



UNIVERSITY OF GOTHENBURG

This is an author produced version of a paper published in **Computational Statistics & Data Analysis**

This paper does not include the final publisher proof-corrections or journal pagination.

Citation for the published paper:

Redenbach, C. and Särkkä, A

Parameter estimation for growth interaction processes using spatio-temporal information

Computational Statistics & Data Analysis, 57 (1) s. 672-683

<http://dx.doi.org/10.1016/j.csda.2012.08.006>

Access to the published version may require subscription. Published with permission from: **Elsevier**

GUP

Gothenburg University Publications

<http://gup.ub.gu.se/gup/>

Parameter estimation for growth interaction processes using spatio-temporal information

Claudia Redenbach

*University of Kaiserslautern, Mathematics Department, 67663 Kaiserslautern, Germany,
redenbach@mathematik.uni-kl.de*

Aila Särkkä

*Mathematical Sciences, Chalmers University of Technology and Mathematical Sciences,
University of Gothenburg, 412 96 Göteborg, Sweden, aila@chalmers.se*

Abstract

Methods for the parameter estimation for a spatio-temporal marked point process model, the so-called growth-interaction model, are investigated. Least squares estimation methods for this model found in the literature are only concerned with fitting the mark distribution observed in the data. These methods are unable to distinguish between models which have the same birth, death, interaction and growth functions and parameters but different arrival strategies for the points. Hence, they are extended such that also the spatial structure of a point pattern is taken into account. The suggested methods are evaluated in a simulation study and applied to a small data set from forestry.

Keywords: L-function, least squares estimation, logistic power-law function, parameter estimation, Scots pines, spatio-temporal marked point process

1. Introduction

During the last 20 years the development of techniques such as Global Positioning Systems has facilitated the record of spatio-temporal data. The availability of these data has raised an interest in spatio-temporal modelling. As a consequence, substantial progress has been made in the development of analysis methods involving geostatistical, hierarchical and multivariate time series approaches, together with the implementation of space-time dynamic models and point process models (see, e.g., Cressie and Wikle (2011), Diggle and Gabriel (2010)). Here, we concentrate on the so-called growth-interaction process, which is a spatial marked point process evolving in time (see Renshaw and Särkkä, 2001). New immigrants (points) arrive randomly in time according to a Poisson process, have uniformly distributed locations on the study area, and are

*Corresponding author. Tel.: +49-631-205 3620, Fax: +46-631-2748

assigned some initial marks. In the successive small time intervals, each individual either dies according to a simple death process, or changes its size deterministically. The growth term includes an individual growth function and an interaction function, which depends on the locations and sizes of the neighbouring points. The parameter estimation for the growth-interaction process is discussed in several publications. Särkkä and Renshaw (2006) suggest to estimate the immigration and death rates by the maximum likelihood method, and the growth and interaction parameters by a least squares approach. Cronie (2010) suggests an improved estimator for the immigration rate, and Cronie and Yu (2010) an improved maximum likelihood approach for the immigration and death rates. Furthermore, Cronie and Särkkä (2011) add an edge correction to the least squares estimation procedure of the growth and interaction parameters. A separate estimation of one of the growth parameters, namely the carrying capacity (maximal size), is suggested in Cronie *et al.* (2011) when fitting the GI model to data from a young forest.

Besides the choice of the growth and interaction functions, also the arrival strategy for new immigrants may influence the spatial arrangement of the points. In the original growth-interaction process the new immigrants are located uniformly and independently of each other in the study region. Given that the locations of mature trees often form a regular pattern, uniform and independent locations may not give the spatial structure that is observed.

Cronie *et al.* (2011) suggest placing new trees according to a hardcore strategy such that the center of a newly arrived tree is not inside another tree. Comas (2009) studies two different forest regeneration strategies through the development of a growth-interaction process. Instead of uniform locations for the new arrivals, he suggests that each tree produces seedlings located according to a Gaussian distribution around the tree. His main purpose is to compare the two regeneration strategies and to study how they affect the size distribution of the trees and timber production. The spatial pattern of the tree locations is not of interest. Here, we will study the effect of different arrival strategies on the spatial structure of a point pattern.

The least squares estimation of the growth and interaction parameters is based on minimizing the difference between observed and predicted marks of the points, while the spatial information is not taken into account. Consequently, different arrival strategies cannot be detected by this method. Furthermore, the analysis of a small forest data set in Cronie and Särkkä (2011) shows that the spatial structures of the real data and of point patterns simulated from the fitted model do not match. The pattern of tree locations in the real data seems to be more regular than in the simulated data. This may be due to the uniformly located new arrivals in the simulated data. To be able to distinguish between two models with the same birth, death, interaction and growth functions and parameters but different arrival strategies of new points, it may not be enough to base the estimation only on marks, but the spatial structure needs to be included as well.

In this paper we suggest how to estimate the parameters based on both mark information and information on the point locations. This makes the parameter

estimation capable of dealing with alternative arrival strategies. In particular, the case of hardcore arrivals is investigated which is considered more suitable for the Scots pine data studied in Cronie and Särkkä (2011) than uniform arrivals.

The paper is organized as follows. The growth-interaction process is introduced in Section ???. In Section ??? we recall the estimation approaches presented in Särkkä and Renshaw (2006) and Cronie (2010), and refine the existing estimation methods by adding spatial information to the procedure. The influence of the arrival strategy and the new estimation approach are investigated by a simulation study in Section ???. Finally, in Section ??? we fit a growth-interaction model to forestry data.

2. Growth-interaction process

2.1. Description of the process

The growth-interaction (GI) process was introduced by Renshaw and Särkkä (2001) as a model for tree growth. Trees are modelled by marked points in a given window W . The marks are positive real numbers which can be interpreted as sizes.

The GI model starts with an initial, and possibly empty, marked point pattern. New immigrants arrive according to a Poisson process, with rate α , at uniformly distributed locations x_i . They are assigned initial marks m_i^0 which are either fixed or follow some distribution. In small time intervals $(t, t + dt)$, each individual either dies naturally with probability μdt , or undergoes the deterministic size change

$$m_i(t + dt) = m_i(t) + f(m_i(t))dt + \sum_{j \neq i} h(m_i(t), m_j(t), \|x_i - x_j\|)dt, \quad (1)$$

where $f(\cdot)$ is an individual growth function, $h(\cdot)$ a spatial interaction function and $\|x_i - x_j\|$ the distance between the locations x_i and x_j . A simulation algorithm for this model is given in Särkkä and Renshaw (2006). Note that in (??) h is a pairwise interaction function since we consider situations, where there is competition between the individuals. However, also higher order interactions, i.e. interactions between more than two points, can be included in the interaction function.

The growth function and the interaction function can be chosen freely. Here, we study the family of so-called logistic power-law functions

$$f(m_i(t)) = c_1 m_i(t) - c_2 (m_i(t))^{p+1},$$

where c_1 , c_2 and p are given parameters. As special cases we get the logistic growth function ($c_1 = \lambda$, $c_2 = \lambda/K$ and $p = 1$)

$$f(m_i(t)) = \lambda m_i(t) (1 - m_i(t)/K),$$

and the linear growth function ($c_1 = -\lambda/K$, $c_2 = -\lambda$ and $p = -1$)

$$f(m_i(t)) = \lambda (1 - m_i(t)/K).$$

In these two special cases, the parameter λ is the intrinsic growth rate and K is the carrying capacity, i.e. an upper bound for the value of the marks.

There is also a lot of freedom when choosing the interaction function. Here, we will give some examples of pairwise interaction functions. Our first example is the symmetric function

$$h(m_i(t), m_j(t), \|x_i - x_j\|) = -bI(\|x_i - x_j\| < r(m_i(t) + m_j(t))), \quad (2)$$

where $b > 0$ is an interaction parameter and $I(\cdot)$ denotes the indicator function. We assume that the area over which a point interacts with other points can be represented by a disk of radius $rm_i(t)$, where $r > 0$ is the range of interaction. As soon as two disks overlap, i.e., $\|x_i - x_j\| < r(m_i(t) + m_j(t))$ then competitive interaction takes place with force b . Note that this function is symmetric, i.e. points with the larger and smaller mark in an interactive pair are affected equally. An asymmetric interaction function of a similar type is obtained via

$$h(m_i(t), m_j(t), \|x_i - x_j\|) = -bm_j(t)I(\|x_i - x_j\| < r(m_i(t) + m_j(t))).$$

In this case, interaction force is given by a certain percentage of the current mark size of the neighbouring point.

Another non-symmetric interaction function is obtained if the interaction is assumed to be proportional to the relative size of the influence zone (Särkkä and Renshaw, 2006). Let $D(x_i, s)$ denote the disk with centre x_i and radius s . If the so-called area-interaction function is given by

$$h(m_i(t), m_j(t), \|x_i - x_j\|) = -b \text{ area}\{D(x_i, rm_i(t)) \cap D(x_j, rm_j(t))\} / \pi r^2 m_i^2(t), \quad (3)$$

we see that the smaller of two interacting points will be affected more than the larger one. Points of equal size are affected equally. This function is of particular interest in forestry applications.

In the forestry setting, one might wonder whether the choice of uniform locations for immigrants is a good one. Different tree species have different tolerance to shade which defines whether the seedlings can survive under the crown of other trees. Shade tolerant species require the shadow of adult trees to survive, whilst shade intolerant species need to have more space. For shade tolerant species, uniformly located or even clustered arrivals seem reasonable. However, for shade intolerant species, like Scots pine, some other arrival approach could be more appropriate. Therefore, we consider an alternative arrival strategy, namely that immigrants are only placed in locations where they do not suffer interaction with older individuals right away. That means that new immigrants are placed according to a simple-sequential inhibition (SSI) procedure. Figures ?? and ?? show that the spatial structure of the resulting pattern is somewhat dependent on the chosen arrival strategy.

2.2. Estimation of parameters

Estimation of the parameters of the GI process based on observations at n different time points is discussed in Särkkä and Renshaw (2006). The immigration rate α and the death rate μ are estimated by the maximum likelihood

(ML) method. For the growth and interaction parameters a least squares (LS) method is applied. For the LS estimation, Särkkä and Renshaw (2006) suggests the following approach. Each of the parameters is given a grid of values. Starting with the data at the first observed time point, the sum of the squared differences between the observed ($m_i(t)$) and predicted ($\tilde{m}_i(t)$) marks

$$S_M = \sum_{t=2}^n \sum_{i \in \Omega_t} (\tilde{m}_i(t) - m_i(t))^2 \quad (4)$$

is computed for each combination of the parameter values. Here Ω_t denotes the marked point pattern at time t , and n is the number of time points where data have been recorded. The values of the parameters that minimize S_M are chosen as the estimates. One can continue with a finer grid close to these “initial” estimates in order to obtain better estimates.

In Cronie (2010), a similar procedure as the one in Särkkä and Renshaw (2006) is used. However, the growth and interaction parameters are estimated first by the LS method. In the second step, these estimated values are used to estimate the immigration rate α and the death rate μ using the ML method. Furthermore, the grid search is replaced by randomised parameter updates. Recommendations for starting values of the parameters and for the step sizes are given as well.

When estimating α and μ , Särkkä and Renshaw (2006) do not take into account that points can arrive and die within the same observation interval without being observed. Consequently, both parameters are underestimated. Cronie (2010) improves the estimation of α and μ by taking the unobserved points into account. While Särkkä and Renshaw (2006) consider the probability of a natural death to be equal for all individuals, Cronie (2010), Comas (2009), and Cronie and Särkkä (2011) allow it to depend on the mark size via a function ρ . The improved estimators are as follows:

Let $T_j, j = 1, \dots, n$, denote the sampling times and let N_{T_j} be the number of individuals observed until time T_j . Then set $\Delta T_{j-1} = T_j - T_{j-1}$. Furthermore, let t_i^0 denote the birth time of the i -th individual, $T_{j,i}$ the last sampling time at which individual i was alive, and $\tilde{m}_i(T_{j,i})$ the prediction of its mark at time $T_{j,i}$. Finally, denote the number of individuals who have died before time T by n_T and let m_T be the number of individuals who are still alive at time T . The estimators for μ and α from Cronie (2010) are then given by

$$\hat{\mu} = n_T \left/ \left(\sum_{i=1}^{n_T} \rho(\tilde{m}_i(T_{j,i})) (T_{j,i} - t_i^0) + \sum_{l=1}^{m_T} \rho(m_l(T)) (T - t_l^0) \right) \right. \quad (5)$$

and

$$\hat{\alpha} = \underbrace{\frac{N_{T_n}}{T_n \nu(W)}}_{=\hat{\alpha}_0} + \frac{1}{T_n \nu(W)} \sum_{j=1}^n \left[N_{T_n} \frac{\Delta T_{j-1}}{T_n} \left(1 - e^{-\hat{\mu} \rho(m_i^0) \Delta T_{j-1}} \right) \right]. \quad (6)$$

Here, $\hat{\mu}$ reduces to the estimator from Särkkä and Renshaw (2006) for $\rho \equiv 1$ and $\hat{\alpha}_0$ is the estimator proposed by Särkkä and Renshaw (2006). In practice, the birth times t_i^0 are unknown. Therefore, they have to be simulated (see Cronie and Särkkä, 2011).

The estimator $\hat{\mu}$ includes predicted marks, so the growth and interaction parameters have to be estimated prior to estimating μ (and α). On the other hand, knowledge of μ and α is not required for the estimation of the growth and interaction parameters. In cases where the starting time T_0 is unknown, T_n and N_{T_n} should be changed to $T_n - T_1$ and $N_{T_n} - N_{T_1}$, respectively. One reason is that the interval $[T_0, T_1]$ is typically longer than the others which makes it more unreliable since it may contain many points that are never observed.

3. New estimation approach

Direct information on the spatial structure of the point locations is not included in the above estimation procedure. Hence, the influence of the arrival strategy is not taken into account when estimating the model parameters. Our intention is to introduce an estimation procedure which also takes spatial structure into account. Instead of minimizing S_M which only depends on marks, one can minimize

$$S = w_1 S_1 + w_2 S_2 + \dots + w_k S_k, \quad (7)$$

where S_1, S_2, \dots, S_k are functions depending on marks, locations or on both. Furthermore, $w_j, j = 1, \dots, k$, are non-negative weights such that $\sum_{j=1}^k w_j = 1$. The estimates for the parameters of the growth and interaction functions are found by minimizing S with respect to the parameters.

3.1. Estimation algorithm

Contrary to the established estimation approach, we also use summary statistics based on the point locations. In this case the evaluation of (??) requires the simulation of the complete pattern, hence knowledge of all parameters including μ and α . Consequently, a separate estimation of the growth and interaction parameters and of μ and α is no longer possible. Therefore, we suggest the following estimation procedure which we demonstrate with the logistic power law growth function and an interaction function with two parameters, the strength of interaction b and the range of interaction r :

1. Choose a set (c_1, c_2, p, b, r) of growth and interaction parameters.
2. Estimate μ and α using (??) and (??) where the predicted marks are computed using the given parameters (c_1, c_2, p, b, r) .
3. Simulate the GI process for the given time period using the parameters (c_1, c_2, p, b, r) and the estimated α and μ . Evaluate the distance (??) for each sampling time point.
4. Store the best parameter values.

5. Update the vector (c_1, c_2, p, b, r) using a given optimisation scheme and return to step 2.

This algorithm is repeated until either S is smaller than a prescribed bound or S does not change during a given number of estimation steps.

3.2. Definition of distance measure

An essential part of our new estimation procedure is the choice of the distance measure S . Here, we have chosen to use $S = 0.5(S_M + S_L)$, where S_M measures the difference between the mark distribution in the data and the model while S_L measures the difference in the spatial arrangement of the points using the L -function. The L -function is a variant of Ripley's K -function whose value at $s > 0$ is proportional to the expected number of further points within distance s of a typical point of the process. In \mathbb{R}^2 , the L -function is defined as $L(s) = \sqrt{\widehat{K}(s)/\pi}$. For the estimation of L we use the estimator $\hat{L}(s) = \sqrt{\hat{K}(s)/\pi}$ where $\hat{K}(s)$ is the standard estimator for K using translation edge correction. We have chosen the cumulative L -function to describe the spatial structure since it yields more stable information than density functions such as the pair correlation function. Therefore, cumulative functions are often recommended for model fitting purposes although density functions are more informative for visual interpretation (Illian *et al.*, 2008). Furthermore, estimation of density functions is not as standardized as estimation of cumulative ones.

In experiments, it turned out that it is important to scale both contributions S_M and S_L suitably such that neither of them dominates the overall distance. Furthermore, the range of values of the distance measure should not depend on the "size" of the problem such as the number of observed time steps, the number of points in a pattern or the scale of the marks. Therefore, the sum of squared differences between observed and predicted marks given in (??) is modified to

$$S_M = \frac{1}{n-1} \sum_{t=2}^n \frac{1}{|\Omega_t|} \sum_{i \in \Omega_t} \left(\frac{\tilde{m}_i(t) - m_i(t)}{\max(m_i)} \right)^2, \quad (8)$$

where $\max(m_i)$ is the largest observed mark. The distance of the L -functions is measured by

$$S_L = \frac{1}{(n-1)M} \sum_{t=2}^n \sum_{i=1}^M \left(\frac{\tilde{L}_t(s_i) - L_t(s_i)}{L_t(s_i)} \right)^2, \quad (9)$$

where \tilde{L}_t and L_t are the L -functions of the simulated model and of the data, respectively, in the t -th observed time step. The L -functions are evaluated in M suitably chosen distances s_1, \dots, s_M . Since the simulation of the model is started with the data at the first observed time point, this time step is not included in the computation of the distances. Therefore, (??) gives a mean weighted difference of the predicted and observed marks over all marks at a time point and over all time points. Equation (??) is the mean relative difference of the

predicted and observed L -functions over all distances the L -function has been estimated (at a time point) and over all time points.

Note that in order to compute S_M , the observed marks are compared to the predicted ones which can be computed deterministically when the values of the growth and interaction parameters are given. To compute S_L , we need to estimate the L -function from the observed pattern as well as from a realisation simulated from the fitted model. To do the latter, simulation of the complete process is needed. Note also that the choice of arrival strategy affects the spatial pattern of points, and therefore, also the L function.

4. Simulation study

Based on simulation studies Särkkä and Renshaw (2006) conclude that the parameters λ and K of the logistic and linear growth functions, as well as the range of interaction r , are estimated quite well but the strength of interaction b is severely biased. Adding an edge-correction to the estimation procedure improves the estimate of b , see Cronie and Särkkä (2011). However, when GI model was fitted to some forestry data, the spatial structure of the locations in the data did not match the structure in the patterns simulated from the fitted model. This insufficient fit of the spatial structure may be caused by the fact that a wrong arrival strategy of new trees was used, and only differences between the observed and predicted marks, not spatial structure, were considered. In our simulation study, the behaviour of the distance measure $S = \frac{1}{2}(S_M + S_L)$, where S_M and S_L are as in (??) and (??) respectively, is evaluated. In particular, the influence of the chosen arrival strategy on S_L is investigated.

For the simulation study, we have chosen linear and logistic growth functions with the parameters $\lambda = 1$, $K = 10$, $\alpha = 5.0$, and $\mu = 0.02$. For the interaction function we have chosen the area interaction function (??) with $r = 0.01$. The value of b was chosen as 0.1, 1.5, and 7.0. Both uniform and hardcore (SSI) arrivals were used. In the hardcore case, new points may only be placed in locations where they do not experience interaction with any of the existing points. The models were simulated starting with an empty window $W = [0, 1]^2$ (which is wrapped onto a torus) at $t = 0$, and the initial mark size is $m_i^0 = 0.01$. See Figures ?? and ?? for realisations of these models. The data set used for the estimation consisted of the model realisations at $n = 10$ time points from $t = 10$ to 100 in steps of 10. For each combination of the model parameters, 100 realisations of the model were generated. Here, the L -function was evaluated in $M = 16$ points chosen equidistantly on the interval $[0, 0.15]$.

It can be seen in Figures ?? and ?? that the number of points decreases and the point pattern becomes more regular as the interaction parameter b increases. Under logistic growth there are fewer points that survive but they are larger than under linear growth. The same is true for the different arrival strategies: Under uniform arrival less points survive than under hardcore arrival. In particular, the number of small points is reduced.

The L -functions of the point patterns shown in Figure ?? indicate that with medium ($b = 1.5$) or strong ($b = 7$) interaction, both uniform and hardcore

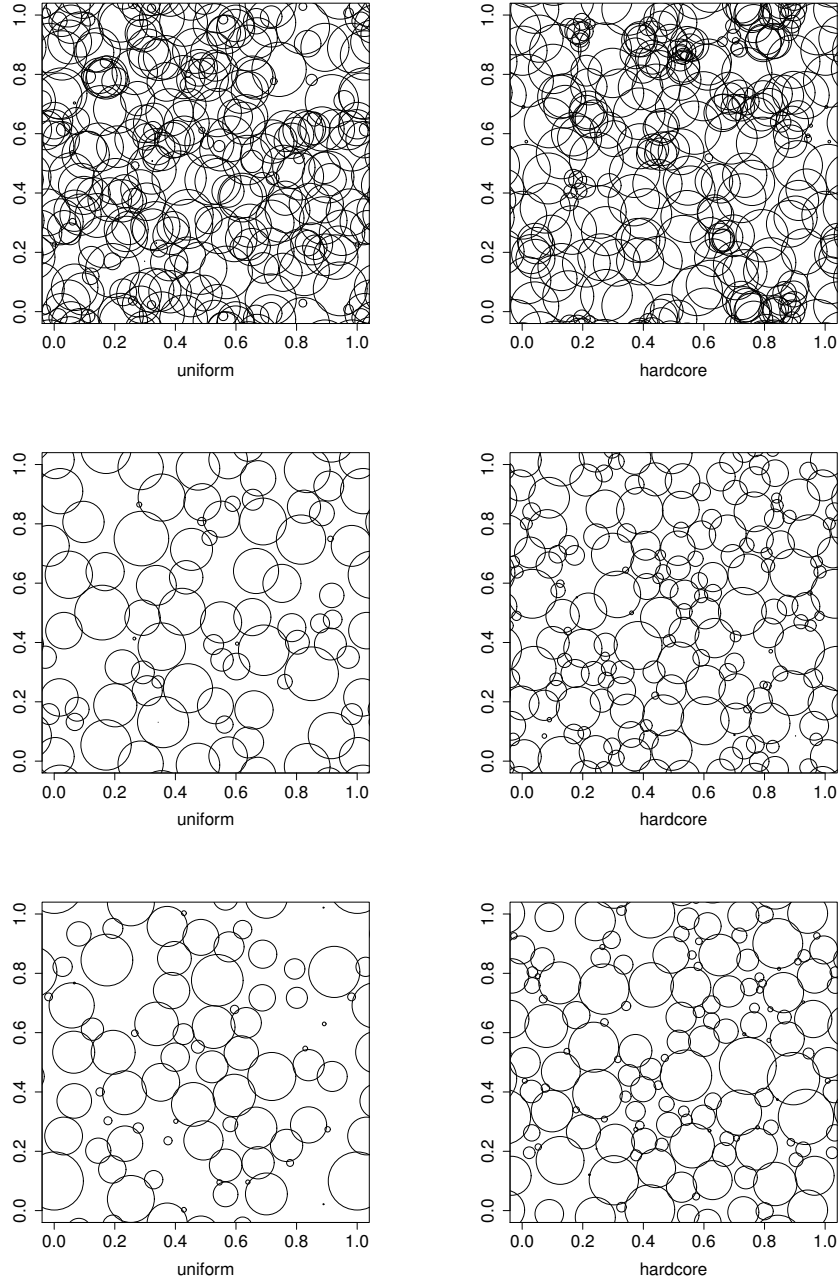


Figure 1: Realisations of the GI process with linear growth and uniform (left) and hardcore (right) arrivals for low ($b = 0.1$, top), intermediate ($b = 1.5$, middle), and high ($b = 7.0$, bottom) interaction at time $t = 100$. The remaining parameters are set to $K = 10$, $r = 0.01$, $\alpha = 5.0$, $\mu = 0.02$. Radii plotted are rm_i .

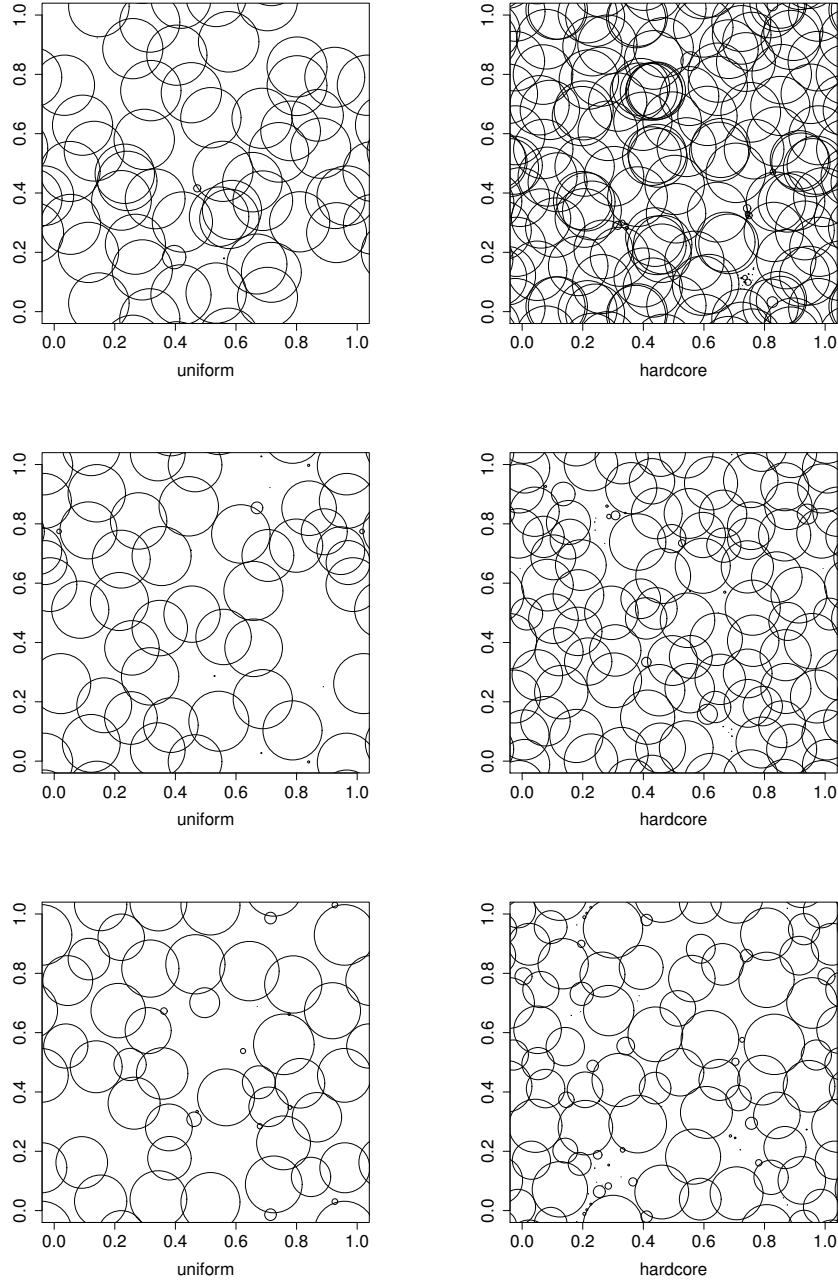


Figure 2: Realisations of the GI process with logistic growth and uniform (left) and hardcore (right) arrivals for low ($b = 0.1$, top), intermediate ($b = 1.5$, middle), and high ($b = 7.0$, bottom) interaction at time $t = 100$. The remaining parameters are set to $K = 10$, $r = 0.01$, $\alpha = 5.0$, $\mu = 0.02$. Radii plotted are rm_i .

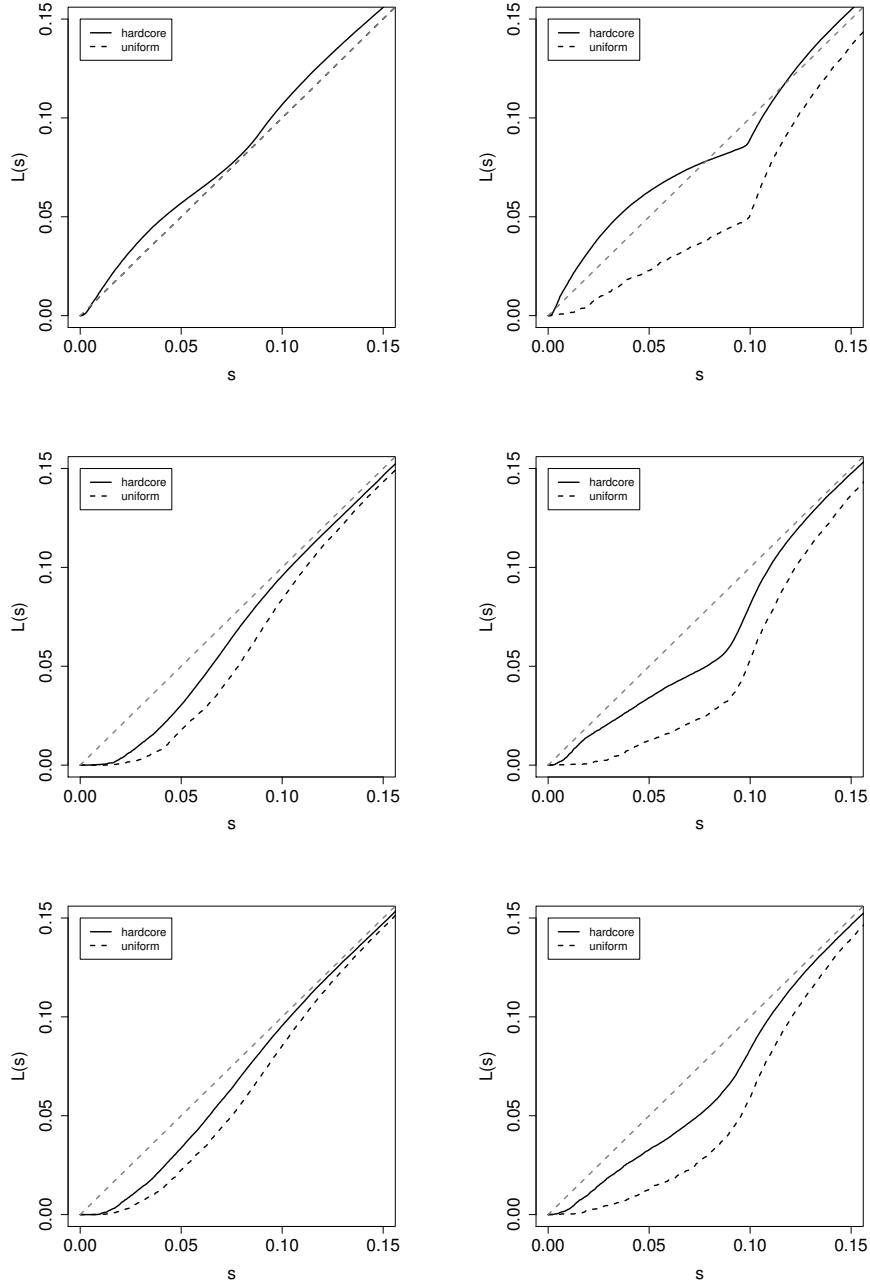


Figure 3: The mean L -functions based on 100 simulations for the GI process with hardcore (solid) and uniform (dashed) arrivals with linear (left) and logistic (right) growth function, and for low ($b = 0.1$, top), intermediate ($b = 1.5$, middle), and high ($b = 7.0$, bottom) interaction at time $t = 100$. The grey curve corresponds to the Poisson process. The remaining parameters are set to $K = 10$, $r = 0.01$, $\alpha = 5.0$, $\mu = 0.02$.

arrivals lead to a clearly regular pattern of points both under linear and logistic growth. Interestingly, uniform arrivals tend to lead to a slightly more regular pattern than hardcore arrivals. This can be due to the fact that under uniform arrivals new points can be located very close to the established points and they will die (due to interaction) almost immediately after their arrival. In contrast, under hardcore arrivals the new points have a greater chance to survive since they do not compete with other points when they arrive. The behaviour under weak interaction ($b = 0.1$) is quite interesting as well. Under linear growth uniform arrivals lead to a Poisson pattern of locations while hardcore arrivals lead to a slightly clustered pattern. However, under logistic growth uniform arrivals cause a clearly regular pattern while hardcore arrivals lead to repulsive pattern at small distances but some clustering at larger distances. This may be due to the fact that under linear growth it is easier for very small points to survive under the shadow of some larger ones than under logistic growth (Särkkä and Renshaw, 2006).

Our main interest here is to investigate how the chosen arrival strategy of new immigrants affects S_L . Table ?? shows the means and standard deviations of the contributions of S_L and S_M to the distance measure using the true parameter values. S_L is estimated using both the correct and the wrong arrival strategy. The values of S_M only require predicted marks and are not affected by the arrival strategy.

growth function	arrival strategy	b	$S_L/2$ correct		$S_L/2$ wrong		$S_M/2$	
			mean	sd	mean	sd	mean	sd
linear	uniform	0.1	0.0085	0.0049	0.0302	0.0153	0.0562	0.0037
linear	uniform	1.5	0.0632	0.0177	0.0348	0.0137	0.0372	0.0030
linear	uniform	7.0	0.0743	0.0186	0.0441	0.0168	0.0326	0.0027
linear	hardcore	0.1	0.0181	0.0107	0.0359	0.0196	0.0534	0.0030
linear	hardcore	1.5	0.0257	0.0087	0.0688	0.0145	0.0225	0.0017
linear	hardcore	7.0	0.0326	0.0097	0.0703	0.0134	0.0207	0.0022
logistic	uniform	0.1	0.0929	0.0345	0.0457	0.0257	0.1816	0.0137
logistic	uniform	1.5	0.1025	0.0330	0.0575	0.0231	0.0943	0.0077
logistic	uniform	7.0	0.1288	0.0234	0.0726	0.0228	0.0604	0.0077
logistic	hardcore	0.1	0.0267	0.0104	0.1344	0.0303	0.1339	0.0106
logistic	hardcore	1.5	0.0626	0.0167	0.1508	0.0281	0.0640	0.0058
logistic	hardcore	7.0	0.0598	0.0168	0.1404	0.0258	0.0365	0.0046

Table 1: Means and standard deviations of $S_L/2$ and $S_M/2$ obtained from estimation on 100 realisations of the model. Estimation of S_L under both the correct (as written in the second column) and the wrong (hardcore if uniform in the second column and vice versa) arrival strategy is considered.

It would be expected that S_L would have a smaller value under the correct arrival strategy compared to the incorrect strategy. When the hardcore arrivals are used, Table ?? shows that the difference measure S_L is indeed smaller under

the hardcore arrivals than under the uniform arrivals indicating that the former would be the correct strategy. However, this does not seem to be the case when the true arrival strategy is uniform. Only under linear growth and for small values of the interaction parameter b , uniform arrivals lead to a smaller value of S_L . Otherwise (especially for larger values of b) hardcore arrivals lead to a smaller value of S_L than the uniform arrivals independently of what the correct arrival strategy is: When b is large, the small newly arrived points tend to die due to interaction soon after they have arrived without being observed. Therefore, only the new points that are not too close to the existing points are able to survive. The hardcore strategy may then look like being the correct arrival strategy.

Similar results were also obtained when using the symmetric interaction function (??).

5. Application

As an example we study a Scots pine stand in Sweden (Area number 1562, Stand number 2060 - The "Lilla Edet" area), which was also studied by Cronie and Särkkä (2011). The data consist of locations and diameters at breast height (dbh) of trees recorded at three time points (years), namely $T_1 = 1985$, $T_2 = 1990$, and $T_3 = 1996$. The approximate age of the stand in 1985 was 22 years, so we set $T_0 = 1963$. The study area is a circular region of radius 10 meters. Only the trees having reached 0.1 m dbh are included in the data set. The tree patterns can be seen in Figure ???. The arrangement of the trees seems to be quite homogeneous even though there is an empty area on the right-hand side of the plot.

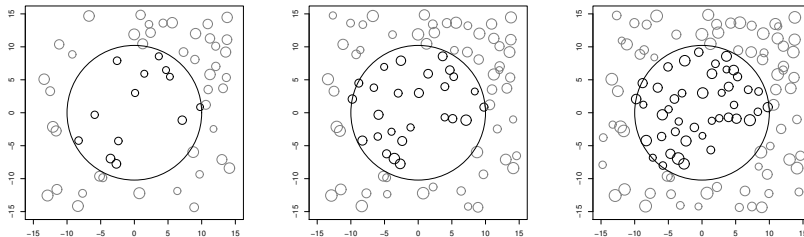


Figure 4: Swedish Scots pines (black) recorded in 1985 (left), 1990 (middle) and 1996 (right). The surroundings (grey) are simulated from the GI process fitted in Cronie and Särkkä (2011). The radii of the trees are scaled with factor 10.

Cronie and Särkkä (2011) fitted a GI model with logistic growth function, area-interaction function and uniform arrivals to these data using the LS approach described in Section ??. Due to the small size of the study region they introduced three methods for edge correction based on simulation of tree patterns

outside the study area. They found that the spatial pattern of the locations in the data was more regular than in the pattern simulated from their fitted model. Furthermore, even the mark distributions in the data and in a simulated pattern were different. The difference is clearly visible in Figure ?? where the surroundings of the study area were simulated using the model estimated by Cronie and Särkkä. We would like to emphasise that the data are used mainly to illustrate the effect of the changing arrival strategy. We have data only on a very small plot, and the results cannot be generalised. Especially, a thorough analysis of real forest dynamics would require much larger data sets.

Here, we investigate whether the new estimation approach together with the choice of hardcore arrivals improves the model fit. For this purpose, we fit a GI model with logistic power-law growth function and area-interaction function (??). We use the new estimation approach based on (??), (??) and (??) and a stochastic search algorithm as described in Cronie (2011). Choosing $\rho \equiv 1$ in the estimation of α and μ we get $\hat{\alpha} = 0.00415$ and $\hat{\mu} = 0$ independently of the growth and interaction parameters. The initial radius of new arrivals is set to the minimal recorded tree size $m_i^0 = 0.05$ m and the cases of uniform and hardcore arrivals are investigated. Since the differences between the three edge correction methods in Cronie and Särkkä (2011) were not very pronounced we decided to choose a simple approach: The estimation was run on the data sets with simulated surroundings as shown in Figure ?? where several realisations of the surrounding pattern were taken into account. However, only points in the original study area are included when computing the distance measure. The stochastic parameter search was stopped when the results had not improved during the last 500 steps.

The estimation procedure was repeated for three realisations of the surroundings. It turned out that the model with hardcore arrivals yields a better fit than the one with uniform arrivals. The estimated parameters are given in Table ?. Figure ?? shows a realisation of the model and a histogram of its mark distribution in comparison with the data.

Interestingly, the interaction strength is estimated as $b = 0$ in our case. This means that the growth of the trees is not constrained by interaction with their neighbours. In this case, the role of the parameter r is only of importance when placing new arrivals. Nevertheless, the regularity of the tree pattern is fitted more closely by our model than by the model from Cronie and Särkkä (2011). One should also notice that the two interaction parameters, r and b , depend on each other: The smaller value for b leads to a larger value of r . The estimated model still has a logistic growth function and the histogram of tree sizes looks similar to the one in Cronie and Särkkä (2011). Hence, our more general choice of the growth function did not improve the fit of the size distribution.

For a further evaluation of the goodness-of-fit of the model, the pair correlation functions and the mark correlation functions (see Illian *et al.*, 2008) estimated from both the model and the data are shown in Figure ?. The pair correlation function describes the spatial arrangement of the trees. For a completely spatially random point pattern, the pair correlation function equals 1. A value above 1 indicates clustering at the given scale, while a value below 1 is

	c_1	c_2	p	r	b
Cronie and Särkkä (2011)	0.0781	0.8230	1	4.0680	3.1626
this work	0.1000	1.0576	1	10.127	0

Table 2: Estimated parameters from Cronie and Särkkä (2011) and from this work.

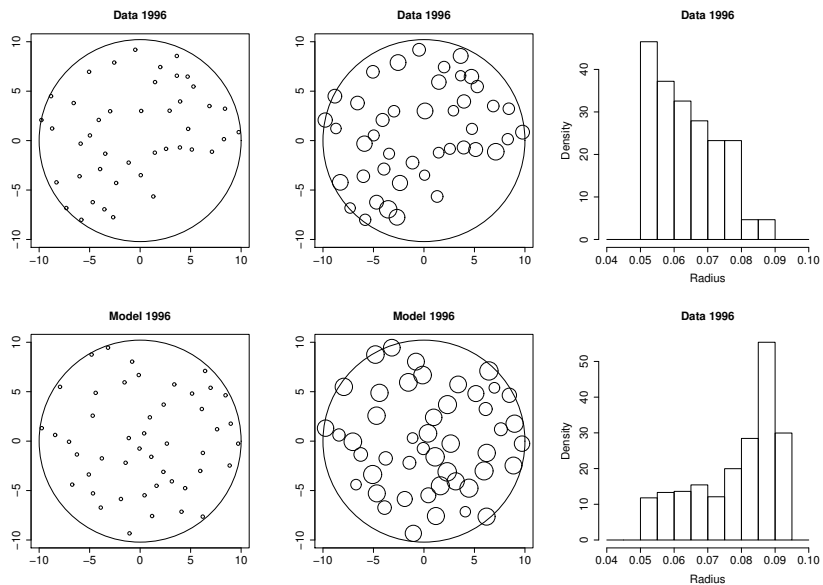


Figure 5: Comparison of the pine data in 1996 (top) and a realisation of the model (bottom). Locations of the trees (left), locations with scaled radii (factor 10, middle), and histogram of the radii (right). The histogram for the model is computed from five realisations of the model.

obtained in the regular case. The mark correlation function detects correlations between the marks. A value of 1 is obtained for no correlation at the given scale, values larger/smaller than 1 indicate positive/negative correlation.

The pair correlation functions for the data and the model show a similar structure with three peaks separated by larger valleys at roughly the same distances. A difference is observed in the height of the peaks. The model realisations are slightly more regular than the data at small distances and less regular at intermediate distances. Nevertheless, the differences are quite small, and the spatial pattern of the model is closer to that of the data than the spatial pattern of the model from Cronie and Särkkä (2011). For both the data and the model the mark correlation functions indicate only weak correlation between the marks. At small distances, slightly higher correlations are observed for the data. This is in line with the results from Cronie and Särkkä (2011).

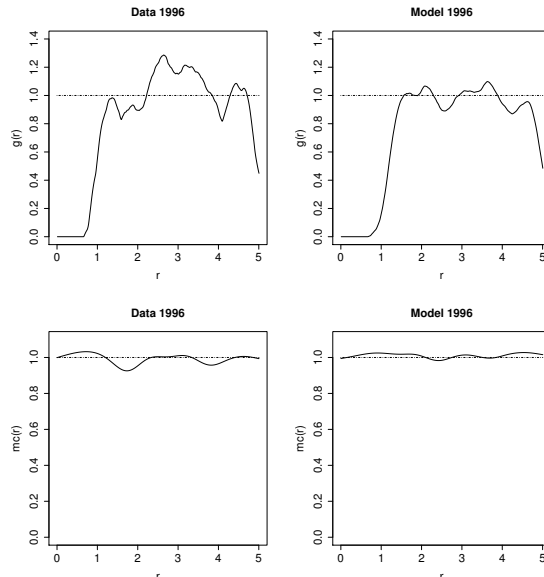


Figure 6: Pair correlation functions (top) and mark correlation functions (bottom) for the pines in 1996 (left) and averaged over five simulated realisation of the model (right).

6. Discussion

We have investigated an alternative strategy for choosing the locations of the new arrivals in the growth-interaction process introduced by Renshaw and Särkkä (2001). In the original model, the new points are located uniformly and independently of each other to the study area (uniform arrivals). Here, we also investigate a strategy, where the locations of the new arrivals are chosen such that they do not interact with the existing points as they arrive (hardcore

arrivals). Särkkä and Renshaw (2006) and Cronie (2011) suggested a least-square estimation procedure for the estimation of the growth and interaction parameters of the growth-interaction process. The difference between observed and predicted marks of the points is minimised, while the spatial information is not taken into account. We have suggested an improved estimation approach, which is based on both mark information and information on the point locations and which is able to distinguish between different arrival strategies.

According to our simulation study, it is important to use hardcore arrivals in the model if the arrivals are hardcore in reality. On the other hand, when the true arrivals are uniformly located, it may still be better to use hardcore arrivals in the model in order to obtain the correct spatial structure. This is the case when there is strong interaction between the points. The points that arrive very close to the existing ones are most likely killed very soon causing a hardcore pattern. However, our simulation study is very limited, and we do not want to generalise that it would always be the best strategy to use hardcore arrivals. This study shows that it is important to think which arrival strategy is the most appropriate one given the question at hand. Furthermore, changing the arrival strategy does not only influence the spatial location but also the survival probability, and hence the intensity of the pattern. As a next step the distance measure could be modified such that this would be taken into account.

Cronie and Särkkä (2011) fitted a growth-interaction process with logistic growth function and area-interaction function to a Scots pine data using the original least squares estimation approach and edge correction. They conclude that neither the mark structure nor the spatial structure of the real data and the data simulated from the fitted model match. The tree locations in the real data seem to be more regular than in the simulated data. By changing the arrival strategy from uniform to hardcore and by using our refined model fitting technique we were able to improve the fit of the spatial structure of the data. Significant differences in the mark distribution are still observable which might be due to the fact that the growth function was not changed compared to Cronie and Särkkä (2011). One should also note that the data collection strategy may play a role when choosing the arrival strategy. Here, trees were only recorded after they had reached a particular size. Therefore, our model describes the arrival of young trees of a particular size, not the arrival of seedlings.

In this paper, the point processes under investigation were assumed to be stationary. For the pine data this assumption could be justified due to the small size of the sampling area. Typical forest data, however, might show an inhomogeneous structure due to, e.g., local variations in soil fertility. For such data, the model could be generalised to the inhomogeneous case by using locally varying model parameters. In this case, the inhomogeneous K -function (Baddeley et al. (2000)) should be used for the model fit. Since the data are spatio-temporal, one could also use the space-time K -function introduced by Diggle et al. (1995). However, in the data we have only three time points making the estimation of the space-time function difficult. Furthermore, since comparison between the model based and observed marks is done for each observed time point, it feels natural to compare the difference in spatial structure correspondingly. We can

conclude that the estimation approach could be further generalized by using the inhomogeneous or time-space K -function.

Acknowledgements

Financial support from the Swedish Research Council and the Swedish Foundation for Strategic Research is highly appreciated. We would like to thank Ottmar Cronie for interesting discussions and Andrea Sanger for assistance with the simulation study. We are also grateful to the two anonymous referees for giving valuable comments and suggestions that led to an improvement of the paper.

References

- Baddeley, A., Møller, J., Waagepetersen, R., 2000. Non- and semi-parametric estimation of interaction in inhomogeneous point patterns. *Statistica Neerlandica* 54, 329–350.
- Comas, C., 2009. Modelling forest regeneration strategies through the development of a spatio-temporal growth interaction model. *Stochastic Environmental Research and Risk Assessment* 23, 1089–1102.
- Cressie, N., Wikle, C. K., 2011. *Statistics for Spatio-Temporal Data*, Hoboken, Wiley.
- Cronie, O., 2010. Some edge correction methods for marked spatio-temporal point process models. Preprint 2010:9, Preprint series - Department of Mathematical Sciences, Chalmers University of Technology and University of Gothenburg.
- Cronie, O., Sarkka, A., 2011. Some edge correction methods for marked spatio-temporal point process models. *Computational Statistics and Data Analysis* 55 (7), 2209-2220.
- Cronie, O., Yu, J., 2010. Maximum likelihood estimation in a discretely observed immigration-death process. Preprint 2010:10, Preprint series - Department of Mathematical Sciences, Chalmers University of Technology and University of Gothenburg.
- Cronie, O., Nystrom, K., Yu, J., 2011. Spatio-Temporal Modelling of Swedish Scots Pine Stands. In *Modelling and Inference for Spatio-Temporal Marked Point Processes*, PhD thesis by Ottmar Cronie - Department of Mathematical Sciences, Chalmers University of Technology and University of Gothenburg.
- Diggle, P.J., Chedwynd, A.G., Haggkvist, R., Morris, S., 1995. Second-order analysis of space-time clustering. *Statistical Methods in Medical Research* 4, 124-136.
- Diggle, P.J., and Gabriel, E., 2010. Spatio-temporal point processes, in *Handbook of Spatial Statistics*, Chapman and Hall, 449-461.
- Illian, J., Penttinen, A., Stoyan, H., Stoyan, D., 2008. *Statistical Analysis and Modelling of Spatial Point Patterns*, Wiley & Sons, Chichester.

Renshaw, E., Särkkä, A., 2001. Gibbs point processes for studying the development of spatial-temporal stochastic processes. *Computational Statistics and Data Analysis* 36, 85-105.

Särkkä, A., Renshaw, E., 2006. The analysis of marked point patterns evolving through space and time. *Computational Statistics and Data Analysis* 51, 1698-1718.