Spatio-Temporal Modelling of Swedish Scots Pine Stands

Ottmar Cronie, Jun Yu, and Kenneth Nyström

Research Report
Centre of Biostochastics

Swedish University of Agricultural Sciences

Report 2011:03
ISSN 1651-8543

Spatio-Temporal Modelling of Swedish Scots Pine Stands

Ottmar Cronie

Mathematical Sciences
Chalmers University of Technology and University of Gothenburg, Göteborg, Sweden

Jun Yu[1]

Centre of Biostochastics
Swedish University of Agricultural Sciences, Umeå, Sweden

Kenneth Nyström

Department of Forest Resource Management
Swedish University of Agricultural Sciences, Umeå, Sweden

[1]E-mail address to the correspondence author: jun.yu@slu.se

Abstract

Considering measurements of locations and radii at breast height made at three different time points of the individual trees in ten Swedish Scots pine plots, we employ the so called growth-interaction (GI) process for the spatio-temporal modelling of the plots. The GI-process places trees at random locations in the study region and assigns radii (sizes) to the trees, which interact and grow with time. It has been used to model Scots pine plots in previous studies, and to improve the fit we suggest some modifications of the model: A different location assignment strategy and a different function for the open-growth (growth in absence of competition). We believe also that the space-time data contain too small trees to reflect the open-growth properly, which primarily affects the carrying capacity parameter. We evaluate the open-growth from a separate set of data which consists of size and age measurements of older and larger single Scots pines. This data set better represents the open-growth of Scots pines than the space-time data sets. A linear relationship is found between the estimated site indexes of the plots and the sizes, and this relationship is exploited in the estimation of the carrying capacity. For each of the ten space-time data sets (plots) we estimate the remaining parameters of the GI-process and finally, by means of some Monte Carlo tests, we test the goodness-of-fit of simulated predictions from the fitted model.

Keywords: Basal area, Carrying capacity, Goodness-of-fit, Growth-Interaction process, Immigration-death process, L-function, Mark-correlation function, Open-growth, Richards growth function, Scots pines, Site productivity index, Spatio-temporal point process.

1 Introduction

For a long time, statistical methods for (marked) spatial point processes have been used extensively to determine various characteristics of forest stands (see e.g. [5, 7, 11, 22]). For instance, different summary statistics, such as Ripley's K-function (see e.g. [5]), have been able to cast light on the inherent mechanisms which govern the way individual trees are located spatially. Additionally, the temporal development of the sizes of individual trees has also been explicitly modelled (see e.g. [26]). Note, however, that when we are considering different spatial features of a forest stand at a fixed time, we are in fact, to some extent, also considering the temporal development of each individual tree which has been present up to the current time point. Put differently, we always regard some aspect of the spatio-temporal development of all the trees when we consider either just one single tree, or when we consider all the trees in the whole stand. Hence, if one wants a deeper understanding of the inherent growth structures which govern the growth of individual trees in a forest stand, it is reasonable to instead extend the study to consider the full spatio-temporal development of the stand, since the spatial domain and the temporal domain clearly are intertwined.

Our main objective here is to describe the development of young Scots pine stands in Sweden. The specific approach chosen here is to fit the so called growth-interaction (GI) process (see e.g. [3, 23]) to a collection of data sets. More specifically, each data set under consideration consists of measurements of individual trees in a single Scots pine stand at a few different time points. For each individual tree we have recorded its location as well as its radius at breast height (rbh) at all measurement times of the stand.

The GI-process is a spatio-temporal marked point process in which new points (trees) arrive to the study region at random times and receive random locations in the study region. Once they arrive they additionally receive sizes (radii of disks centred on their locations) and they start growing and competing until they finally die, either by competition or naturally.

A previous (initial) study similar to the current one has been conducted in [3], where the GI-process was used to model one Scots pine data set of the same type as the data sets considered here. It was indicated that the GI-process is an appropriate model for such data. However, simulated point patterns of the GI-process, which were generated based on the parameter estimates obtained by fitting the GI-process to the data, had nearest neighbour distances which were smaller than what was observed in the actual data. Furthermore, the estimated open-growth, i.e. the growth of a tree when the competition with other trees is negligible, suggested that after approximately 40 years (in open-

growth) a Scots pine had reached its theoretical maximal attainable size – its carrying capacity – and this should not be the case. In addition, the observed size distributions in the data and in the simulations were different. The suggested remedy was to change the form of the open-growth function in the model, which controls the growth of a tree in absence of competition in the model, to a more flexible one. Additionally, we also believe that by changing the way we assign the locations to the new trees in the model, we may improve the fit of the model. More specifically, instead of letting the location of a new tree to be uniformly distributed in the study region, as was previously the case (see e.g. [3, 23]), we now let the location of the tree be uniformly distributed on the part of the study region which is not occupied by the other trees present at its arrival time. Note that this is a natural assumption since trees do not grow inside each other.

As previously mentioned, the carrying capacity, denoted by K, is one of the parameters in the GI-process which governs the rate of the so called open-growth of a tree in the model. By looking closer at the carrying capacity estimates obtained in [3], we believe that the actual carrying capacity was underestimated because the forest stand considered was so young. Since biased carrying capacity estimates result in biased estimates of the open-growth behaviour, it is important to obtain fairly adequate estimates of K if we wish to say something about the individual growth ability of a tree. We believe that the underestimation of the carrying capacity which was seen in [3] is an effect of not considering trees which are close to their full potential maximum sizes, and this in turn is a consequence of the trees in the data set being young. Since the data sets we consider in this study also only contain quite young trees, we base the estimation of the carrying capacity on a completely separate set of data which contains older trees – the open-growth data set. We point out that the stands in both data sets grow in different regions of Sweden, and it is most unlikely that they follow the exact same underlying growth patterns. In particular, the soil on which one stand grows may be more fertile than the soil of the next stand, and we would like to account for such variation between the stands. The approach suggested here is to let the carrying capacity be expressed through *the site productivity index* SI, which is expressed as expected dominant height at 100 years total age, and in the current data sets it has been estimated according to [9]. The open-growth data set contains measurements made from a large number of Scots pine stands at one given time point, and each measurement consists of three parts: The size (diameter) of the largest tree, the age of the largest tree and the stand's SI-value. We believe that such a data set reflects the open-growth behaviour fairly well since by considering

the largest tree in a stand we most likely consider the tree in the stand which has been subject to the least amount of competition. We note that in [18] separate estimation of the carrying capacity has already been considered for the GI-process, and a fixed value of 0.25 meter rbh for the carrying capacity parameter was chosen.

The paper is structured in the following way. In Section 2 we present the two data sets and in Section 3 we define the GI-process. In Section 4 the estimation approach for the open-growth is presented and the estimation approach for the GI-process is recalled, and the obtained estimation results are presented. In Section 5, we evaluate the goodness-of-fit of predictions of the model. Finally, some discussions can be found in Section 6.

2 Scots pine data

In this paper we will consider two sets of data, the *space-time data set* and the *open-growth data set*. Both are taken from the Swedish National Forest Inventory (NFI). The NFI is an annual sparse stratified sample plot inventory with partial replacement [16].

2.1 Space-time data

The space-time data set, which we model by the GI-process, consists of permanent sample plots (radius 10 m) which have been established between 1983 and 1987 and re-measured for the first time after five years, i.e. between 1988 and 1992. About half of the sample was then re-measured a second time between 1993 and 1997. However from 1994 on the re-inventory interval was altered to 6 years. Registrations consist of stand, site and tree variables [24]. Tree species, diameter at breast height (1.3 m above ground), and position (with coordinates) were measured for all trees with diameter of at least 10 cm on the plots. Smaller trees were registered on a reduced plot and mapped only for a few trees. Sample trees were selected from callipered trees with probabilities given by their basal area. On sample trees also height, crown height and cause of damage were registered.

In the present study 10 stands (plots) established 1985 were used for the fitting of the GI-process, with data from 1985, 1990, and 1996. Note that the years of measurements are the same for all 10 plots, but the ages (sample times), $T_{j,1}$ (1985), $T_{j,2}$ (1990), $T_{j,3}$ (1996), $j = 1, \ldots, 10$, of the 10 plots at the measurement times differ. Additionally, a sub sample of these plots (5 plots) re-measured a fourth time $T_{j,4}$ (2005) were used to evaluate the model.

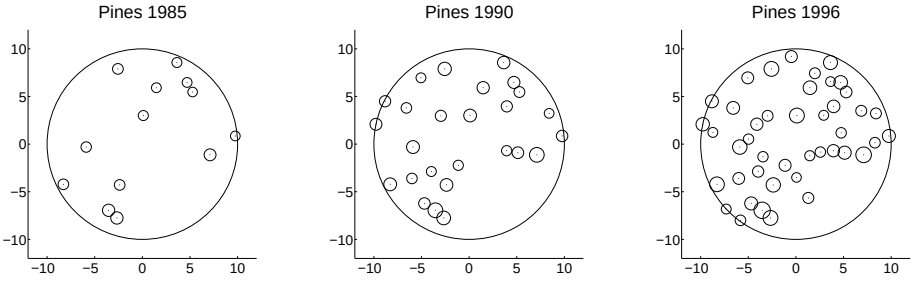

Figure 1: Swedish Scots pine plots recorded in 1985 (left), 1990 (middle) and 1996 (right). The radii of the pines are scaled by a factor of 10.

Each plot considered primarily consist of Scots pines (at least 90%) and contains at least 10 trees. It should be noted that the plots considered are quite young (see Table 1) and since only trees with a radius of at least (rbh) 0.05 m are considered, we restrict our study to the modelling of Scots pines larger than 0.05 m rbh at early ages. It should further be pointed out that for technical convenience, in the remaining part of the paper, we use rbh to describe tree sizes rather than the more common diameter at breast height. Furthermore, we will often simply refer to the rbh of a tree as its radius or size. We note that a plot may contain trees which are not Scots pines, and we have chosen to include the non-pine trees in the modelling since they affect the spatial structure of the plot as well as its temporal development. In Figure 1 we find an example of a data set. Here all tree radii have been scaled by a factor of 10 for increased visibility.

Regarding our notations, the circular spatial study region of radius 10 meters is denoted by W and recall that the three sample time points at which the plot has been measured (ages in 1985, 1990 and 1996) are denoted by $T_{j,1}, T_{j,2}, T_{j,3}$. Furthermore, the jth data set is denoted by $\mathbb{X}_j = (\mathbb{X}_j(T_{j,1}), \mathbb{X}_j(T_{j,2}), \mathbb{X}_j(T_{j,3}))$, $j = 1, \ldots, 10$, and at each time point $T_{j,k}$ we have that $\mathbb{X}_j(T_{j,k}) = \{(\mathbf{x}_i, m_{ik})\}$, where $\mathbf{x}_i$ and m_{ik} denote, respectively, the location of the ith tree and the rbh at time $T_{j,k}$ of the ith tree. We also denote by $n_{T_{j,k}}$ the number of trees which are alive at time $T_{j,k}$ and by $N_{T_{j,k}}$ the total number of distinct trees which have been observed at the sample times $T_{j,1}, \ldots, T_{j,k}$, $k = 1, 2, 3$. Besides the information given in $\mathbb{X}_j$, we also have attached to each plot a value of the site index SI. Note that for some data sets we have measurements $\mathbb{X}_j(T_{j,4})$ from an additional time point $T_{j,4}$, which will be used to evaluate predictions of the fitted GI-process. Some plot

Table 1: Information about the data sets $\mathbb{X}_j$, $j = 1, \ldots, 10$: SI is the site index, $T_{j,k}$, $k = 1, 2, 3$, is the kth inventory time (stand age), $n_{T_{j,k}}$ denotes the number of trees which are alive at time $T_{j,k}$ and $N_{T_{j,k}}$ is the number of distinct trees which have been observed at $T_{j,1}, \ldots, T_{j,k}$.

j	SI	$T_{j,1}$	$T_{j,2}$	$T_{j,3}$	$n_{T_{j,1}}$	$n_{T_{j,2}}$	$n_{T_{j,3}}$	$N_{T_{j,1}}$	$N_{T_{j,2}}$	$N_{T_{j,3}}$
1	13	23	28	34	15	21	29	15	21	29
2	14	22	27	33	13	26	43	13	26	43
3	16	45	50	56	12	15	17	12	15	17
4	17	30	35	41	2	15	23	2	15	23
5	21	29	34	40	27	45	50	27	45	50
6	19	32	37	43	24	36	48	24	36	48
7	18	25	30	36	34	39	40	34	39	40
8	20	23	28	34	40	51	52	40	51	53
9	14	45	50	56	11	14	16	11	14	16
10	15	45	50	56	9	15	15	9	15	15

characteristics are given in Table 1.

2.2 Open-growth data

Provided that two trees in a forest are close enough to each other, they will compete for resources (e.g. light and nutrients). However, if the distance between them is large enough, their competition becomes negligible. This type of growth, i.e. growth without competition/interaction, is often referred to as *open-growth* and it is often modelled by means of growth functions (see e.g. [14, 20]).

When we model the growth of trees, one major part is to model the open-growth and when we model the open-growth an important part is to estimate the theoretical maximal attainable size – the *carrying capacity*. However, since the plots in the space-time data set contain only fairly young trees, which have not reached their full sizes, these plots do not fully reflect the potential open-growth of Scots pines.

Hence, we choose to gain information about the open-growth from a separate set of data, *the open-growth data set*. The open-growth data set, which will be used to explore the open-growth relationships of Scots pines, is based on data collected in the NFI during 2003-2007. In the study 2579 pure pine plots have been used. The open-growth data set has a wide distribution of tree and stand level characteristics, i.e site index (SI) and tree age, see Fig-

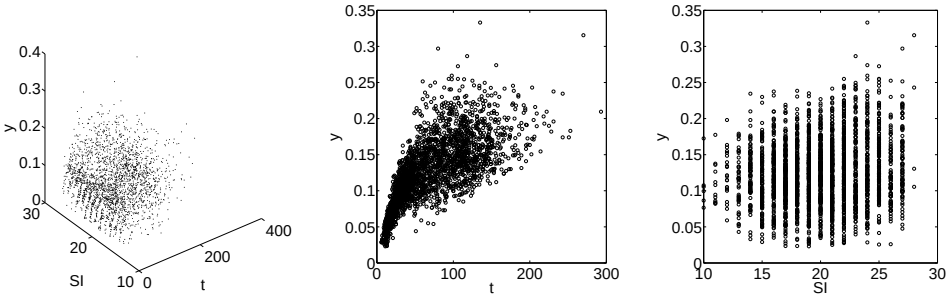

Figure 2: The open-growth data set: $\{(t_i, SI_j, y_j)\}_{j=1}^n$ (left), $\{(t_j, y_j)\}_{j=1}^n$ (middle) and $\{(SI_j, y_j)\}_{j=1}^n$ (right). For the largest tree of plot j, y_j, t_j and SI_j denote the rbh and the age of the tree, and the value of the site index SI of the jth plot, respectively.

ure 2. Hence, it contains trees which have larger sizes and higher ages than the trees in the space-time data set, and thus better reflects the open-growth. To motivate this further, we note that one of the main components of the GI-process is a growth function which controls the open-growth of the radii. Hence, when modelling the GI-process it is essential that we are able to create a good estimate of the open-growth of the pine trees.

Each observation in the open-growth data set $\{(y_j, t_j, SI_j)\}$ comes from a pure Scots pine plot which has been measured at one time point. Specifically, SI_j denotes the SI-value of the jth plot, t_j denotes the age (year) of the largest tree present in the jth plot and y_j denotes the largest tree's rbh (meter). Additionally, some initial filtering has to be performed in the open-growth data in order to ensure that we are considering trees which have been subject to as little competition as possible. We remove the observations which correspond to the largest 1% and the smallest 1% of the ratios $\{y_j/t_j\}$. This is done since a too large or too small values of y_j/t_j imply that measurement errors are likely to be present. Next, we group the remaining observations by age, so that all trees which have the same age are in the same group. Then, in each age group, we remove the observations which correspond to the largest 2.5% and the smallest 2.5% of the ratios $\{y_j/t_j\}$. As a result we have $n = 2465$ observations $\{(t_j, SI_j, y_j)\}_{j=1}^n$, where $10 \leq SI_j \leq 28$ and $6 \leq t_j \leq 293$. It should further be pointed out that there is some uncertainty present in the SI-values assigned to each plot/tree.

In Figure 2 it can be seen that the size development y over time (age) t resembles the typical shape of a growth curve model (see e.g. [20]). It may

further be observed that there tends to be a linear relationship between the maximal attainable size y and the site index SI.

3 The GI-process

Consider the scenario where we have measured a plot at the time points $0 < T_1 < \ldots < T_n = T$ within a study region W of size $\nu(W)$. Recall that in the current study, W is given by a circular region. As a model for its spatio-temporal development we suggest the so-called growth-interaction (GI) process [1, 3, 18, 23].

In the GI-process the arrivals of new trees (points) to W occur according to a Poisson process on $[0,T]$ with rate $\alpha\nu(W)$, $\alpha > 0$. At their arrival times $B_1 < \ldots < B_N$, the N trees which arrive during $[0,T]$ are assigned the locations (stock centres) $X_1, \ldots, X_N$ on W.

The size/radius (rbh) of the ith tree at time $t \in [0,T]$ is denoted by $M_i(t)$ and since we only observe the trees once their radii have reached a certain size (recall that our modelling data consists of trees of at least 0.05 m rbh), we let $M_i(t) = 0$ for all $t < B_i$ and assign the initial size $M_i(B_i) = M_0 > 0$ to the ith tree. As time evolves the radii $\{M_i(t)\}_{i=1}^N$ will grow and interact with each other. Given that the ith tree has size $M_i(t)$ at time t, its size-change over the (infinitesimal) time interval $(t, t+dt)$ is given by

$$\begin{aligned} M_i(t+dt) &= M_i(t) + dM_i(t) \\ &= M_i(t) + \left(f\left(M_i(t);\theta\right) - \sum_{\substack{j=1 \\ j\neq i}}^{N} h\left(M_i(t), M_j(t), X_i, X_j;\theta\right) \right) dt, \end{aligned} \tag{3.1}$$

where θ is a parameter vector which controls the growth and interaction pattern of the model and $f\left(M_i(t);\theta\right)$ is the growth function which governs the open-growth of the ith tree. The function $h\left(M_i(t), M_j(t), X_i, X_j;\theta\right)$ is the spatial interaction function which handles the spatial pairwise interaction/competition of the ith tree with the other (neighbouring) trees.

In accordance with [1, 3], a tree can also experience a (size dependent) *natural death*: Given $M_i(t)$, the probability that the ith tree suffers a natural death during the (infinitesimal) time interval $(t, t+dt)$ is given by $\mu/(1 + M_i(t)) + o(dt)$, $\mu > 0$ (equivalently, at each time t the potential remaining lifetime of the ith tree is $Exp(\mu/(1+M_i(t)))$-distributed), and we note that a tree becomes more viable as it grows in size. Additionally, the competition for resources may cause a tree to die and we say that the ith tree has suffered a *competitive death* if at some time $t > B_i$ we have $M_i(t) \leq 0$. Note that hereby

the death time of the ith tree is given by the smallest of the natural death time and the competitive death time, and we let $M_i(t) = 0$ once the tree has died.

Returning to the locations of the trees, we let the location X_i of the ith tree be uniformly distributed on the part of W which is not occupied by other (the previous) trees at time $t = B_i$, i.e. $X_i \sim Uni(W \setminus \bigcup_{j=1}^{i-1} B_{X_j}[M_j(B_i)])$, where the closed disk $B_{X_j}[M_j(t)]$, with centre X_j and radius $M_j(t)$, represents the (breast height) space which is occupied by the jth tree at time t. Note that we hereby avoid pairs of points which are too close to each other (trees do not grow inside each other). We further also note that this approach differs from the earlier arrival strategy, where the locations are uniformly distributed on all of W (see e.g.[23]).

We next discuss the specific choices made for the functions $f(\cdot)$ and $h(\cdot)$ and details concerning them.

3.1 Spatial interaction

We now turn to the second function contained in expression (3.1) – the spatial interaction function – which controls the competition between the trees. Many options are available for $h(\cdot)$ (see e.g. [13, 19, 23]) and we here follow, among others, [23] by employing the so called *area interaction function* (see e.g. [6]). Its exact form is given by

$$h\left(M_i(t), M_j(t), X_i, X_j; \theta\right) = c\, \frac{\nu\left(B_{X_i}\left[rM_i(t)\right] \cap B_{X_j}\left[rM_j(t)\right]\right)}{\nu\left(B_{X_i}\left[rM_i(t)\right]\right)}, \tag{3.2}$$

where $\nu\left(B_{X_i}\left[rM_i(t)\right]\right)$ denotes the size of the closed disk $B_{X_i}\left[rM_i(t)\right]$ and the elements $c \geq 0$ and $r \geq 1$ of the parameter vector θ are referred to, respectively, as the *force of interaction* and the *scale of interaction*. The region $B_{X_i}\left[rM_i(t)\right]$, which represents the region in which the tree competes for resources, will be referred to as the *influence zone* of the ith tree at time t (see e.g. [25]). We note that the ratio on the right hand side of expression (3.2) is the proportion of the ith influence zone which is covered by the jth influence zone at time t. Moreover, we note that if the distance $\|X_i - X_j\|$ is larger than $rM_i(t) + rM_j(t)$, the influence zones do not overlap, whereby no interaction takes place between the ith and the jth tree during $(t, t + dt)$.

The motivation for this choice of interaction function mainly comes from its non-symmetry. More specifically, large trees influence small trees more than small trees influence large trees.

3.2 Open-growth

As previously mentioned, the *open/individual growth function* $f(M_i(t);\theta)$ in expression (3.1) controls the growth of a tree in absence of competing neighbouring trees. We note that in the case of no interaction, i.e. when $h(\cdot) = 0$, expression (3.1) turns into the equation $dM(t)/dt = f(M(t);\theta)$, $M(0) = M_0$, which has $M(t)$ as its solution (for simplicity we here write $M(t)$ for $M_i(t)$). In the literature many different applications of different growth functions can be found (see e.g. [20, 26]). We here present two general growth functions which will be applied in this paper. Note that the parameters λ, K and δ, which figure in the growth functions below, are elements of the parameter vector θ.

The first model considered is the Richards growth function (RGF) (see e.g. [12, 20, 19]) and we have that

$$M(t) = K\left(1 + \left((M_0/K)^{\delta} - 1\right)\mathrm{e}^{-\lambda t}\right)^{1/\delta}, \tag{3.3}$$

where $\delta \neq 1$, $K > 0$ and $\lambda > 0$. We note that through the RGF other growth functions may be derived [12, 20]. For instance, the *logistic growth function* (LGF) may be obtained by setting $\delta = -1$ in (3.3). We point out that the LGF previously has been evaluated in the context of the GI-process (see e.g. [3, 19, 23]).

By transforming cumulative distribution functions (cdf's), different growth functions may be obtained [20]. Another model which we will consider is the *Weibull growth function* (WGF) which is obtained through the cdf of the Weibull distribution. More specifically, we have that

$$M(t) \quad = \quad K - (K - M_0)\,\mathrm{e}^{-(\lambda t)^{\delta}},$$

where $K > 0$, $\lambda > 0$ and $\delta > 0$.

Both the RGF and the WGF are strictly increasing functions with *carrying capacity* (theoretical upper bound) K and parameters λ and δ controlling the growth rate/speed of $M(t)$.

4 Estimation

4.1 Estimating the open-growth

We recall the open-growth data set $\{(t_j, SI_j, y_j)\}_{j=1}^{n}$, $n = 2465$, from Section 2.2, and also the observed linear relationship between the maximal attainable size and the site index SI, which was seen in Figure 2. We argue that since the carrying capacity K tells us how large a tree is allowed to become, it

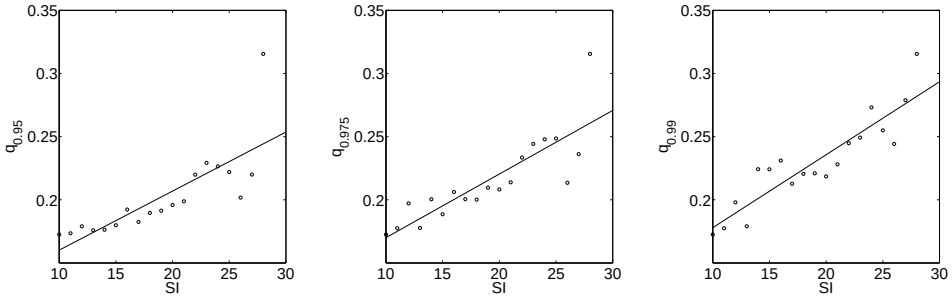

Figure 3: Fitted linear regressions w.r.t. the pairs $(SI, q_\alpha(SI))$. Left $\alpha = 0.95$, middle $\alpha = 0.975$ and right $\alpha = 0.99$.

seems sensible to let K be reflected by the fertility, i.e. the site index SI. This linear relationship becomes even more clear in Figure 3, where we have divided the observations into groups based on their SI-values, calculated the quantiles $q_\alpha(SI)$, $\alpha = 0.95, 0.975, 0.99$, of the y_j-values in each SI-group and plotted the pairs $(SI, q_\alpha(SI))$, $SI = 10, \ldots, 28$, together with the fitted linear regression models.

We additionally see from Figure 2 that the size increases with time and therefore, when α is large, $q_\alpha(SI)$ will only be concerned with the older trees in the SI-group. Furthermore, from Table 2 we see that the group with SI-value 28 only contains four measurements and the largest rbh is 0.3155, which explains why $q_\alpha(28)$ becomes quite extreme. When ignoring $q_\alpha(28)$ and fitting linear regression models, we obtain $q_{0.95}(SI) = 0.1353 + 0.0033SI$, $q_{0.975}(SI) = 0.1365 + 0.0040SI$, and $q_{0.99}(SI) = 0.1305 + 0.0051SI$. All three linear trends (intercept and slope) are significant at the 0.05 significance level and QQ-plots together with Lillie-tests (H_0 : data is Gaussian, H_1 : data is non-Gaussian) suggest that the residuals are normally distributed (Lillie-test p-values: $p_{0.95} = 0.2082$, $p_{0.975} = 0.0987$ and $p_{0.975} > 0.5$). We have thus established that there is some linear relationship between SI and K.

The usual approach to fitting a growth curve model to size measurements made over time is to fit the (nonlinear) regression model $Y_j = M(t_j; \theta) + \epsilon_j$ to the time-size measurements (t_j, Y_j), $j = 1, \ldots, n$, where $M(t_j; \theta)$ is a growth function and the ϵ_j's are iid random variables with mean 0 and variance σ^2. We note that by accepting the linear relationship $K = a_0 + a_1 SI$ for the upper bound of the rbh, we are considering the situation where we fit the regression model

$$Y_j = M(t_j; K, \lambda, \delta) + \epsilon_j = M(t_j; a_0 + a_1 SI_j, \lambda, \delta) + \epsilon_j \tag{4.1}$$

Table 2: Information about each SI-group, where n is the number of observations in the group and $\max y_j$ is the size of the largest tree in the group.

SI	10	11	12	13	14	15	16
n	6	12	21	45	92	123	154
$\max y_j$	0.1725	0.1775	0.1980	0.1790	0.2345	0.2300	0.2375
SI	17	18	19	20	21	22	23
n	210	207	254	232	246	216	157
$\max y_j$	0.2190	0.2385	0.2350	0.2270	0.2415	0.2550	0.2865
SI	24	25	26	27	28		
n	156	153	87	90	4		
$\max y_j$	0.3330	0.2725	0.2525	0.2970	0.3155		

to our data. Here $M(t;\cdot)$ either is given by the RGF, the WGF or the LGF, and we have chosen to include the LGF (as a comparison) since it is included in most of the previous studies which consider the GI-model (see e.g. [3, 19, 23]).

Before fitting the growth curve model (4.1) to the open-growth data, we merge the three first SI-groups as well as the last two SI-groups since there is some uncertainty present in the SI-values, and since some of the SI-groups are quite small. The new SI-values assigned to the new groups are obtained in the following way. We take a weighted mean of the SI-values of the included (old) SI-groups and the weights are determined by the number of observations in each of the SI-groups included. More specifically, from Table 2 we find that merging the SI-groups 10, 11 and 12, results in the SI-group SI_{10-12} which has SI-value $(6\cdot 10+12\cdot 11+21\cdot 12)/(6+12+21) = 11.3846$. Furthermore, by merging the SI-groups 27 and 28, we obtain the new SI-group SI_{27-28} which has SI-value $(90\cdot 27+4\cdot 28)/(90+4) = 27.0426$. We note that the number of observations in the groups SI_{10-12} and SI_{27-28} are 39 and 94, respectively.

Considering these new SI-groups, the results after fitting the regression model (4.1) to the open-growth data set, in the context of the RGF, the WGF and the LGF, can be found in Table 3.

Moreover, the corresponding (normal) QQ-plots of the residuals r_j are given in Figure 4. It is worth noticing that the parameter estimates do not differ much from each other in the case of the RGF and the WGF. Furthermore, we see in Figure 4 that the residuals are not normally distributed and, additionally, we see that the variance grows with the size of $\hat{y}_j$ or, equivalently, that the variance grows with increasing t_j (since $M(t;\cdot)$ is an increasing function). We further also note that the RGF and the WGF generate almost identical residuals whereas the LGF behaves a bit differently and not quite as well man-

Table 3: The results after fitting the regression model (4.1) to the open-growth data set, in the context of the RGF, the WGF and the LGF.

Param. est.	$\hat{a}_0$	$\hat{a}_1$	$\hat{M}_0$	$\hat{\lambda}$	$\hat{\delta}$
RGF	0.07418	0.00554	$6.5 \cdot 10^{-253}$	0.01752	1.29434
WGF	0.07602	0.00567	$1.5 \cdot 10^{-183}$	0.02078	0.83233
LGF	0.05164	0.00586	0.04078	0.04399	-

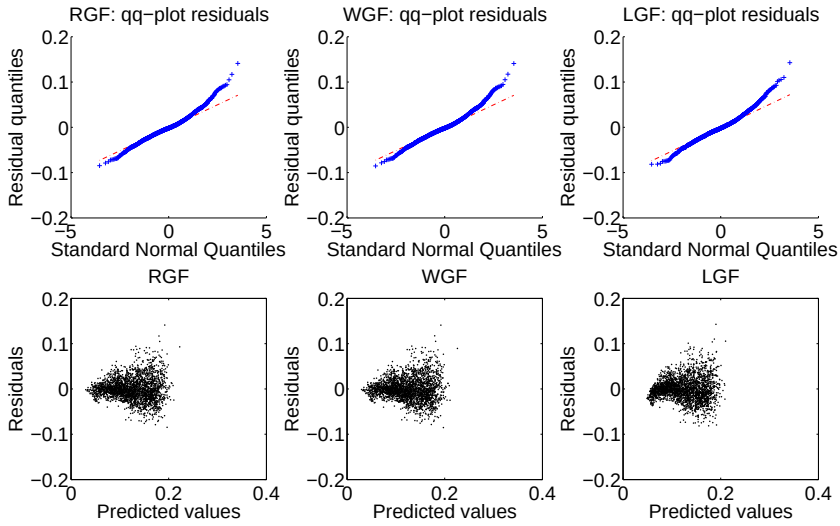

Figure 4: The residuals r_j obtained in the estimation (see Table 3) of the RGF (left), the WGF (middle) and the LGF (right). Upper row: (normal) QQ-plots of the residuals. Lower row: Plots of the predicted values versus the residuals.

ages to capture the shape of the growth over time (the plot in the bottom right corner of Figure 4 is slightly curved). Possibly a more accurate model would be to consider a multiplicative error, i.e. $Y_j = M(t_j; a_0 + a_1 SI_j, \lambda, \delta)\epsilon_j$. However, since we are mainly concerned with the estimation of K, the specification of the error term is of less importance to us.

To give an idea of the fitted growth curves, in Figure 5 we have plotted them together with the data for the SI-groups SI_{10-12}, SI_{17}, SI_{22} and SI_{27-28}. Note that we obtain $\hat{K}|SI_{10-12} = 0.13730$ and $\hat{K}|SI_{27-28} = 0.22412$ for the RGF, $\hat{K}|SI_{10-12} = 0.14063$ and $\hat{K}|SI_{27-28} = 0.22948$ for the WGF, and

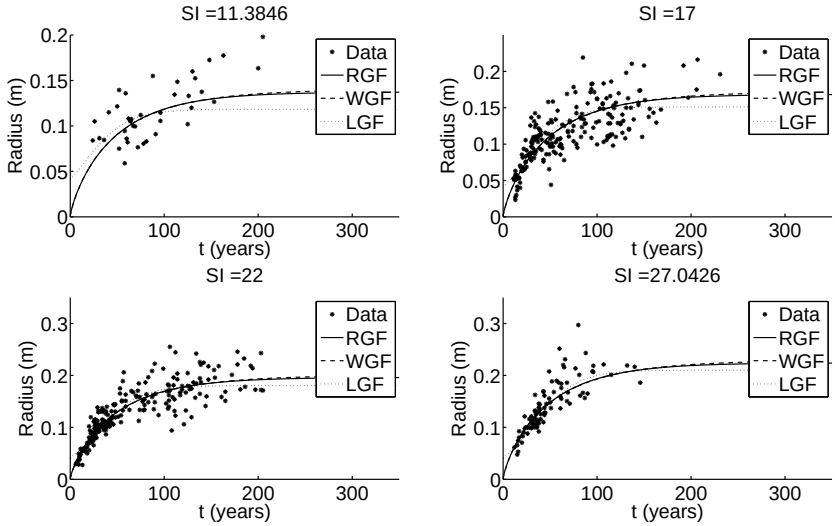

Figure 5: Plots of the estimated growth curves (see Table 3) together with the data for the SI-groups SI_{10-12}, SI_{17}, SI_{22} and SI_{27-28}.

$\hat{K}|SI_{10-12} = 0.11831$ and $\hat{K}|SI_{27-28} = 0.20100$ for the LGF. Also note that all of these are much lower than the observed maximum radii.

4.2 Estimation of the carrying capacity

As we saw in the previous section, we were unable to successfully fit the three growth curves to our open-growth data set. However, the observed linear relationship between K and SI may still be exploited in order to obtain some estimate for K.

To further explore the estimation of K, we now compare the estimates $\hat{K} = \hat{a_0} + \hat{a_1}SI$ in Table 3 with the three linear regressions $\hat{K} = q_\alpha(SI) = b_0 + b_1 SI$, $\alpha = 0.95, 0.975, 0.99$, which, just as before, are based on the α-quantiles of each (new) SI-group. The results obtained can be found in Table 4 and just as before the linear relationship is significant in each case and the residuals can be assumed to be normally distributed (Lillie-test p-values: $p_{0.95} > 0.5$, $p_{0.975} = 0.0587$ and $p_{0.99} = 0.3147$).

We thus conclude that, for all SI-values, the maximal attainable size of a tree suggested by $\hat{K} = q_\alpha(SI) = \hat{b}_0 + \hat{b}_1 SI$ is larger than the maximal attainable size suggested by any of the (growth function) estimates of K in

Table 4: Linear relationship between the α-quantiles $q_\alpha(SI)$ and SI ($\alpha = 0.95, 0.975, 0.99$).

α	$\hat{b}_0$	$\hat{b}_1$	$q_\alpha(SI_{10-12})$	$q_\alpha(SI_{27-28})$
0.95	0.12761	0.00367	0.16944	0.22698
0.975	0.12920	0.00441	0.17936	0.24835
0.99	0.13248	0.00517	0.19138	0.27240

Table 5: Estimates obtained after using the estimate $\hat{K} = q_{0.975}(SI) = 0.12920 + 0.00441SI$ and fitting the RGF, the WGF and the LGF to the open-growth data set.

Param. est.	$\hat{M}_0$	$\hat{\lambda}$	$\hat{\delta}$
RGF	$1.8 \cdot 10^{-12}$	0.00821	1.80736
WGF	$6.7 \cdot 10^{-23}$	0.01456	0.68572
LGF	0.06689	0.01789	-

Table 3. As we can see from Figure 5, for SI_{10-12} the largest y_j-values reach approximately 0.2 and for SI_{27-28} they reach approximately 0.3, whence, it seems more reasonable to use e.g. $\alpha = 0.975$ when we employ $q_\alpha(SI) = \hat{b}_0 + \hat{b}_1 SI$ to estimate K. Note that we do not use $\alpha = 0.99$ because extreme values in the data may contain measurement errors. We note that this choice of α corresponds, more or less, to the fixed estimate $\hat{K} = 0.25$ used in [19], which was motivated by the study conducted in [15]. In Table 5 we find the estimates obtained after plugging $\hat{K} = q_{0.975}(SI) = 0.12920 + 0.00441SI$ into the three growth curve models and, subsequently, estimated the remaining parameters, as explained in Section 4.1.

In Figure 6 we have illustrated the fitted models and conclude that also here the performance of the RGF and the WGF are very similar, whereas the LGF behaves quite differently. The LGF has only two parameters (instead of three) and it does not seem to manage to capture the growth at early ages. Since the LGF can be obtained as a special case of the RGF and since its performance, arguably, is a little poorer than the performance of the RGF and the WGF, we choose not to proceed with the LGF in the spatio-temporal evaluation. Furthermore, since the RGF and the WGF perform equally well, we choose to only proceed with the evaluation of the RGF in the remainder of the paper.

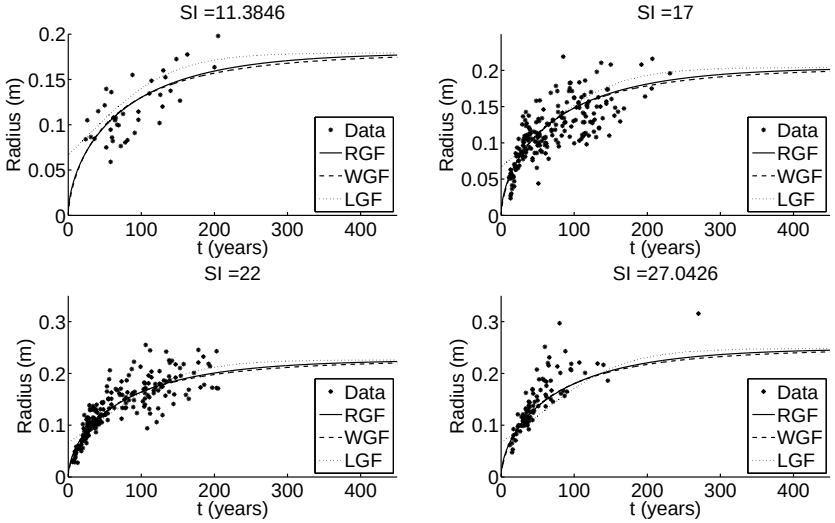

Figure 6: The estimated growth curves (see Table 5) in the context of the estimate $\hat{K} = q_{0.975}(SI) = 0.12920 + 0.00441SI$, for SI_{10-12}, SI_{17}, SI_{22} and SI_{27-28}.

4.3 Estimation of the GI-process parameters

Having estimated K separately by $\hat{K} = q_{0.975}(SI) = 0.12920 + 0.00441SI$, we now turn to the space-time data and the estimation of the remaining parameters, i.e. $\theta^* = (\lambda, \delta, c, r)$, μ and α. The estimation approach previously used to fit the GI-process to space-time data was first introduced in [23] and in [3] it was adjusted to accommodate the size dependent natural deaths. It consists of two separate parts: The arrival and death rates α and μ are estimated using a maximum likelihood (ML) estimation procedure while the growth and interaction parameters θ are estimated (separately) through a least squares scheme. In the case of θ, we here follow the approach suggested in [3], with the exception that we use the estimate $\hat{K} = q_{0.975}(SI)$ for the carrying capacity and estimate θ^* by means of the least squares approach.

In order to obtain the estimates $\hat{\theta}^* = (\hat{\lambda}, \hat{\delta}, \hat{c}, \hat{r})$ for a given data set $\mathbb{X}_j$, we minimise the sum of squared differences $S(\theta^*) = \sum_k \sum_i (\tilde{m}_{ik}(\theta^*, \hat{K}) - m_{ik})^2$ between the observed sizes m_{ik} and the predicted (model based) sizes $\tilde{m}_{ik}(\theta^*, \hat{K})$. The first sum is taken over sample time points and the second sum is taken over all trees which are alive at a specific sample time point. In finding $\hat{\theta}^*$ we

have to rely on some numerical minimisation routine to minimise the sum of squares $S(\theta^*)$, and we here choose to employ the same MCMC type of method as was used in [3] (more details are given in [2]). Additionally, since there are no observations available outside the study region W, we need some approach to impede the consequences of the so called edge effects (see e.g. [5, 11]). Three spatio-temporal edge correction methods, which were all presented in the context of the GI-process, are suggested in [3] and here we employ the so called *simple edge correction method* of [3] when we estimate θ^*.

Recall now the two previously mentioned ML-estimation approaches considered for the estimation of α and μ. The first approach, which was derived in [2] and later employed in [3], is an approximate ML-approach which takes into consideration that the size dependent natural deaths are governed by the probability $\mu/(1+M_i(t))+o(dt)$. The second approach, which was developed in [4], is a full ML-estimation scheme based on the discretely sampled *immigration-death process* (see e.g. [4, 8, 17]). We note that using the latter approach is equivalent to letting the natural death probability be given by $\mu+o(dt)$, hence not taking the size dependent natural deaths into consideration. Furthermore, in this approach, when it compares the increase/decrease of number of trees between two consecutive time points, it does not account for the competitive deaths which have taken place. Although the approach of [4] not fully matches the current model formulation, it manages better to take into account all the unobserved arrival/death scenarios where some trees arrive and die during the same time interval $(T_{j,k-1}, T_{j,k})$, $k = 1, \ldots, n$, without being observed, and as a result it generally performs better than the estimators of [2]. Hence, we will use the ML-estimators of [4] to estimate α and μ.

4.4 Estimation results

Concerning the starting values used when we run the edge corrected estimation procedure, we use the edge-corrected estimates obtained for the Scots Pine data set considered in [3] as starting values for the interaction parameters c and r, i.e. $\hat{c} = 3.5$ and $\hat{r} = 4$. Moreover, since [3] used the LGF as open-growth function, for the RGF-part $f(\cdot;\theta)$ we choose as starting values $\hat{\delta} = -1$ and $\hat{\lambda} = 0.1$ (the latter being close to the value obtained in [3]). Furthermore, we keep the estimate $\hat{K} = q_{0.975}(SI) = 0.12920 + 0.00441SI$ and the (known) initial size $M_0 = 0.05$ fixed throughout the whole estimation. Regarding the specifics of the edge correction, for each data set we use three simulated surroundings in each iteration and we average over the estimates of the last four iterations to obtain the final estimates (see [3] for details). Furthermore, we let the region on which we simulate the surrounding trees be given by

Table 6: Edge corrected estimates of the growth and interaction parameters together with their means and standard deviations (S.d.). SI is the value of the site index.

j	$\hat{K}$	$\hat{\lambda}$	$\hat{\delta}$	$\hat{c}$	$\hat{r}$	SI
1	0.18653	0.03450	0.07144	7.38389	4.25188	13
2	0.19094	0.05483	-0.47340	3.24789	3.74719	14
3	0.19976	0.08825	-0.83724	4.53837	5.92735	16
4	0.20417	0.02873	-0.46213	4.88429	5.14329	17
5	0.22181	0.08086	-0.86521	5.56370	4.57572	21
6	0.21299	0.07421	-0.83360	3.57790	2.98364	19
7	0.20858	0.05826	-0.59818	4.53696	2.57067	18
8	0.21740	0.04112	-0.35130	3.66954	4.86772	20
9	0.19094	0.06175	-0.88204	3.12824	5.35041	14
10	0.19535	0.02824	-0.75410	2.80455	3.44803	15
Mean	-	0.05508	-0.59858	4.33353	4.28659	-
S.d.	-	0.02165	0.30408	1.38185	1.08610	-

a square region of side length 25 m. In Table 6 we give the edge corrected parameter estimates for all the data sets.

We see that the parameter estimates in Table 6 for the 10 data sets are quite similar, except for plot 1, as was to be expected since the plots are quite similar. In the case of the open-growth parameters, we see that SI does not vary much between the plots and thus the estimates $\hat{K}(SI) = 0.12920 + 0.00441SI \approx 0.2$ also do not vary much. We point out that the estimates of the open-growth rates λ and δ on average behave a bit differently here where there is competition present, compared to the case of the open-growth data set. In the cases where δ is estimated to approximately -1, we have indications that the open-growth behaves almost like the LGF. Furthermore, when $\hat{\delta} \approx 0$ (as in the case of plot 1) we obtain an estimated open-growth which, in the competitive settings of these plots, behaves approximately like the so called Gompertz model (see e.g. [12]). We further also note that λ is estimated much larger than the open-growth data estimate $\hat{\lambda} = 0.00821$ of Table 5. However, this was expected since the inclusion of competition (which inhibits the open-growth) forces the open-growth rate to be higher. Additionally, since all estimates of δ were significantly smaller than the estimate $\hat{\delta} = 1.80736$ in Table 5, this also suggests that δ is decreased in order to compensate for the competition.

We see that $\hat{r} \approx 4.3$, which means that the range of the competition of a tree is estimated to be approximately 4.3 times the radius of the tree. As

an example we see that a newly arrived tree competes for resources within a distance of approximately $\hat{r}M_0 = 4.3 \cdot 0.05 = 0.215$ m from its stem centre, whereby its influence zone has size 0.145 m^2. Similarly, if the estimates of Table 6 also would be valid for the stands when they are older, whereby the individual trees are older and larger, we see that the competition distance is approximately $\hat{r}\hat{K}(SI) \approx 4.3 \cdot 0.2 = 0.86$ m and its influence zone has size 2.323 m^2. For the force of interaction, we see that that the strength of the competition between a tree and its neighbour (within competing distance), i.e. the amount by which we inhibit the open-growth during $(t, t + dt)$, is approximately given by $\hat{c} \approx 4.3$ times the proportion of the tree's influence zone which is overlapped by its neighbour's influence zone. We point out that the interaction parameters c and r (and their estimates) strongly depend on each other [3] and, there is also a dependence between the open growth parameters λ and δ (and their estimates) since they both control the open-growth rate. Now, if we were to fix, say, λ and δ in the estimation, as a result the estimates of c and r would be changed/adjusted to fit the growth of the observed trees (at least to some extent). Hence, there is dependence between all parameters λ, δ, c, r.

Note further that the numerical algorithm used to minimise the sum of squares $S(\theta^*)$ stepwise samples random values for θ^* as proposed estimates, and then determines if the new proposal results in a smaller $S(\theta^*)$ than the previous minimising θ^* does. The stopping criterion used in this stepwise procedure stops the algorithm once we have seen a threshold number of consecutive proposals without any minimisation of $S(\theta^*)$. Hence, more precise estimates would possibly be obtained if this threshold were chosen larger. However, using a larger threshold would be more time consuming since more calculations of $S(\theta^*)$ would be needed.

We now turn to the estimates of the arrival and death rates α and μ. They can be found in Table 7, and we note that they are based on the counts n_{j,T_1}, n_{j,T_2} and n_{j,T_3} in Table 1.

If we were to assume that the natural death probabilities are given by $\mu + o(dt)$ (immigration-death process arrivals and natural deaths) and that no trees are present at time 0, the number of trees present at time $T_{j,k}$ will be Poisson distributed with parameter $\hat{\alpha}\nu(W)(1 - \mathrm{e}^{-\hat{\mu}T_{j,k}})/\hat{\mu}$, where $\nu(W) = 10^2\pi \approx 314$ (see [4]).

In the case of, say, plot 9, by considering Table 7 this translates to the estimated expected number of alive trees at $T_{j,4} = 65$ being 19.4979 and, equivalently, the estimated standard deviation of the number of trees alive at $T_{j,4}$ being $\sqrt{19.4979} = 4.4157$. Further characteristics of $\mathbb{X}_j(T_{j,4})$, for the data

Table 7: Estimates of α and μ for each plot.

j	$\hat{\alpha}$	$\hat{\mu}$
1	0.00633	0.07304
2	0.19439	2.34876
3	0.00134	0.01038
4	0.11781	2.46738
5	0.40926	2.85716
6	0.49576	4.32636
7	0.00335	0.01580
8	0.41909	2.58157
9	0.00136	0.01161
10	0.13438	2.81448
Mean	0.17831	1.75065
S.d.	0.19458	1.57865

sets which have a four inventory time, can be found in Table 8. Furthermore, to have some idea of how the data sets behave on average w.r.t. the number of alive trees, we may use the means in Table 7 to obtain estimates of the expectation and standard deviation of the number of alive trees at, say, $T_{j,4} = 65$, and we obtain, respectively, the values 31.9976 and 5.6567.

Since the GI-process that we are fitting uses size dependent natural deaths, these estimates of the behaviour of the number of trees alive are not totally correct (although they are almost correct in many cases). However, by comparing the number of alive trees $n_{T_{j,4}}$ at the fourth time point $T_{j,4}$ with these estimates, we gain some insight to whether the estimates $\hat{\alpha}$ and $\hat{\mu}$ are too far off. For instance, in the case of data set 9 we have that $n_{T_{9,4}} = 23$, so from the simple (approximate) prediction of 19.4979 we see that our estimates seem fairly acceptable. We further point out that when we consider really large times $T_{j,4}$, the expected number of trees alive at $T_{j,4}$ will be approximately $\hat{\alpha}\nu(W)/\hat{\mu}$ (see [4]). One may exploit this fact in future studies to improve the estimates $\hat{\alpha}$ and $\hat{\mu}$. This follows since if there is information available about the (approximate) maximum amount of trees which occupy a study region W in old Scots pine stands, we can use this maximum number (divided by the size $\nu(W)$) to estimate the ratio α/μ, and this estimate may then in turn be used as a condition/restriction when we estimate α and μ.

It is likely that the estimates of Table 7 are quite biased since we base the estimation on only the three observations $n_{T_{j,1}}$, $n_{T_{j,2}}$ and $n_{T_{j,3}}$. Furthermore, it was seen in [3], in the evaluation of the size-dependent estimators of [2],

that μ became heavily biased and this further strengthens the belief that the estimates in Table 7 are biased.

5 Goodness-of-fit of the fitted predicted model

Having fitted the GI-process to the data, we now want to test the goodness-of-fit of the model and for this purpose we use measurements of the data at a later time point $T_{j,4} > T_{j,3}$ as reference.

To study, among other things, the spatial structure, we generate, say, 999 predictions of the jth data set at its fourth sample time $T_{j,4}$ by starting the simulation of the GI-process with the estimated parameters in the marked point pattern at time $T_{j,3}$, $\mathbb{X}_j(T_{j,3})$, and from here running independent simulations up to the subsequent sample time point $T_{j,4}$. We then compare the observed and the simulated data at time $T_{j,4}$ by looking at empirical distributional properties of some summary statistics (by means of e.g. Monte-Carlo test; see [5, 11]).

We have chosen a variance stabilised version of Ripley's K-function, the L-function, together with the mark-correlation function (mcf) as summary statistics to describe the spatial behaviour of locations and sizes (see e.g. [5, 11]). Note that the value of $K(r)$ is proportional to the expected number of further trees (points) within distance r of the typical tree, and the L-function is defined as $L(r) = \sqrt{K(r)/\pi}$. The mcf, $k(r)$, is a measure of the dependence between the mark (radius) sizes of points at a given distance r from each other. More specifically, in a marked point pattern with mean mark (average radius size) $\bar{m}$, by letting m_1 and m_2 be the marks (radii) of two typical points (trees) which are distance r from each other, we define $k(r)$ as the expected size of the product $m_1 m_2$, divided by $\bar{m}^2$. In both cases we compare the observed summary statistics with envelopes based on the predicted summary statistics (see e.g. [5, 11]).

We have also chosen to look at some summary statistics typically used in forestry. The first statistic considered is the *total basal area per hectare* $TBA = 10^4 \sum_i m_{i4}^2 \pi / \nu(W)$. The other statistics considered are the *basal area weighted mean diameter* $WMD = \sum_i (2m_{i4})(m_{i4}^2\pi) / \sum_i m_{i4}^2 \pi$, the mean diameter $\bar{m}$ and the *number of trees alive* at $T_{j,4}$, $n_{T_{j,4}}$.

In Table 8 we find the observed and the expected number of alive trees, for those plots that have a fourth inventory occasion. We see clearly that our predictions (expectations) tend to underestimate the number of alive trees.

Below follow the results for only one plot, namely plot 9. We use the estimates of Table 6 and Table 7 to generate the predictions, and the L-

Table 8: Observed and expected number of alive trees, as well as stand ages, for the plots with a fourth inventory occasion.

j	2	6	8	9	10
$T_{j,4}$	42	52	43	65	65
$n_{T_{j,4}}$	52	54	36	23	16
$\hat{\alpha}\nu(W)(1-\mathrm{e}^{-\hat{\mu}T_{j,4}})/\hat{\mu}$	26	36	51	19.5	15

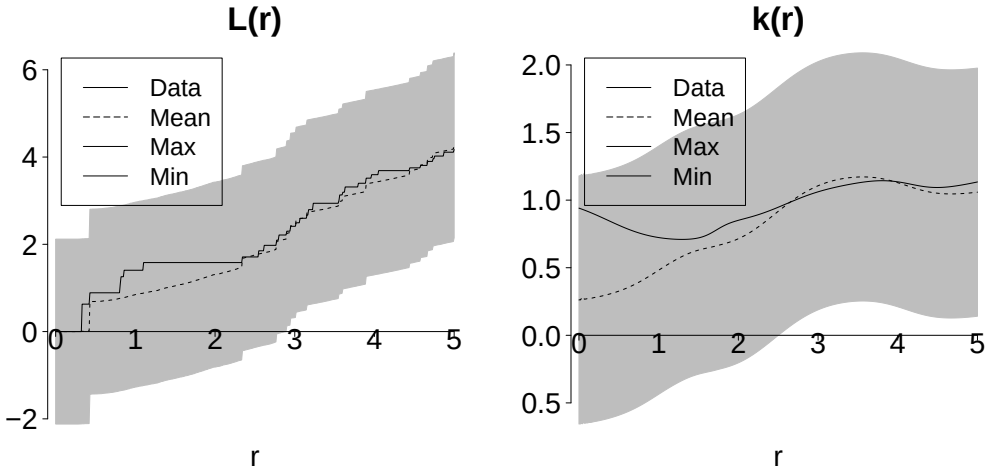

Figure 7: Left: L-function, $L(r)$, for pine data set 9 at the last sample time $T_4 = T_{9,4} = 65$, together with simulated envelopes. Right: Mark correlation function, $k(r)$, for pine data set 9 at the sample time T_4, together with simulated envelopes. The predictions (realizations) which are used to generate the envelopes are created by starting 999 simulations of the GI-process in the initial state $\mathbb{X}_9(T_3)$ and, based on the parameters in Tables 6 and 7, running the simulations up to the subsequent sample time point T_4.

function and mcf plots together with the prediction based envelopes can be found in Figure 7.

Both in the case of the L-function and the mcf, we see from Figure 7 that the curves estimated from the data clearly are inside the envelopes, and we hereby conclude, based on the spatial structure, that the point pattern and its marks can be generated by the estimated GI-process. Note, however, that in general most data trees at time $T_{9,3}$ and $T_{9,4}$ are the same, but some new

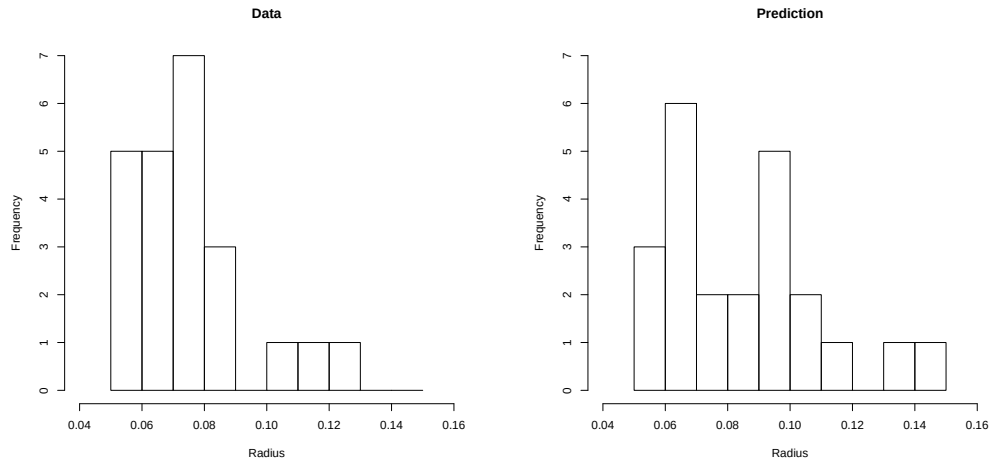

Figure 8: Radius histograms of the data (left) and a prediction (right). The prediction has the same number of alive trees as the data.

trees appear and some existing ones may die between the two time points. Moreover, it has been observed that the estimates of α and μ affect the spatial structure to a large extent.

In the case of the remaining summary statistics, the comparison of the data and the predictions can be made in the following way. Given one of the summary statistic, S, let $\hat{S}$ denote the estimate of S based on the data and let $\hat{S}_1, \ldots, \hat{S}_{999}$ denote the summary statistic estimated from the 999 predictions. In order to assess whether to accept the fit of the model (w.r.t. S), we order the 1000 estimated summary statistics $\hat{S}, \hat{S}_1, \ldots, \hat{S}_{999}$ according to their increasing sizes, and check if the *rank* (position) of $\hat{S}$ is either very small or very large. Note that the rank can be used to formally test the fit (test statistic S) by means of Monte Carlo tests (see e.g. [5, 10]). Note also that in the case of $L(r)$ and $k(r)$ tests can be constructed which are based on whether the estimated curves fall outside the envelopes at any r (max-min simultaneous/global envelopes).

In Figure 8 we find mark histograms for the data and one of the predictions. Both have the same number of alive trees.

We see that the prediction (right) suggests that the radii can become larger than the radii in the data (the largest tree in the prediction has radius 0.141 m and in the data it is 0.127 m). Furthermore, we also see that the radius distribution of the data has a larger proportion of smaller trees than the pre-

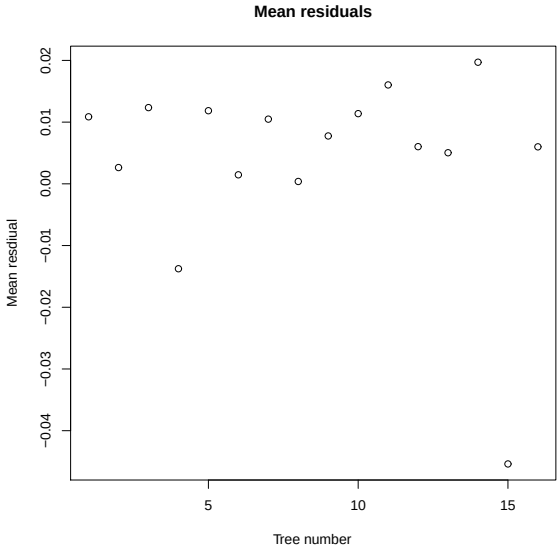

Figure 9: Mean residuals for the predictions for each tree.

dictions. It seems that the estimated model either has an open-growth which is too strong, or the competition is too weak. Moreover, this might also suggest that $\hat{\alpha}$ should be increased to increase the number of small trees.

In order to see how each tree behaves individually, we would like to measure the deviations between the actual radii and the predicted radii, for each single data tree. For a given tree i alive at $T_{9,3}$, denote by $m_{i4}^{(1)}, \ldots, m_{i4}^{(999)}$ its predicted sizes at time $T_{9,4}$, and consider the corresponding residuals. In Figure 9 we show the estimated mean residuals for each of the $n_{T_3} = 16$ trees which are alive at $T_{9,3}$.

We see that almost all predictions are larger than zero, and we again confirm that we have overestimated the growth. Note that since the trees in $\mathbb{X}_9(T_3)$ are already well established at T_3, their predicted growth during the time interval (T_3, T_4) will become almost deterministic. This follows since the newcomers in (T_3, T_4) are small and by the form of the spatial interaction function in expression (3.2), small trees do not affect larger trees much. We also note that if $\hat{\mu}$ is too high, most of the predictions of a given radius m_{i4} will have $m_{i4}^{(j)} = 0$ since these predicted trees will have suffered natural deaths. During the prediction simulation, when a data tree dies, it will leave room for simulated newcomers to grow more rapidly than they would have done

Table 9: The rank of the observed radii m_{i4} among their predictions $m_{i4}^{(1)}, \ldots, m_{i4}^{(999)}$, $i = 1, \ldots, 16$.

i	1	2	3	4	5	6	7	8
rank	93	93	90	1000	107	102	113	99
i	9	10	11	12	13	14	15	16
rank	98	78	86	91	106	91	999	92

Table 10: Observed summary statistics together with their ranks. For each summary statistic is also given the mean and standard deviation of the predicted summary statistics.

	Obs.	Rank	Mean	S.d.
TBA	13.63825	143	15.11195	1.43486
WMD	0.16978	1	0.20670	0.00596
$\bar{m}$	0.14939	1	0.18392	0.00727
n_{T_4}	23	993	16.89690	2.15259

otherwise (since they are in a place with little interaction), whence the radius distribution depends on $\hat{\mu}$.

We can also compare the predicted sizes with the actual sizes m_{i4} by checking their ranks among their predictions. In Table 9 we find the ranks of the sizes m_{i4} of the $n_{T_3} = 16$ trees which were present at T_3. Note that a low rank means that the predicted sizes $m_{i4}^{(1)}, \ldots, m_{i4}^{(999)}$ generally are larger than the actual size m_{i4}.

As we can see most trees rank between the 8th and the 11th percentile, which leads us to accepting the fit in the case of a specific tree at the significance level $(rank + 1)/1000$, in all cases except for $i = 4$ and $i = 15$ (if simultaneous testing is done the significance level must be adjusted for multiple testing). Furthermore, by considering also Table 9 we see that two trees tend to always stick out. An explanation for this could be that these trees have small sizes and are neighbours of large trees, and since there are indications that the estimated growth is strong, it is likely that these predicted trees are killed by their predicted large neighbours.

In Table 10 we have summarized the values and the ranks of the observed summary statistics, as well as the estimated means and standard deviations of the predicted summary statistics.

We see that for the total basal area TBA we cannot reject the hypothesis

that the process which has generated the TBA of the forest at time T_4 and the process which has generated the TBA of the predictions (the GI-process) are the same. Furthermore, as was expected, due to the radius distributions and the over-prediction of the radii, the fit of the model cannot be accepted w.r.t. any of WMD or $\bar{m}$. From the quick comparison of the expected number of alive trees and the observed number of alive trees shown in Table 8, we saw that the two did not deviate too much from each other. However, we here have clear indications that the number of alive trees suggested by the model is lower than its observed counterpart. Hence, either α is underestimated or μ is overestimated, and we note that this can be redeemed by including more sample (inventory) times in the estimation of α and μ.

6 Discussion

In this paper, by using the open growth data set, a linear relationship was found between the site productivity index and the large tree sizes, which provides a good approach estimating the carrying capacity. The performances of some open growth models were compared, and it was shown that the Richards or Weibull growth functions better capture the open growth behaviour for Scots pines, than the logistic growth function used in the earlier studies. Evaluation of our growth-interaction process for spatio-temporal modelling of forest stands based on the space-time data was conducted. The preliminary results indicated that, based on the spatial structure, the point patterns and marks can be generated by the estimated GI-process, according the L-function and the mark-correlation function tests. The forest stand characteristics were also evaluated. The predicted total basal area per hectare performed reasonably well, but not the mean diameter or the basal area weighted mean diameter. Both of them were overestimated compared to the observed values. The number of stems in the plots for which we have made predictions were underestimated, except for one plot. For unknown reasons the observed mortality (death rate) between the third and the fourth inventory times was unusually high on that plot.

Since the estimates of the GI-model parameters are dependent on the expected tree population at a given time point, a starting point in our continuing work will be to further investigate the arrival and death process. Hereby, it is also required that we have more frequently sampled data. We have chosen to model the tree population with a diameter at breast height (dbh) equal to or larger than 10 cm because we have complete spatial information of these trees. Note that today's forestry focuses on low death rate and rather regular spatial

arrangement of established seedlings, which differs from naturally regenerated forests. Additional information about potential arrivals (i.e. trees ¡ 10 cm dbh) at a given time point would probably gain the model evaluation and development. The modelling data in the present study are all rather young at the first inventory time with a large amount of new arrivals exceeding the threshold diameter (10 cm dbh). To better describe and capture the development of pine stands during the whole life cycle, the data must be supplemented with older stands.

References

[1] 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.

[2] Cronie, O. (2010). Some edge correction methods for marked spatio-temporal point process models. Preprint 2010:9, Mathematical Sciences - Chalmers University of Technology and University of Gothenburg. www.math.chalmers.se/Math/Research/Preprints/2010/9.pdf

[3] Cronie, O., Särkkä, A. (2011). Some edge correction methods for marked spatio-temporal point process models. *Computational Statistics & Data Analysis* **55**, 2209–2220.

[4] Cronie, O., Yu, J. (2010). Maximum likelihood estimation in a discretely observed immigration-death process. *Research Report 2010:1, Centre of Biostochastics, Swedish University of Agricultural Sciences.*

[5] Diggle, P. (2001). *Statistical Analysis of Spatial Point Patterns.* Second ed. Oxford university press.

[6] Gerrard, D.J. (1969). Competition quotient - a new measure of the competition affecting individual forest trees. *Mich. State Univ. Agr. Exp. Stn. Res. Bull.* **20**.

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

[8] Grimmett, G., Stirzaker, D. (2001). *Probability and Random Processes.* Third ed. Oxford University Press.

[9] Hägglund, B., Lundmark, J.-E. (1977). Site index estimation by means of site properties - Scots pine and Norway spruce in Sweden. *Studia Forestalia Suecica* **138**, 38 pp. ISSN 0039-3150.

[10] Hope, A.C.A. (1968). A simplified monte carlo significance test procedure. *J. Roy. Statist. Soc. B* **30**, 582–598.

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

[12] Lei, Y.C., Zhang, S.Y. (2004). Features and partial derivatives of Bertalanffy-Richards growth model in forestry. *Nonlinear Analysis: Modelling and Control* **9**, 65-Ű73.

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

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

[15] Pukkala, T., Kolström, T., Miinaa, J. (1994). A method for predicting tree dimensions in Scots pine and Norway spruce stands. *Forest Ecology and Management* **65**, 123–134.

[16] Ranneby, B., Cruse, T., Hägglund, B., Jonasson, H., Swärd, J. (1987). Designing a new national forest survey for Sweden. *Studia Forestalia Suecica* **177**, 29 pp.

[17] Renshaw, E. (1994). The Linear Spatial-Temporal Interaction Process and Its Relation to $1/\omega$-noise. *J. Roy. Statist. Soc. B* **56**, 75–91.

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

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

[20] Seber, G., Wild, C. (1989). *Nonlinear Regression*. Wiley.

[21] Stoyan, D., Kendall, W., Mecke, J. (1995). *Stochastic Geometry and Its Applications*. Second ed. John Wiley & sons.

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

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

[24] Söderberg, U. (1997). Country report for Sweden. In Päininen, R. & Köhl, M. (eds.) *Study on European forestry information and communication system. Reports on forestry inventory and survey systems.* Volume 2. Office for official publications of the European Communities. Luxembourg. pp. 955-1017. ISBN 92-827-9846-1.

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

[26] Zeide, B. (1993). Analysis of growth equations. *Forest Science* **39**, 594–616.