passes an individual's $Exp(\mu\rho(m_i(t)))$ -distributed remaining lifetime will change with its size. An alternative way of expressing the behaviour of the death process is to say that the conditional probability that an individual i dies naturally during $(t, t+dt)$ given $m_i(t)$ equals $\mu\rho(m_i(t))\, dt + o(dt)$. While [16] uses $\rho(m_i(t)) \equiv 1$ we here consider $\rho(m_i(t)) = 1/(1 + m_i(t))$, implying that individuals become more viable as they grow; a choice motivated by our forestry applications. If, on the contrary, one wishes to consider individuals who become less viable as they grow in size then, for instance, $\rho(m_i(t)) = m_i(t)/(1 + m_i(t))$ would be a better candidate.

The Von Bertalanffy-Chapman-Richards growth function has previously been used to model the development of the radii of isolated Scots pines [10]. This growth function has as special case the logistic growth function [14] and its shape resembles the shape of the Von Bertalanffy-Chapman-Richards growth function fitted in [10]. We therefore consider the logistic

growth function, given by

$$f\left(m_i(t);\theta\right) = \lambda\, m_i(t)\left(1 - \frac{m_i(t)}{K}\right), \tag{2}$$

both a good and a tractable candidate for our purposes (see e.g. [14] and [16]). Expression (2) contains the two parameters $\lambda > 0$ and $K > 0$ which, respectively, denote the growth rate of a mark and its upper bound (carrying capacity). If we consider an individual in absence of interacting neighbouring individuals then (1) together with (2) gives rise to the ordinary differential equation $dm_i(t)/dt = \lambda\, m_i(t)\,(1 - m_i(t)/K)$ for which the solution is given by

$$m_i(t) = \frac{K}{1 + (K/m_i^0 - 1)\,\mathrm{e}^{-\lambda t}}. \tag{3}$$

Note that (3) (and thereby (2)) requires that $m_i^0 > 0$.

Just as for the individual growth function the possible choices of spatial interaction functions are many (c.f. [8], [14] and [16] for examples of interaction functions and related discussions). Here, we consider the so called area interaction function, given by

$$h\left(m_i(t), m_j(t), \mathbf{x}_i, \mathbf{x}_j; c, r\right) = -c\,\frac{\nu\left(B\left[\mathbf{x}_i, rm_i(t)\right] \cap B\left[\mathbf{x}_j, rm_j(t)\right]\right)}{\nu\left(B\left[\mathbf{x}_i, rm_i(t)\right]\right)}, \tag{4}$$

where $c \in \mathbb{R}$ is the force of interaction and $r > 0$ is the scale of interaction. Furthermore, $B\left[\mathbf{x}_i, rm_i(t)\right]$ is a closed disk centred at $\mathbf{x}_i$ with radius $rm_i(t)$ and it is referred to as the 'influence zone' of the individual. Since competition for resources takes place only within influence zones ([3] and [20]), individuals i and j will compete only when their influence zones overlap, i.e. when $B\left[\mathbf{x}_i, rm_i(t)\right] \cap B\left[\mathbf{x}_j, rm_j(t)\right] \neq \emptyset$. This non-symmetric soft core interaction has the effect that large marks influence small marks more than the other way around, yet allowing the small marks to play their part. This interaction model is more realistic in tree modelling applications than symmetric interaction models ([14] and [16]). Depending on the choice of parameters, this area interaction function has the ability to generate regular as well as aggregated point patterns (despite the underlying uniform distribution of the locations) [13].

3 Estimation

An expression of the full likelihood function is not known for this model and although likelihood methods are generally highly desirable due to their asymptotic properties, under certain regularity conditions (see e.g. [2] and [19]), other more tractable estimation methods often generate estimates of similar quality. We here follow [16] by estimating $\theta = (\lambda, K, c, r)$ using the least squares approach. The death rate, μ, and the arrival intensity, α, are estimated separately by the ML-method.

Regarding the simulation of the process, [16] presents an algorithm where W is the unit square wrapped onto a torus and $\rho(m_i(t)) \equiv 1$, which is easily modified to suite any choice of $\rho(m_i(t))$ (in particular $\rho(m_i(t)) = 1/(1 + m_i(t))$) and any $W \subseteq \mathbb{R}^2$. When computing $m_i(t)$ it should be noted that one does not have to include all $j \in \Omega_t \setminus \{i\}$ in the sum in expression (1), rather only those within the maximal interaction range, i.e. $j \in \Omega_t \setminus \{i\}$ such that $||\mathbf{x}_i - \mathbf{x}_j|| \leq 2rK$.

Given that T_j and N_{T_j}, $j = 1, \ldots, n$, respectively, denote the jth sample time and the total number of individuals observed by time T_j, we let our data set be represented by $\mathbb{X} = \{\mathbb{X}(T_j)\}_{j=1}^n = \left\{ \left(\mathbf{x}_{ij}, m_{ij}, I_{ij}\right) : i = 1, \ldots, N_{T_j} \right\}_{j=1}^n$, where $\mathbf{x}_{ij} = \mathbf{x}_i(T_j)$, $m_{ij} = m_i(T_j)$ and $I_{ij} = I_i(T_j)$. The functions $I_i(\cdot)$, $i = 1, \ldots, N_{T_n}$, are indicator functions such that $I_i(t) = 1$ if individual i is alive at time t and $I_i(t) = 0$ if the individual is dead at time t. As before $\mathbf{x}_i(\cdot)$ and $m_i(\cdot)$ denote the location and the size of individual i, respectively. Note also that the index set comprising the individuals alive at time t can be written as $\Omega_t = \{i \in \{1, \ldots, N_t\} : I_i(t) = 1\}$.

3.1 Least squares estimation of λ, K, c, and r

Considering a set of parameters $\theta = (\lambda, K, c, r)$ and a configuration $\mathbb{X}(T_j)$, let $\tilde{m}_i\left(T_{j+1}; \theta, \mathbb{X}(T_j)\right)$, $i \in \Omega_{T_{j+1}}$, denote the prediction of $m_{i(j+1)}$ from $\mathbb{X}(T_j)$, based on calculating equation (1). If an individual has $\tilde{m}_i\left(T_{j+1}; \theta, \mathbb{X}(T_j)\right) > 0$ while $I_{i(j+1)} = 0$ it will be treated as having died by natural causes during (T_j, T_{j+1}). Our least squares estimates are then found by minimizing

$$S(\theta) := \sum_{j=1}^{n-1} \sum_{i \in \Omega_{T_j}} I_{i(j+1)} \left[\tilde{m}_i\left(T_{j+1}; \theta, \mathbb{X}(T_j)\right) - m_{i(j+1)}\right]^2,$$

with respect to $\theta = (\lambda, K, c, r) \in \mathbb{R}_+ \times \mathbb{R}_+ \times \mathbb{R} \times \mathbb{R}_+$.

In order to minimize $S(\theta)$ some optimization procedure is required. We here adopt an MCMC-type method (see [12]) where we start by choosing initial parameter estimates, i.e. let $\lambda = \lambda_0 > 0$, $K = K_0 > 0$, $c = c_0 \in \mathbb{R}$ and $r = r_0 > 0$, for which we calculate $S(\theta) = S(\lambda, K, c, r)$. We also define the step sizes $\delta_\lambda > 0$, $\delta_K > 0$, $\delta_r > 0$, and $\delta_c > 0$. Now, in each round we

1. randomly choose one of the parameters λ, K, r, c;
2. for our parameter of choice, say λ, let $\lambda' = \lambda + Z$, for Z drawn from $Uni(-\delta_\lambda, \delta_\lambda)$;
3. calculate $S(\theta') = S(\lambda', K, r, c)$;
4. if $S(\theta') < S(\theta)$ let $\lambda = \lambda'$, otherwise let $\lambda = \lambda$;
5. return to step 1.

We continue to run the algorithm until either $S(\theta)$ is less than some predefined minimum value, say, $S_{min} = 10^{-5}$ or until we have not seen any decrease in $S(\theta)$ for a predefined number of consecutive runs, say, $N_{max} = 200$. We let our final estimates $\hat{\theta} = (\hat{\lambda}, \hat{K}, \hat{c}, \hat{r})$ be given by the last θ obtained in the algorithm above. Note that we here utilize the information obtained in the previous step in order to stepwise get closer to the final estimate.

When minimizing $S(\theta)$, in the case of a simulated data set, it can be seen that $S(\theta)$ may not attain its minimum at the true parameter set but instead at some biased θ. This 'incorrect' shape of $S(\theta)$ is mainly due to edge effects (discussed further in Section 5) and dependence between certain parameters. For instance, two different sets of c and r may result in similar interactions, due to the form of (4). In order to control the estimation routine, so that this risk of bias is reduced, our approach is to find good starting values, $(\lambda_0, K_0, c_0, r_0)$, (as opposed to arbitrarily chosen ones) and to choose sensible step sizes, δ_λ, δ_K, δ_c, δ_r. The exact forms and derivations of these are given in Appendix A.1 and A.2.

3.2 Estimation of μ

Let $f_{L_k}(t_k|\mu)$, $k = 1, \ldots, n_T$, denote the densities of the random lifetimes $L_1, \ldots, L_{n_T}$ (observed as $t_1, \ldots, t_{n_T}$) of the n_T individuals who have died from

natural causes by time T (determined during the minimization of $S(\theta)$), given some natural death rate function $\mu\rho(m_i(t))$, and let $t^0_{i(L_1)}, \ldots, t^0_{i(L_{n_T})}$ denote the birth times of the individuals having these life times. Also, under the same natural death rate regime, let $S_1, \ldots, S_{m_T}$ denote the m_T random life-times of the individuals who are still alive at time T (observed as $s_1, \ldots, s_{m_T}$). Then the likelihood of the death rate, μ, is (approximately) given by

$$\begin{aligned} L(\mu) &= \prod_{k=1}^{n_T} f_{L_k}(t_k|\mu) \prod_{l=1}^{m_T} \mathbb{P}(S_l > s_l|\mu) \\ &= \prod_{k=1}^{n_T} \mu\rho\left(m_{i(L_k)}\left(t^0_{i(L_k)} + t_k\right)\right) \exp\left\{-\mu\rho\left(m_{i(L_k)}\left(t^0_{i(L_k)} + t_k\right)\right) t_k\right\} \\ &\quad \times \prod_{l=1}^{m_T} \mathrm{e}^{-\mu\rho(m_{i(S_l)}(T))s_l}, \end{aligned}$$

where $m_{i(L_k)}(t)$ denotes the observed mark, at time t, of the individual having life time L_k. Similarly $m_{i(S_l)}(T)$ denotes the observed mark size at time T, of the individual having lived time S_l at time T. By solving with respect to μ in $\mathrm{d}\log\left(L(\mu)\right)/\mathrm{d}\mu = 0$ we get the ML-estimator

$$\hat{\mu} = n_T \Bigg/ \left(\sum_{k=1}^{n_T} \rho\left(m_{i(L_k)}\left(t^0_{i(L_k)} + t_k\right)\right) t_k + \sum_{l=1}^{m_T} \rho(m_{i(S_l)}(T))s_l \right). \quad (5)$$

In the case of $\rho(m_i(t)) \equiv 1$ this reduces to the estimator, $\hat{\mu}_0$, found in [16]. Since we sample the process only at $0 = T_0 < T_1 < \ldots < T_n = T$, neither the actual death times, $t^0_{i(L_k)} + t_k$, $k = 1, \ldots, n_T$, nor the sizes at these death times, $m_{i(L_k)}(t^0_{i(L_k)} + t_k)$, $k = 1, \ldots, n_T$, will be known. Recall that we label an individual as naturally dead once the predicted mark $\tilde{m}_i\left(T_{j+1}; \theta, m_i(T_j)\right) > 0$ while $I_{i(j+1)} = 0$, during the calculation of $S(\theta)$. Let $T_{j,i(L_k)}$ be the last sample time at which individual $i(L_k)$ was observed alive and let $\tilde{m}_{i(L_k)}(T_{j,i(L_k)})$ denote the prediction of its mark at $T_{j,i(L_k)}$. This censoring forces us to approximate (5) by

$$\begin{aligned} \hat{\mu}_1 = n_T \Bigg/ \Bigg(& \sum_{k=1}^{n_T} \rho\left(\tilde{m}_{i(L_k)}\left(T_{j,i(L_k)}\right)\right)\left(T_{j,i(L_k)} - t^0_{i(L_k)}\right) \\ & + \sum_{l=1}^{m_T} \rho(m_{i(S_l)}(T))\left(T - t^0_{i(S_l)}\right) \Bigg). \end{aligned} \quad (6)$$

As pointed out earlier, the process is observed only at the sampled time points $0 = T_0 < T_1 < \ldots < T_n = T$ so that the actual birth times (and death times) of the individuals remain unknown. Conditioned on the number of individuals arriving during $(T_{j-1}, T_j]$ the arrival times of the individuals will be uniformly distributed on (T_{j-1}, T_j) (see e.g. [9]). Thus, when estimating μ, for each interval $(T_{j-1}, T_j]$, we simulate $\Delta N_{T_{j-1}} = N_{T_j} - N_{T_{j-1}}$ birth times having a $Uni(T_{j-1}, T_j)$ distribution, provided that $\Delta N_{T_{j-1}} > 0$, which in turn are assigned to all individuals being observed for the first time at T_j. The question regarding which arrival time to assign to which individual is solved by giving the first arrival time to the individual who is the largest at time T_j, the second arrival time to the individual which is the second largest at time T_j and so forth. This will have the consequence that the life times will be random. Hence, by repeating this procedure a suitable number of times, each time simulating new random birth times, we generate a set of estimates of μ which are used to estimate a standard error for $\hat{\mu}$.

3.3 Estimation of α

Let $B(t) \geq 0$ denote the actual number of immigrants by time t and let $N_{T_j} = \left|\cup_{j=1}^{n} \Omega_{T_j}\right|$, $j = 1, \ldots, n$, denote the number of individuals observed at sample times up to T_j. Concerning the estimation of α, the approach of [16] is to ignore all the unobserved individuals who arrive and die within the same time interval (T_j, T_{j+1}), resulting in the immigration-increments $\Delta B(T_{j-1}) = \Delta N_{T_{j-1}}$, $j = 1, \ldots, n$, where $\Delta B(T_{j-1}) = B(T_j) - B(T_{j-1})$ and $\Delta N_{T_{j-1}} = N_{T_j} - N_{T_{j-1}}$. Since $B(t)$ is a Poisson$(\alpha\nu(W))$-process, its (independent) increments are $Poisson\,(\alpha(T_{j+1} - T_j)\nu(W))$-distributed. This being the scenario, an ML-estimator for α (see [16]) is provided by

$$\hat{\alpha}_0 = \frac{N_{T_n}}{T_n \nu(W)}. \tag{7}$$

This estimator is unbiased under the hypothesis that $N^{obs}(t) = B(t)$ since $\mathbb{E}[N_{T_n}/T_n\nu(W)] = \mathbb{E}[N_{T_n}]/T_n\nu(W) = \alpha T_n \nu(W)/T_n\nu(W) = \alpha$. This approach, however, underestimates α since we do not account for the individuals who arrive and die in the same sample interval, (T_k, T_{k-1}), (see [16]).

One possible way of partially compensating for this bias is to add to each increment of the observed process, $\Delta N_{T_{j-1}}$, the expected number of

individuals suffering a natural death among the expected number of individuals arriving during (T_{j-1}, T_j). Since the expected number of arrivals during (T_{j-1}, T_j) is unknown it will be replaced by an estimate hereof, provided by expression (7). Regarding the expected number of natural deaths, provided by μ, it will be governed by $\hat{\mu}$, the estimate of μ found in the previous subsection. The estimator takes the form

$$\hat{\alpha} = \underbrace{\frac{N_{T_n}}{T_n \nu(W)}}_{=\hat{\alpha}_0} + \frac{1}{T_n \nu(W)} \sum_{j=1}^{n} \left\lfloor N_{T_n} \frac{\Delta T_{j-1}}{T_n} \left(1 - \mathrm{e}^{-\hat{\mu}\rho(m_i^0)\Delta T_{j-1}}\right) \right\rfloor,$$

where $\lfloor x \rfloor$ denotes the integer part of x and $\Delta T_{j-1} = T_j - T_{j-1}$. The derivation of the estimator as well as some characteristics of it and its relation to $\hat{\alpha}_0$ can be found in Appendix A.3.

4 Data

Before presenting the edge correction methods we will introduce the specific tree data set under consideration. The data set we consider consists of measurements of locations and diameters at breast height (dbh) in a west Swedish Scots pine stand[1]. Recordings have been made in the years $T_1 = 1985$, $T_2 = 1990$, and $T_3 = 1996$ and the approximate age of the stand in 1985 was 22 years, thereby setting $T_0 = 1963$. Note that only the time intervals in which births and deaths occur are known, leaving the actual birth and death times unknown. All measurements have been made on a circular region of radius 10 meters where trees having reached 0.01 m dbh are included in the data set. Figure 1 illustrates plots of the data set with scaled radii (factor 10), for improved visualization, together with the appurtenant radius histograms.

Note how the size histogram tends to change as time elapses, with an increasing number of large trees. This is further confirmed by Table 1.

The RS-model has previously been fitted to data sets such as this [16]. However, since the number of trees present at each time point is fairly low it is important to take the edge effects into account, i.e. we have to somehow, for each sample time, estimate the behaviour of the unobserved trees surrounding

[1] Area number ("Trakt") 1562, Stand number ("Pålslag") 2060 - The "Lilla Edet" area.

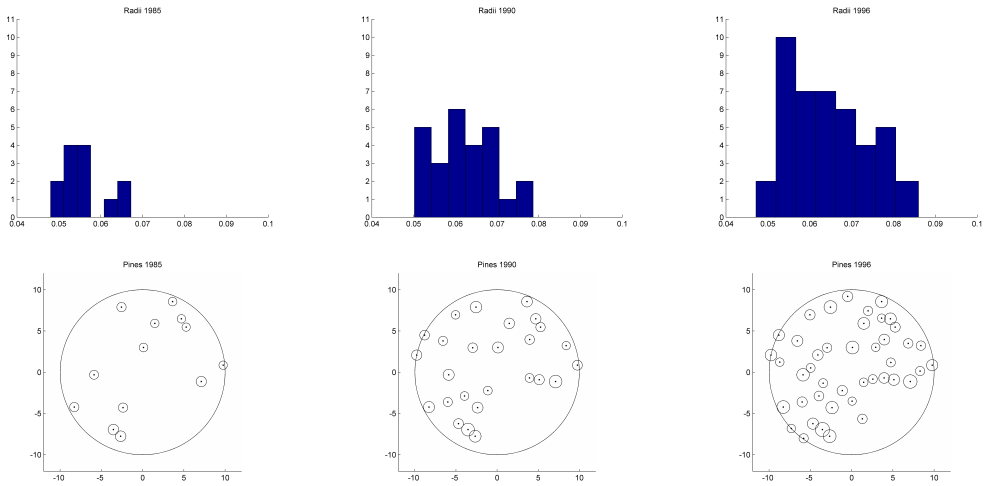

Figure 1: Swedish Scots pines recorded in 1985 (left), 1990 (middle) and 1996 (right). Upper row: Histograms of the radii. Lower row: Locations of the pines with scaled radii (factor 10).

T_j	1984	1990	1996
N_{T_j}	13	26	43
Mean radius	0.0557	0.0619	0.0640
Radius s.d.	0.0050	0.0074	0.0096
$\max_{i \in \Omega_{T_j}} m_{ij}$	0.0645	0.0775	0.0860

Table 1: Total number of trees, estimated mean, estimated standard deviation (s.d.) and maximum of the Scots pine radii at each sample time.

our region of interest. Given this estimated information one can then correct the estimates such that the unobserved interaction between the region of interest and its surrounding area is compensated for.

5 Spatio-temporal edge correction

When sampling real data, $\mathbb{X}$, one usually considers all individuals within some region A (here circular) which is part of some larger region W. The individuals in A interact with each other but simultaneously also with the individuals present outside A, i.e. the individuals in $B = W \setminus A$. So, if one were to estimate some statistics and/or model parameters in a situation where the interaction among (neighbouring) individuals plays a role, by only taking into consideration the individuals in A the estimators may generate biased estimates since the interaction between the individuals in A and those in B would be neglected. The effects of the absence of the information regarding this interaction are commonly referred to as *edge effects*. The risk that the edge effects generate biases rapidly increases when one deals with small quantities of data in A, as is the case with our tree data set introduced in Section 4. Hence, some type of correction method is needed (see e.g. [4], [7] and [21]).

A simple edge correction method would be the so called minus sampling method (see e.g. [17]). First one finds all individuals who fall within a buffer zone, $C \subseteq A$, consisting of all points $\mathbf{x} \in A$ located less than some distance $d_0 > 0$ from the boundary of A. Then one carries out the estimation based only on the individuals in $A \setminus C$, yet taking into account the locations and marks of the individuals in C. In doing this we let the individuals in C and $A \setminus C$ affect each other, yet basing the computation of the statistic or the parameter estimate in question only on the individuals in $A \setminus C$. However, in situations where there is a limited amount of data in region A, as in our pine data set, removing data is not an option and this method therefore is not applicable.

A more sensible way of doing (spatio-temporal) edge correction in situations where there is little data available is to utilize the features of the parametric model which one attempts to fit to the data. We here give the idea behind the edge correction methods presented in this section. One starts

by finding initial (possibly biased) estimates of the model parameters, $\hat{\Theta}_*$, based on our original data set (region A). Then, under the regime of $\hat{\Theta}_*$, we wish to find the expected model behaviour when restricted to region B (possibly conditioned on the actual data in A), $\mathbb{E}_{\hat{\Theta}_*}[\mathbb{X}_{[0,T]}|_B]$. By doing so we wish to establish the expected interaction between $\mathbb{E}_{\hat{\Theta}_*}[\mathbb{X}_{[0,T]}|_B]$ and the individuals in region A. With $\mathbb{E}_{\hat{\Theta}_*}[\mathbb{X}_{[0,T]}|_B]$ at hand we now re-estimate the model parameters from the actual data (region A), however, this time allowing for $\mathbb{E}_{\hat{\Theta}_*}[\mathbb{X}_{[0,T]}|_B]$ to interact with the actual data during the estimation. Once these new estimates have been obtained, we let them replace $\hat{\Theta}_*$ and repeat the above procedure again. By continuing in this fashion we have an iterative procedure which we stop once it has fulfilled a certain predefined convergence criterion.

The question still remains, however, regarding how to find the expected behaviour, $\mathbb{E}_{\hat{\Theta}_*}[\mathbb{X}_{[0,T]}|_B]$. We here suggest three methods based on the idea described above where $\mathbb{E}_{\hat{\Theta}_*}[\mathbb{X}_{[0,T]}|_B]$ is estimated from successive simulations of a (possibly interacting) version of our parametric model, restricted to region B. All three methods are similar to the ideas presented by Geyer in [1] in the sense that they all use simulated data as interacting data in region B. At each iteration step, at the sample times $T_1, \ldots, T_n$, all three methods sample a series of simulated process realisations which all live in region B. Thereafter each such sampled simulated outer realisation is combined with the actual data, $\mathbb{X}$, to form a full data set, $\mathbb{X}^*$, on $W = A \cup B$. From each new data set $\mathbb{X}^*$ we carry out our estimation procedure, however, as opposed to using the full data set $\mathbb{X}^*$ in the estimation we here only include $\mathbb{X}$ (region A) in the calculation of the estimates/statistics while we simultaneously let the (simulated) individuals in B interact with $\mathbb{X}$ (thereby influencing the estimates generated from $\mathbb{X}$). Now, in a given iteration step, by averaging over all estimates generated from each simulated outer realisation we get the final estimates for that specific iteration (this is how the simulated outer regions are considered to create an estimate of $\mathbb{E}_{\hat{\Theta}_*}[\mathbb{X}_{[0,T]}|_B]$ and its interaction with $\mathbb{X}$). This averaged set of estimates now replaces $\hat{\Theta}_*$ and by repeating the whole procedure once again we have executed the next iteration step.

A further question, yet to be explained in detail, is the stopping criterion used in the algorithms. Note that the estimates may be vector-valued. For each of the algorithms, given that we use N simulated outer realisations in each iteration, we will keep running it until the estimates, $\hat{\Theta}_*$, generated in

two consecutive iterations differ by at most a distance $\epsilon > 0$. Once this has occurred we save these estimates and run the algorithm for another $M - 1$ iterations and average over the M estimates hereby generated, in order to get our final estimates. Another possible stopping criterion which may be used is the following. We run M iterations of our edge correction, hence generating a set of M estimates, $\Xi_1 = \{\hat{\Theta}_*^1, \ldots, \hat{\Theta}_*^M\}$, for which we estimate the variance, $\hat{\sigma}_1^2 = \widehat{Var(\Xi_1)}$, component wise. By running one more iteration of the edge correction, thus getting a new vector of estimates, $\hat{\Theta}_*^{M+1}$, we create the set $\Xi_2 = \{\hat{\Theta}_*^2, \ldots, \hat{\Theta}_*^M, \hat{\Theta}_*^{M+1}\}$ for which we estimate the variance, $\hat{\sigma}_2^2 = \widehat{Var(\Xi_2)}$. We continue in this fashion, i.e. creating $\Xi_{i+1} = (\Xi_i \setminus \{\hat{\Theta}_*^i\}) \cup \{\hat{\Theta}_*^{M+i}\}$, $i = 2, 3, \ldots$, to get $\hat{\sigma}_{i+1}^2 = \widehat{Var(\Xi_{i+1})}$, until $||\hat{\sigma}_{i+1}^2|| < \epsilon$ for some $\epsilon > 0$, where $||\cdot||$ is the Euclidean norm. Since the second approach considers the variation of a large number of estimates it is generally preferable to the first method. However, the first stopping criterion is less computationally demanding than the second one (since we have to wait M iterations before we can judge whether to stop or not in the second one) and it does a good enough job for the illustrative purposes we have here. Hence, in what follows we choose to apply the first of the two stopping criteria.

We will use the remainder of the section to present, discuss and evaluate the different methods.

5.1 Edge correction methods

The three edge correction methods we will present are explained for the RS-model but they may be applied to other spatial and spatio-temporal (marked) point processes as well. In the algorithms presented here the large rectangular window W will be wrapped onto a torus when we generate the individuals in the outer region, B, (see e.g. [4], [15], [11] or [21]). Recall that we sample the process as $\mathbb{X} = \{\mathbb{X}(T_j)\}_{j=1}^n = \left\{\left(\mathbf{x}_{ij}, m_{ij}, I_{ij}\right) : i = 1, \ldots, N_{T_j}\right\}_{j=1}^n$, where $\mathbf{x}_{ij} = \mathbf{x}_i(T_j)$, $m_{ij} = m_i(T_j)$ and $I_{ij} = I_i(T_j)$ is an indicator function such that $I_i(t) = 1$ if individual i is alive at time t and $I_i(t) = 0$ if the individual is dead at time t. Also recall that $\tilde{m}_i\left(T_{j+1}; \theta, \mathbb{X}(T_j)\right)$, $i \in \Omega_{T_{j+1}}$, denotes the prediction of $m_{i(j+1)}$ from $\mathbb{X}(T_j)$ generated by equation (1), under $\theta = (\lambda, K, c, r) \in \mathbb{R}_+ \times \mathbb{R}_+ \times \mathbb{R} \times \mathbb{R}_+$.

For a general process the three edge correction methods would have

been presented in such a way that the whole parameter set would have been considered in each iteration. But in the case of the RS-model we may in fact omit the re-estimation of μ and α since their estimates tend not to change significantly between two iterations, despite the fact that we anew label individuals as naturally dead once $S(\theta)$ is evaluated for a new set of parameters θ, possibly leading to other life-times (of the naturally dead individuals) used in the estimator generating the new estimate $\hat{\mu}$ (hence also leading to a new estimate $\hat{\alpha}$). Note that below $||\hat{\theta}_* - \hat{\theta}||$ represents the Euclidean distance between $\hat{\theta}_*$ and $\hat{\theta}$.

5.1.1 Simple simulation of the outer region

We here present the first of the three methods; an algorithm which illustrates the basic idea on which all three methods are based.

1. Choose some small $\epsilon > 0$ and positive integers M and N.
2. Estimate the parameters from the data set $\mathbb{X}$ (region A) to generate a set of (non-edge-corrected) estimates $\hat{\theta}_* = \left(\hat{\lambda}_*, \hat{K}_*, \hat{c}_*, \hat{r}_*\right)$.
3. For $i = 1, \ldots, N$:
   (a) Simulate the process on $W = A \cup B$, based on $\hat{\theta}_*$ and $(\hat{\mu}, \hat{\alpha})$, and sample it at $T_1, \ldots, T_n$ (where W is wrapped onto a torus).
   (b) Create the data set $\mathbb{X}^*$ by removing what has been simulated in region A (for the sample times $T_1, \ldots, T_n$) and then replacing it with the data, $\mathbb{X}$.
   (c) Least squares estimation of $\theta = (\lambda, K, c, r)$ based on $\mathbb{X}^*$:
   Minimize

$$S(\theta) = \sum_{j=1}^{n-1} \sum_{i \in \Omega_{T_j} \cap \{k \in \mathbb{Z}_+ : \mathbf{x}_k \in A\}} I_{i(j+1)} \left[\tilde{m}_i\left(T_{j+1}; \theta, \mathbb{X}(T_j)\right) - m_{i(j+1)}\right]^2$$

   w.r.t. θ to get the estimates in this iteration, $\hat{\theta}_i = (\lambda_i, K_i, c_i, r_i)$.
   Note that we include only the individuals in $\mathbb{X}$ (region A) in the sum of squares $S(\theta)$. Also note that we must generate the predictions $\tilde{m}_i\left(T_{j+1}; \theta, \mathbb{X}(T_j)\right)$ for all the individuals in $\mathbb{X}^*$ (the individ-

uals in B in A hereby interact) each time we evaluate $S(\theta)$ for a new θ.

4. Calculate $\hat{\theta} = \left(\frac{1}{N} \sum_{i=1}^{N} \lambda_i, \frac{1}{N} \sum_{i=1}^{N} K_i, \frac{1}{N} \sum_{i=1}^{N} c_i, \frac{1}{N} \sum_{i=1}^{N} r_i \right)$.

5. If $\left|\left| \hat{\theta}_* - \hat{\theta} \right|\right| < \epsilon$ set $\hat{\theta}^{(1)} = \hat{\theta}$ and go to step 6, otherwise go to step 3. Also set $\hat{\theta}_* = \hat{\theta}$.

6. For $j = 1, \ldots, M-1$:

   (a) Repeat steps 3 and 4 to generate the estimates $\hat{\theta}$ and set $\hat{\theta}_* = \hat{\theta}$.

   (b) Denote these estimates by $\hat{\theta}^{(j)} = \left(\lambda^{(j)}, K^{(j)}, c^{(j)}, r^{(j)}\right)$.

7. Let the final estimates be given by

$$\hat{\theta} = \left(\frac{1}{M} \sum_{i=1}^{M} \lambda^{(j)}, \frac{1}{M} \sum_{i=1}^{M} K^{(j)}, \frac{1}{M} \sum_{i=1}^{M} c^{(j)}, \frac{1}{M} \sum_{i=1}^{M} r^{(j)} \right).$$

Since the algorithm averages over all estimates $\hat{\theta}_1, \ldots, \hat{\theta}_N$ in a given iteration, it reduces the risk of having surrounding areas of too artificial nature generating the estimates. For instance, it is possible that some large individual(s) in B, close to the boundary of A, end up within the interaction range of some large individual(s) in A for a given simulated surrounding area. Such a scenario would not be encountered if the two individuals had been interacting naturally with each other throughout time. The algorithm above, through its averaging effect, reduces the strong impact which an extreme situation such as the aforementioned may have on some of the estimates.

5.1.2 Rotations of the outer region

We now consider a modifications of the previous algorithm which differs in the way it generates the surrounding realisations. Instead of simulating several outer realisations at each iteration, the idea here is that we instead use only one simulated outer region which we rotate a number of times, relative to the actual data, $\mathbb{X}$. By combining $\mathbb{X}$ with each rotation of the outer region we get a series of full data sets on W on which we base the estimation.

More specifically we replace step 2 and step 3 in the algorithm presented in Section 5.1.1 by

2*. Estimate the parameters from the data set $\mathbb{X}$ (region A) to generate a set of (non-edge-corrected) estimates $\hat{\theta}_* = \left(\hat{\lambda}_*, \hat{K}_*, \hat{c}_*, \hat{r}_*\right)$.

Choose the angles $\omega_1 < \ldots < \omega_N$ either according to $\omega_{i+1} - \omega_i = 2\pi/N$ or $\omega_i \sim Uni(0, 2\pi)$. For all $i = 1, \ldots, N$, perform counterclockwise rotations (around the centre of A) of all locations, $\mathbf{x}_k = (x_k, y_k)$, in $\mathbb{X}$:

$$\mathbf{x}_k(\omega_i) = (x_k \cos(\omega_i) - y_k \sin(\omega_i), x_k \sin(\omega_i) + y_k \cos(\omega_i)) .$$

We get the rotated data sets $\mathbb{X}_{\omega_1}, \ldots, \mathbb{X}_{\omega_N}$.

3*. Simulate the process on $W = A \cup B$, based on $\hat{\theta}_*$ and $(\hat{\mu}, \hat{\alpha})$, and sample it at $T_1, \ldots, T_n$ (where W is wrapped onto a torus).

For $i = 1, \ldots, N$:

1. Create the data set $\mathbb{X}^*_{\omega_i}$ by removing what has been simulated in region A (for the sample times $T_1, \ldots, T_n$) and then replacing it with the rotated data, $\mathbb{X}_{\omega_i}$.
2. Least squares estimation of $\theta = (\lambda, K, c, r)$ based on $\mathbb{X}^*_{\omega_i}$:
   Minimize
   $$S(\theta) = \sum_{j=1}^{n-1} \sum_{i \in \Omega_{T_j} \cap \{k \in \mathbb{Z}_+ : \mathbf{x}_k(\omega_i) \in A\}} I_{i(j+1)} \left[\tilde{m}_i\left(T_{j+1}; \theta, \mathbb{X}(T_j)\right) - m_{i(j+1)}\right]^2$$
   w.r.t. θ to get the estimates in this iteration, $\hat{\theta}_i = (\lambda_i, K_i, c_i, r_i)$.
   Note that we include only the individuals in $\mathbb{X}$ (region A) in the sum of squares $S(\theta)$. Also note that we must generate the predictions $\tilde{m}_i\left(T_{j+1}; \theta, \mathbb{X}(T_j)\right)$ for all the individuals in $\mathbb{X}^*_{\omega_i}$ (the individuals in B in A hereby interact) each time we evaluate $S(\theta)$ for a new θ.

As mentioned in step 2* one possibility is to use random angles. Although this adds an extra component of randomness to the procedure it has the drawback of allowing for situations where two or more of the angles become nearly the same, hence increasing the risk of the type of extreme estimates mentioned in Section 5.1.1. We therefore choose not to evaluate the version with random angles any further.

5.1.3 Outer region influenced by the growth of the data

Instead of rotating the surrounding area to avoid estimates based on the artificial surroundings described in Section 5.1.1 one may choose to condition on the development of the individuals in $\mathbb{X}$ (region A) when generating the surrounding individuals in region B. Our third edge correction method tries to overcome the problem of these artificial surroundings by letting the actual data individuals enter region A and directly start growing, alongside the simulation of the surrounding individuals which takes place in region B. During this growth the individuals in region A are allowed to influence the development of the individuals in region B but not the other way around. By doing this we try to mimic the actual underlying growth scenario.

Since the actual arrival times and the exact growth patterns remain unknown, an individual will enter A at an arrival time simulated uniformly over the sample time interval in which it was first observed (jumps of a Poisson process are uniformly distributed over time intervals) and then grow linearly between its observed sizes at the sample times so that it (possibly) affects the growth of the simulated surrounding individuals. The exact algorithm is given by replacing steps 2 and 3 in the algorithm of Section 5.1.1 by

2**. Estimate the parameters from the data set $\mathbb{X}$ (region A) to generate a set of (non-edge-corrected) estimates $\hat{\theta}_* = \left(\hat{\lambda}_*, \hat{K}_*, \hat{c}_*, \hat{r}_*\right)$.

For each time interval $(T_{j-1}, T_j]$, $j = 1, \ldots, n$, we observe ΔN_{j-1} new individuals. Simulate $Uni(t_{j-1}, t_j)$-distributed birth times $b_1^j < \ldots < b_{\Delta N_{j-1}}^j$ and assign these to the individuals in $\mathbb{X}$ (region A) who have arrived in $(T_{j-1}, T_j]$ in such an order that the largest individual gets the smallest time, going upwards until the smallest individual has received the largest time.

3**. For $i = 1, \ldots, N$:

(a) Simulate the process, based on $\hat{\theta}_*$ and $(\hat{\mu}, \hat{\alpha})$, but now only on the region $B = W \setminus A$ (where W is wrapped onto a torus) and sample it at $T_1, \ldots, T_n$. Furthermore, during the simulation, let each individual in $\mathbb{X}$ (region A) enter at its simulated birth time, b_k^j, $k = 1, \ldots, \Delta N_{j-1}$, $j = 1, \ldots, n$, and grow linearly between the sample time points until the last sample time point, T_j, it has

been observed alive. This will have the consequence that these linearly growing individuals will have their actual (observed) sizes at the sample times. The effect acquired here is that the data, $\mathbb{X}$, will affect the growth of the simulated individuals in B (but not the other way around). Refer to this (partially) simulated data set as $\mathbb{X}^*$. Note that the only individuals in $\mathbb{X}^*$ located in A are the ones found in $\mathbb{X}$.

(b) Least squares estimation of $\theta = (\lambda, K, c, r)$ based on $\mathbb{X}^*$:
Minimize

$$S(\theta) = \sum_{j=1}^{n-1} \sum_{i \in \Omega_{T_j} \cap \{k \in \mathbb{Z}_+ : \mathbf{x}_k \in A\}} I_{i(j+1)} \left[\tilde{m}_i \left(T_{j+1}; \theta, \mathbb{X}(T_j) \right) - m_{i(j+1)} \right]^2$$

w.r.t. θ to get the estimates in this iteration, $\hat{\theta}_i = (\lambda_i, K_i, c_i, r_i)$. Note that we include only the individuals in $\mathbb{X}$ (region A) in the sum of squares $S(\theta)$. Also note that we must generate the predictions $\tilde{m}_i \left(T_{j+1}; \theta, \mathbb{X}(T_j) \right)$ for all the individuals in $\mathbb{X}^*$ (the individuals in B in A hereby interact) each time we evaluate $S(\theta)$ for a new θ.

5.2 Evaluation of the estimation methods

In order to be able to evaluate the estimation methods previously presented we simulate what we here will refer to as a 'test set', consisting of a simulated realisation of the process on $W = [0, 30] \times [0, 30]$ (wrapped onto a torus), using step size $dt = 0.01$. We include those individuals alive at the sample times $T_1 = 22$, $T_2 = 27$, $T_3 = 33$ (the age of our pine stand at its sample time points) who are located within the circular region $A = \{\mathbf{y} \in W : ||\mathbf{y} - (15, 15)|| \leq 10\}$. The parameters used are $K = 0.1$, $\lambda = 0.08$, $c = 2$, $r = 2$, $\alpha = 0.007$, $\mu = 0.02$, and $m_i^0 = 0.05$. In order to check the accuracy of our estimation techniques we re-estimate the parameters generating the test set. We do not estimate m_i^0, but instead treat it as known since in forest stands one mostly knows the minimal tree radius from which measurements are being made (see Section A.1 for its estimation). The specific choice of parameters used to generate the test set was made since it generates realisations which resemble our tree data set. However, all methods

we here apply have been evaluated for a range of different parameter values and the results obtained have been similar to those obtained for the test set.

If one gradually decreases the size of W in a series of edge corrections the distance on the torus between some of the individuals in B gradually decreases. This is particularly the case for those individuals located close to and on opposite sides of the boundary of A (the individuals interacting the strongest with $\mathbb{X}$). In this gradual decrease these individuals start interacting more strongly with each other and thereby increasingly inhibit each other's growths, resulting in a gradual decrease in the edge correcting effect. Hence, a small W results in a slow convergence to the final estimates, whereas, a very large W makes the edge corrections computationally demanding. When we edge correct the re-estimation of the test set parameters we have chosen W to be a square region with side length 25, a choice purely based on trials.

Standard error estimates are obtained by re-running the edge correction procedure of choice a large number of times. However, in situations where this is computationally demanding, some resampling technique may be used to obtain the standard error estimates. For each edge correction method we have considered 10 different estimation runs where each of these uses the last $M = 4$ iterations, once $||\hat{\theta}_* - \hat{\theta}|| < \epsilon = 1$, in order to create the averaged final estimates and in each iteration we have considered $N = 3$ simulated surroundings ($N = 3$ angles in the case of the rotation-correction). Furthermore, on the basis of trials we have concluded that for each simulated surrounding it is sufficient to run the edge corrected estimation procedures until no change in $S(\theta)$ has been observed for $N_{max} = 50$ consecutive runs.

Table 2 presents both the initial estimates (see Appendix A.1), the final estimates found when applying no edge correction (stopping criterion for the minimisation of $S(\theta)$, $N_{max} = 3000$) and the estimates found for each of the edge corrections. The estimates of μ and α are given by $\hat{\mu} = 0.0113$ and $\hat{\alpha} = 0.00637$ ($(\hat{\mu} - \mu)/\mu = -43.5\%$ and $(\hat{\alpha} - \alpha)/\alpha = -9.0\%$).

	λ	K	c	r
True value	0.08	0.1	2	2
Initial	0.1006	0.0933	0.007	2.75
Bias	0.0206	−0.0067	−1.9930	0.75
Bias (%)	25.8%	-6.7%	-99.7%	37.5%
Uncorrected	0.0822	0.0995	5.4991	1.8301
Bias	0.0022	−0.0005	3.4991	−0.1699
Bias (%)	2.80%	-0.50%	174.96%	-8.50%
Simple				
Est. mean	0.0822	0.0996	2.7978	1.8694
Bias	0.0022	-0.0004	0.7978	-0.1306
Bias (%)	2.80%	-0.43%	39.89%	-6.53%
Est. s.e.	0.0001	0.0001	0.7755	0.1572
Rotations				
Est. mean	0.0821	0.0995	2.8364	1.7614
Bias	0.0021	-0.0005	0.8364	-0.2386
Bias (%)	2.58%	-0.46%	41.82%	-11.93%
Est. s.e.	0.0004	0.0001	0.5439	0.1416
Influenced growth				
Est. mean	0.0823	0.0996	2.7499	1.7926
Bias	0.0023	-0.0004	0.7499	-0.2074
Bias (%)	2.86%	-0.36%	37.50%	-10.37%
Est. s.e.	0.0005	0.0002	0.5422	0.1437

Table 2: Test set estimates: Initial estimates, non-edge corrected estimates ($N_{max} = 3000$) and estimates obtained through the different edge corrections. We have run each edge correction 10 times in order to get the estimated mean values and standard errors (s.e.). In each run we have used $N = 3$ (simulated surroundings/rotations), $N_{max} = 50$ (stopping criterion for the minimisation of $S(\theta)$), $\epsilon = 1$ (convergence criterion) and $M = 4$ (number of final iterations).

As one can see in the uncorrected estimation, the biases for the estimates of λ, K, r and α are fairly moderate. This, however, cannot be said about c and μ and regarding the under-estimation of μ there is little to be done. The large over-estimation obtained for c in the uncorrected estimation, however, is mainly a result of the edge effects which we correct for. Furthermore, we also see that the small biases of the estimates of λ and K tend not to

change significantly from the uncorrected estimates. The influenced growth correction manages to reduce the bias of $\hat{c}$ slightly more than the other two methods but this comes with a trade off in the form of an increased underestimation of r, compared to both the uncorrected estimates and the simple correction estimates (the main reason being the strong dependence between c and r). A possible reason that the influenced growth generally performs the best in the estimation of c is that it actually takes into consideration the (approximate) behaviour of the actual data and it therefore restricts the previously mentioned artificial surroundings more than the other two methods. The simple correction is the only one of the three methods which reduces the r-bias but it is also the method giving the highest standard error estimates for c and r. As we see the rotation correction performs slightly worse than the other two methods but it has the advantage of reducing the computational time compared to the other two methods. By increasing the number of rotations one may be able to decrease the bias, but this comes with an increase in computation time. If no edge corrections are used, the points of the (data) point patterns likely will have less close neighbours than in reality. This will result in too large estimates of c, which in turn will result in more regular point patterns since c to a large extent controls the regularity of the point patterns generated by the process.

Table 3 gives us the results obtained after the first iteration. Note that the large bias generated by the uncorrected estimate of c directly is reduced by each of the methods.

Since our main concern is correcting the estimate of c, choosing $\epsilon = 1$ more or less implies that the final M iterations start once $|\hat{c}_* - \hat{c}| < \epsilon = 1$. By increasing ϵ a bit one may think that the final estimates get very different. However, since a substantial reduction takes place already after the first iteration and since we average over the final M iterations, if we were to choose ϵ a bit larger than 1 this in fact does not change the results drastically. Note further that one can start with a given ϵ and then increase it after a couple of iterations if the fluctuations between consecutive iterations are larger than initially believed (i.e. if $||\hat{\theta}_* - \hat{\theta}|| < \epsilon$ does not occur). The average number of iterations that were needed in order to reach $||\hat{\theta}_* - \hat{\theta}|| < \epsilon = 1$ in the 10 runs are 2.6 for the simple correction, 3.2 for the rotation correction and 2.8 for the influenced growth correction.

Iteration 1	λ	K	c	r
True value	0.08	0.1	2	2
Uncorrected	0.0822	0.0995	5.4991	1.8301
Bias	0.0022	-0.0005	3.4991	-0.1699
Bias (%)	2.80%	-0.50%	174.96%	-8.50%
Simple	0.0822	0.0995	3.3362	1.7991
Bias	0.0022	-0.0005	1.3362	-0.2009
Bias (%)	2.81%	-0.46%	66.81%	-10.05%
Rotations	0.0822	0.0995	3.1342	1.8770
Bias	0.0022	-0.0005	1.1342	-0.1230
Bias (%)	2.72%	-0.46%	56.71%	-6.15%
Influenced growth	0.0821	0.0996	3.6874	1.8697
Bias	0.0021	-0.0004	1.6874	-0.1303
Bias (%)	2.64%	-0.44%	84.37%	-6.54%

Table 3: Results obtained for the edge corrected estimation of the test set parameters after the first iteration. We have used $N = 3$ (simulated surroundings/rotations) and $N_{max} = 50$ (stopping criterion for the minimisation of $S(\theta)$).

5.3 Fitting the model to the Scots pines

In Section 4 we introduced our data set, a stand of Swedish Scots pines measured at three time points. In Table 4 we give the estimates found after having run the non-edge corrected estimation procedure, together with the results obtained in the three edge corrected estimation procedures. Just as for the test set, in the uncorrected estimation we have run the estimation until no change in $S(\theta)$ has been observed for $N_{max} = 3000$ consecutive runs whereas in all the corrected ones we have used $N_{max} = 50$. In the edge corrections we have chosen W to be a square region with side length 25 and for each edge correction method we have considered 10 different estimation runs where each of these uses the last $M = 4$ iterations to create its final estimates and in each iteration we have considered $N = 3$ simulated surroundings/rotations. However, here we have chosen the less restrictive value 2 for ϵ. In the uncorrected estimation we found $\hat{\alpha} = 0.004148$ and $\hat{\mu} = 0$ and these will be taken as final estimates for α and μ.

	λ	K	c	r
Initial	0.0350	0.0860	0.0195	8.0
Uncorrected	0.0790	0.0943	6.3314	3.7325
Simple				
Est. mean	0.0781	0.0949	3.1626	4.0680
Est. s.e.	0.0019	0.0017	1.0327	0.6351
Rotations				
Est. mean	0.0794	0.0944	3.1010	3.9396
Est. s.e.	0.0025	0.0015	0.7992	0.3802
Influenced growth				
Est. mean	0.0778	0.0954	3.5054	3.6151
Est. s.e.	0.0026	0.0016	0.7911	0.7229

Table 4: Parameter estimates found for the Scots pines: Initial estimates, non-edge corrected estimates ($N_{max} = 3000$) and estimates obtained through the different edge corrections. We have run each edge correction 10 times in order to get the estimated mean values and standard errors (s.e.). In each run we have used $N = 3$ (simulated surroundings/rotations), $N_{max} = 50$ (stopping criterion for the minimisation of $S(\theta)$), $\epsilon = 2$ (convergence criterion) and $M = 4$ (number of final iterations).

Note that, as expected, for all three methods, the edge corrected estimates are quite close to the uncorrected ones, except for c. The estimated values of c show that the point patterns are less regular than the uncorrected estimate suggests.

6 Discussion

We have recalled the Renshaw-Särkkä growth-interaction model (RS-model) – a spatio-temporal point process with interacting marks. The death rate of the underlying immigration-death process here depends on each individual's mark size, as opposed to the approach used in [16] where the death rate is constant.

We have then discussed the estimation of the parameters of the model when the process is sampled discretely in time. The parameters which control the marks' growth and interaction, λ, K, c, and r, are estimated using the same least-squares approach as proposed in [16]. Related to the least-squares estimation, we specify how we minimise the sum of squares numerically and discuss some issues related to that. Parallel to this, a new estimator is derived which takes the size changes of the individuals into consideration. Also a new estimator is suggested for the arrival intensity, which compensates for the unobserved arrivals and deaths of individuals arriving and dying between two consecutive sample time points.

We finally propose three edge correction methods for (marked) spatio-temporal point processes which all are based on the idea of placing an approximated expected behaviour of the process at hand outside the study region. We estimate this expected behaviour by simulating realizations of the process, under a parameter choice based on some non-edge corrected initial estimates, and for each such realisation we generate new edge corrected estimates which we average over to get our edge corrected estimates.

We finally fit the RS-model to a data set of Swedish Scots pines. A thorough study of the RS-model's applicability in forestry will be made later. Regarding further developments, note that the RS-model here is presented for a single species. However, it can easily be extended to include the scenario where interaction takes place also between different species, living and interacting within the same study region. This extension is made by letting

each species be governed by, on one hand, its own individual growth function and, on the other hand, its own mark interaction function. Hereby the amount an individual is affected by its neighbours depends, not only on the distance to the neighbours and the sizes of these neighbours, but also on the species of the neighbours. Another interesting extension would be to add a (Brownian) noise in the mark growth function of the RS-model, for example by letting the marks be governed by $dM_i(t) = dm_i(t) + dB_i(t)$, where the $B_i(t)$'s are independent Brownian motions, so that it incorporates uncertainties in the mark sizes. We then hope to find a full likelihood structure for this multivariate diffusion type RS-model.

7 Acknowledgements

The author wishes to thank Aila Särkkä (Chalmers University of Technology) for wonderful supervision, ideas and feedback. The author is also grateful for useful comments and suggestions (mainly related to the edge correction methods) from Claudia Redenbach (University of Ulm), Eric Renshaw (University of Strathclyde) and Gerald van den Boogaart (TU Bergakademie). Gratitude also goes out to Kenneth Nyström (Swedish University of Agricultural Sciences) for useful comments and discussions related to forestry. This research has been supported by the Swedish Research Council. Grants have been received from Chalmers vänner and Chalmerska forskningsfonden.

A Appendix

A.1 Initial estimates for λ, K, c and r

Since K represents the carrying capacity, an upper bound of the marks, it is sensible to use the largest observed mark value as starting value K_0.

Having found K_0 we can find an initial estimate of λ. Since the least interaction among individuals takes place at early time points, i.e. $dm_i(t) \approx f(m_i(t))dt$ for small t, by neglecting the interaction term in (1) one ends up with expression (3). By solving w.r.t. λ in (3), where K_0 replaces K and the largest observed individual at the first sample time point, m_{max}, replaces

$m_i(t)$, we get as initial estimate of λ

$$\lambda_0 = -\frac{1}{T_1} \log \left(\frac{m_i^0(1 - \frac{m_{max}}{K_0})}{m_{max}(1 - \frac{m_i^0}{K_0})} \right).$$

Recall that $m_i^0 > 0$ is the initial size which, if unknown, can be estimated by the smallest size of all individuals observed throughout all time points.

In the case of r and c, however, no obvious choices of initial values are present. What is possible, though, is to construct appropriate bounds for r, $r \in [r_l, r_u]$, which control the optimization and then choose the starting value for r to be, say, $r_0 = (r_u + r_l)/2$. Once this is done we choose our starting value for c to be

$$c_0 = \arg\max_{c \in \mathbb{R}} S\left(\lambda_0, K_0, r_0, c\right).$$

There is a natural lower bound for r, namely $r_l = 1$, since two trees cannot grow inside each other. To determine the upper bound, consider the mark correlation function of a stationary marked point process in $\mathbb{R}^2$ (see [7]), defined as

$$k(r) = \frac{\mathbb{E}_{or}[m_i m_j]}{\mu_m^2} \qquad \text{for } r > 0.$$

Here μ_m is the mean mark of the process and $\mathbb{E}_{or}[m_i m_j]$ denotes the conditional expectation of the mark-product of a pair of (marked) points of the process, given the existence of two such points distance r apart. It is a measure of dependence between the marks of two arbitrary points of the process a distance r apart. If, for some r, $k(r) = 1$ then the marks having inter-point distance r are uncorrelated whereas values of $k(r)$ smaller than 1 indicate inhibition (competition) at distance r and $k(r) > 1$ is a sign of mutual stimulation (points benefit from having inter-point distance r). Figure 2 illustrates idealized shapes of $k(r)$.

Denote by r^* the smallest value of $r > 0$ for which $k(r) = 1$. This is the shortest inter-point distance at which there are indications of uncorrelated marks. In the context of the RS-model, for a fixed time t, r^* indicates where the expected influence zone ends, i.e. $\mathbb{E}[rm_i(t)] \leq r^*$. Consider now a time point at which the marked point pattern generated by the RS-model has stabilised, here taken as the last sample time point available, T_n. We get that

$$r \leq r_u = r^*/\mathbb{E}[m_i(T_n)].$$

We estimate the mean mark at time T_n, $\mathbb{E}[m_i(T_n)]$, by $\bar{m}(T_n)$, the average size of the marks present at time T_n. In the case of our test set (see Section 5.2) the mark correlation plot at $T_3 = 33$ is given by Figure 2. The mean mark size for $T_3 = 33$ is given by $\bar{m}(T_3) = 0.0743$ and, as can be seen in Figure 2, $r^* \approx 1/3$, implying that $r_u = r^*/\bar{m}(T_3) \approx 4.5$ thus leading to $r_0 = 2.75$.

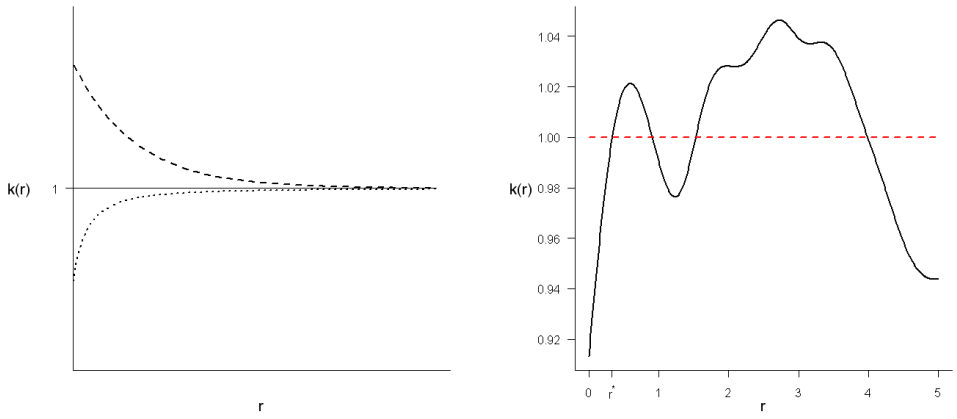

Figure 2: **Left**: Idealized shapes of different mark correlation functions. Mutual stimulation (dashed), uncorrelated marks (solid) and inhibition (dotted). **Right**: Mark correlation plot of our test set (see Section 5.2) at time $T_3 = 33$ ($\lambda = 0.08$, $K = 0.1$, $c = 2$, $r = 2$, $\alpha = 0.007$, $\mu = 0.02$) where $r^* = 1/3$.

A.2 Choosing step-lengths

Another issue of importance here are the step-lengths δ_λ, δ_K, δ_c, δ_r. The simplest way of choosing δ_λ, δ_K, and δ_r is to choose $\delta_\lambda = \lambda_0$, $\delta_K = K_0$, and $\delta_r = r_0 - r_l$ since this way we allow for the estimates of these parameters to reach their minimum values. However, since $c \in \mathbb{R}$, choosing $\delta_c = c_0$ is not in any way self-evident. Choosing a too small δ_c would be more or less equivalent to keeping it fixed which certainly is not desirable. Although letting δ_c be too big may result in slower convergence, trials have shown that it does not affect the convergence of the estimation as much as keeping it too

small. Since c and r do not have as natural choices of initial estimates as λ and K do and because of the strong dependence between them, new starting values for c and r can be found by starting the minimization, keeping $\lambda = \lambda_0$ and $K = K_0$ fixed, and then run the procedure a few times (say $N_{max} = 50$) with δ_c chosen big. This generates new estimates of c and r which in turn can be used as new starting values, c_0 and r_0, and we can then choose δ_c to be this new c_0, which we keep throughout the remaining estimation procedure (including the edge correction parts).

A.3 The estimator for α

When constructing our α-estimator we wish to somehow compensate for the unobserved individuals who arrive and die during the same interval (T_{j-1}, T_j), $j = 1, \ldots, n$.

For each $j = 1, \ldots, n$, let N_{T_j} be the number of individuals observed at sample times up until T_j, i.e. $N_{T_j} = \left|\cup_{i=1}^{j} \Omega_{T_i}\right|$, where Ω_t consists of the indices of the individuals alive at t and $|A|$ denotes the cardinality of a set A. Further, let $B(t) \geq 0$ denote the number of arrivals to W by time t. Instead of considering $\Delta B(T_{j-1}) = \Delta N_{T_{j-1}}$, where $\Delta B(T_{j-1}) = B(T_j) - B(T_{j-1})$ and $\Delta N_{T_{j-1}} = N_{T_j} - N_{T_{j-1}}$, and let our likelihood be based on these independent $Poi\left(\alpha(T_j - T_{j-1})\right)$-distributed increments, as was done in [16], we here consider

$$\begin{aligned} \Delta B(T_{j-1}) &= \Delta N_{T_{j-1}} \\ &+ \underbrace{\mathbb{E}\left[\sum_{k=1}^{\Delta B(T_{j-1})} \mathbf{1}\left\{\text{Individual } k \text{ dies in } (T_{j-1}, T_j)\right\}\right]}_{I}, \end{aligned} \tag{8}$$

where $\mathbf{1}\{\cdot\}$ is an indicator function. In other words, we add to the observed increments the expected number of individuals arriving and dying during (T_{j-1}, T_j).

Let $\eta_k^{\Delta T_{j-1}}$ denote the lifetime of individual $k \in \{1, \ldots, \Delta B(T_{j-1})\}$ in (8) and recall that m_k^0 is its (deterministic) initial size and $t_i^0 \sim Uni(T_{j-1}, T_j)$ its arrival-time (since the jumps of a Poisson process occurring in a given time interval are uniformly distributed on that interval [9]). By the lack of

memory property of the exponential distribution and by Fubini's theorem the expectation in expression (8) can be written as

$$
\begin{aligned}
I &= \mathbb{E}\left[\sum_{k=1}^{\Delta B(T_{j-1})} \mathbf{1}\left\{T_{j-1} < t_k^0 + \eta_k^{\Delta T_{j-1}} < T_j\right\}\right] \qquad (9)\\
&= \mathbb{E}\left[\mathbb{E}\left[\sum_{k=1}^{\Delta B(T_{j-1})} \mathbf{1}\left\{T_{j-1} < t_k^0 + \eta_k^{\Delta T_{j-1}} < T_j\right\} \middle| \Delta B(T_{j-1})\right]\right]\\
&= \mathbb{E}\left[\sum_{k=1}^{\Delta B(T_{j-1})} \frac{1}{\Delta T_{j-1}} \int_{T_{j-1}}^{T_j} \mathbb{E}\left[\mathbf{1}\left\{T_{j-1} < x_k + \eta_k^{\Delta T_{j-1}} < T_j\right\}\right] dx_k\right]\\
&= \mathbb{E}\left[\sum_{k=1}^{\Delta B(T_{j-1})} \frac{1}{\Delta T_{j-1}} \int_{T_{j-1}}^{T_j} P\left(\eta_k^{\Delta T_{j-1}} < T_j - T_{j-1}\right) dx_k\right]\\
&\approx \mathbb{E}\left[\sum_{k=1}^{\Delta B(T_{j-1})} \left(1 - \mathrm{e}^{-\mu\rho(m_i^0)\Delta T_{j-1}}\right)\right]\\
&= \alpha\nu(W)\Delta T_{j-1}\left(1 - \mathrm{e}^{-\mu\rho(m_i^0)\Delta T_{j-1}}\right).
\end{aligned}
$$

Since the actual μ is unknown we will replace it by its estimate, $\hat{\mu}$, found in expression 6. Furthermore, expression (9) also contains α, the parameter we want to estimate. We deal with this by replacing α by an initial estimate, namely, $\hat{\alpha}_0 = N_{T_n}/(T_n\nu(W))$, given by (7).

In order for expression (8) to be treated as an actual Poisson process increment it needs to be integer valued, hence

$$
\Delta B(T_{j-1}) = \Delta N_{T_{j-1}} + \left\lfloor N_{T_n}\frac{\Delta T_{j-1}}{T_n}\left(1 - \mathrm{e}^{-\hat{\mu}\rho(m_i^0)\Delta T_{j-1}}\right)\right\rfloor, \qquad (10)
$$

where $\lfloor x \rfloor$ denotes the integer part of x. For convenience we will denote the right hand side of (10) by $H(\Delta T_{j-1}, \Delta N_{T_{j-1}}, \hat{\mu}, N_{T_n})$. We end up with the

likelihood function

$$\begin{aligned} L(\alpha) &= \prod_{j=1}^{n} \mathbb{P}\left(\Delta B(T_{j-1}) = H(\Delta T_{j-1}, \Delta N_{T_{j-1}}, \hat{\mu}, N_{T_n})\right) \\ &= \prod_{j=1}^{n} \frac{\mathrm{e}^{-\alpha\nu(W)\Delta T_{j-1}}(\alpha\nu(W)\Delta T_{j-1})^{H(\Delta T_{j-1}, \Delta N_{T_{j-1}}, \hat{\mu}, N_{T_n})}}{H(\Delta T_{j-1}, \Delta N_{T_{j-1}}, \hat{\mu}, N_{T_n})!} \end{aligned}$$

and by evaluating $\mathrm{d}\log(L(\alpha))/\mathrm{d}\alpha = 0$ we finally arrive at the estimator

$$\hat{\alpha} = \underbrace{\frac{N_{T_n}}{T_n\nu(W)}}_{=\hat{\alpha}_0} + \frac{1}{T_n\nu(W)} \sum_{j=1}^{n} \left\lfloor N_{T_n} \frac{\Delta T_{j-1}}{T_n} \left(1 - \mathrm{e}^{-\hat{\mu}\rho(m_i^0)\Delta T_{j-1}}\right) \right\rfloor . \quad (11)$$

Since $\hat{\mu} > 0$, $\Delta T_{j-1} > 0$ and $\rho(x) > 0$, for all $x > 0$, and since $f(x) = 1 - \mathrm{e}^{-x}$ is strictly increasing and bounded below by 0 and above by 1, for $x > 0$, it is clear that $\hat{\alpha}$ is increasing with $\hat{\mu}$ and

$$\hat{\alpha}_0 = \lim_{\hat{\mu}\to 0} \hat{\alpha}|_{\hat{\mu}} < \hat{\alpha} < \lim_{\hat{\mu}\to\infty} \hat{\alpha}|_{\hat{\mu}} = \hat{\alpha}_0 + \frac{1}{T_n\nu(W)} \sum_{j=1}^{n} \left\lfloor N_{T_n} \frac{\Delta T_{j-1}}{T_n} \right\rfloor .$$

For a random variable $Z = X + Y$ it holds that $\mathrm{Var}(Z) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\,\mathrm{Cov}(X, Y)$. Let now $X = \hat{\alpha}_0$ and let Y be the sum in expression (11). Since X and Y are positively correlated (both contain N_{T_n}) and since $\mathrm{Var}(Y) \geq 0$ it is clear that $\mathrm{Var}(\hat{\alpha}) > \mathrm{Var}(\hat{\alpha}_0)$ for all $\hat{\mu} > 0$. This implies that the trade off for using $\hat{\alpha}$ instead of $\hat{\alpha}_0$ is a higher standard error. Furthermore, as $\hat{\alpha}$ is increasing with $\hat{\mu}$, so is $\mathrm{Var}(\hat{\alpha})$.

Table 5 gives us the estimated means and standard errors (s.e.) of $\hat{\alpha}$ (and $\hat{\alpha}_0$) for a few values of $\hat{\mu}$, based on 30 simulated realisations from the same parameters as the test set (recall that $\alpha = 0.007$; see Section 5.2).

In estimations of μ based on simulated realisations it has been observed that there seems to be no indication of over-estimation of μ. As one can see in Table 5, on average $\hat{\alpha}_0$ under-estimates α more than $\hat{\alpha}$ does when $\hat{\mu} \leq \mu$, in the above scenario indicating that $\hat{\alpha}$ is preferred to $\hat{\alpha}_0$. Note also the smaller standard error of $\hat{\alpha}_0$.

$\alpha = 0.007$	Est. mean	Est. s.e.	Est. bias (%)
$\hat{\alpha}_0 = \lim_{\hat{\mu}\to 0} \hat{\alpha}\vert_{\hat{\mu}}$	0.0060	0.0008	-0.00099 (-14%)
$\hat{\alpha}$ ($\hat{\mu} = 0.0002$)	0.0060	0.0008	-0.00099 (-14%)
$\hat{\alpha}$ ($\hat{\mu} = 0.002$)	0.0061	0.0009	-0.00089 (-13%)
$\hat{\alpha}$ ($\hat{\mu} = 0.02$)	0.0074	0.0011	0.00044 (6%)
$\hat{\alpha}$ ($\hat{\mu} = 0.1$)	0.0102	0.0014	0.00320 (46%)
$\hat{\alpha}$ ($\hat{\mu} = 0.2$)	0.0111	0.0016	0.00411 (59%)
$\hat{\alpha}$ ($\hat{\mu} = 5$)	0.0119	0.0017	0.00489 (70%)
$\lim_{\hat{\mu}\to\infty} \hat{\alpha}\vert_{\hat{\mu}}$	0.0119	0.0017	0.00489 (70%)

Table 5: Estimated means, standard errors (s.e.), and biases of $\hat{\alpha}$ (and $\hat{\alpha}_0$), based on 30 simulated realisations from the same parameters as the test set (see Section 5.2; recall that $\alpha = 0.007$, $\mu = 0.02$, $\lambda = 0.08$, $K = 0.1$, $c = 2$, and $r = 2$).

References

[1] Barndorff-Nielsen, O.E., Kendall, W.S., van Lieshout, M.N.M. *Stochastic Geometry: Likelihood and Computation*. Monographs on Statistics and Applied Probability 80. Chapman & Hall/CRC, 1999.

[2] Basawa, I.V., Prakasa Rao, B.L.S. *Statistical Inference for Stochastic Processes*. Academic press, 1980.

[3] Berger, U., Hildenbrandt, H. A new approach to spatially explicit modelling of forest dynamics: spacing, ageing and neighbourhood competition of mangrove trees. *Ecological Modelling*, 132:287–302, 2000.

[4] P. Diggle. *Statistical Analysis of Spatial Point Patterns*. Oxford university press, second edition, 2001.

[5] Goreaud, F., Loreau, M., Millier, C. Spatial structure and the survival of an inferior competitor: a theoretical model of neighborhood competition in plants. *Ecological Modelling*, 158:1–19, 2002.

[6] Grabarnik, P., Särkkä, A. Modelling the spatial structure of forest stands by multivariate point processes with hierarchical interactions. *Ecological Modelling*, 220:1232–1240, 2009.

[7] Illian J., Penttinen A., Stoyan H., Stoyan D. *Statistical Analysis and Modelling of Spatial Point Patterns*. Wiley-Interscience, 2008.

[8] Nord-Larsen, T. Modeling individual-tree growth from data with highly irregular measurement intervals. *Forest Science*, 52:198–208, 2006.

[9] J.R. Norris. *Markov Chains*. Cambridge series in statistical and probabilistic mathematics, 1997.

[10] Prévosto, B., Curt, T., Gueugnot, J., Coquillard, P. Modeling mid-elevation Scots pine growth on a volcanic substrate. *Forest Ecology and Management*, 131:223–237, 2000.

[11] E. Renshaw. The linear spatial-temporal interaction process and its relation to $1/\omega$-noise. *J. Roy. Statist. Soc. B*, 56:75–91, 1994.

[12] E. Renshaw. Applying the saddlepoint approximation to bivariate stochastic processes. *Mathematical Biosciences*, 168:57–75, 2000.

[13] Renshaw, E., Comas, C. Space-time generation of high intensity patterns using growth-interaction processes. *Statistics and Computing*, 19:423–437, 2009.

[14] Renshaw, E., Comas, C., Mateu, J. Analysis of forest thinning strategies through the development of space-time growth-interaction simulation models. *Stochastic Environmental Research and Risk Assessment*, 23:275–288, 2009.

[15] Renshaw, E., Särkkä, A. Gibbs point processes for studying the development of spatial-temporal stochastic processes. *Computational Statistics & Data Analysis*, 36:85–105, 2001.

[16] Särkkä, A., Renshaw, E. The analysis of marked point patterns evolving through space and time. *Computational Statistics & Data Analysis*, 51:1698–1718, 2006.

[17] Stoyan, D., Kendall, W., Mecke, J. *Stochastic Geometry and its Applications*. John Wiley & sons, second edition, 1995.

[18] Stoyan, D., Penttinen, A. Recent applications of point process methods in forestry statistics. *Statistical Science*, 15:61–78, 2000.

[19] van der Vaart, A.W. *Asymptotic Statistics*. Cambridge series in statistical and probabilistic mathematics, 1998.

[20] Weiner, J., Damgaard, C. Size-asymmetric competition and size-asymmetric growth in a spatially explicit zone-of-influence model of plant competition. *Ecological Research*, 21:707–712, 2006.

[21] Williams, M.S., Williams, M.T., Mowrer, H.T. A boundary reconstruction method for circular fixed-area plots in environmental survey. *Journal of agricultural, biological, and environmental statistics*, 6:479–494, 2001.