Statistical inference and perfect simulation for point processes observed with noise

Ph.D. Thesis

Jens Lund

**The Royal Veterinary and Agricultural University
Copenhagen**

December 1999

Preface

This thesis is submitted in partial fulfillment of the requirements for the Ph.D. degree at the Royal Veterinary & Agricultural University, Copenhagen. The work has been done from December 1996 to December 1999 at the Department of Mathematics and Physics under supervision of professor Mats Rudemo.

The most recent version of the papers in this thesis can always be obtained from my homepage `http://www.dina.kvl.dk/~jlund`.

It is usual to thank a lengthy list of persons in the preface of your thesis. I could certainly make such a list as I know and I have met a lot of wonderful, nice, and helpful persons during my Ph.D. studies. However, I am afraid that I would forget some of you, so I prefer to simply say THANK YOU without mentioning anybody. Your support, advice, and encouragement has been invaluable to me. I owe you all a great debt of gratitude.

Jens Lund
Frederiksberg, December 1999.

Address: The Royal Veterinary and Agricultural University
Department of Mathematics and Physics
Thorvaldsensvej 40
DK – 1871 Frederiksberg C
Denmark

e-mail: `jlund@dina.kvl.dk`
homepage: `http://www.dina.kvl.dk/~jlund`
fax: +45 35 28 23 50
phone: +45 35 28 22 89

Summary

The main themes of this thesis are spatial statistics and simulation algorithms. The thesis is split into five papers that may be read independently. All five papers deal with spatial models. Lund & Rudemo (1999), Lund et al. (1999), and Lund & Thönnes (1999*b*) deal with the same new model for point processes observed with noise, and Lund et al. (1999), Lund & Thönnes (1999*b*), and Lund & Thönnes (1999*a*) has a simulation aspect.

Lund & Rudemo (1999), Lund et al. (1999), and Lund & Thönnes (1999*b*) develop and analyse a new model for point processes observed with noise. Usually the analysis of spatial point patterns assume that the observed points (the true points) are a realization from a specific model. In contrast our approach is to assume the observed pattern generated by thinning and displacement of the true points, and allow for contamination by points not belonging to the true pattern.

Lund & Rudemo (1999) develop the model for point processes observed with noise. The likelihood function for an observation of a noise corrupted point pattern given the true positions is derived. As data for our analysis is indeed a realization of the underlying true process and its associated noise corrupted point pattern we need not consider a model for the underlying process. The parameters in the model describe how many of the true points are lost, how large the displacements are, and the number of contaminating surplus points. For estimation of the parameters in the noise model a deterministic, iterative, and approximative maximum likelihood estimation algorithm is developed. The likelihood function is a sum of an excessive large number of terms, and the algorithm works by finding large dominating terms. Alternative estimation methods are discussed.

Lund et al. (1999) analyse the model developed in Lund & Rudemo (1999) with respect to the now unobserved true points. We assume a noisy observation of a true point pattern and knowledge of the parameters in the model. A Bayesian point of view is now adopted and we specify a prior distribution for the underlying true process. Given the model, the prior distribution, and the noisy observation, we get the posterior distribution of the true points. This posterior distribution is investigated by samples from the distribution. These samples are obtained from a Markov chain Monte Carlo (MCMC) algorithm extending the Metropolis-Hastings sampler for point processes. A thorough discussion is provided on the choice of prior distribution and how to present the samples from the MCMC runs. The MCMC samples are used to estimate for example the K-function for the unobserved true point pattern. These estimates are clearly better than estimates based on the observed points alone.

The use of the MCMC algorithm in Lund et al. (1999) relies on the fact that a Markov chain run for a long time approaches its stationary distribution. Lund & Thönnes (1999*b*) uses a recent technique called Coupling From The Past (CFTP) to deliver a sample drawn from the exact posterior distribution of the unobserved true points described in Lund et al. (1999), a so-called perfect simulation. This perfect simulation algorithm is based on spatial birth-and-death processes for simulation of point processes. In order to apply CFTP in our problem the simulation is carried out on an augmented state space. The algorithm turns out to be too slow in practice and thus demonstrates possible current limits of CFTP.

Lund & Thönnes (1999*a*) describes a new perfect simulation algorithm for general locally stable point processes. The algorithm is based on CFTP for Metropolis-Hastings simulation of point processes and it is simpler than the previous known perfect algorithm based on Metropolis-Hastings simulation for this class of models. The present state of the algorithm is that it is far too slow to be useful in practice, and it might have some theoretical flaws. These problems are discussed in the introductory part of the thesis.

Lund (1998) develops a model for survival times of trees that take the spatial positions of the trees into account. At a finite number of timepoints it is observed whether a tree is alive or not, and thus we have interval censoring of the even aged trees. The model is a discrete time version of Cox's proportional hazards model. Positions of trees are considered as fixed, and they are used to compute competition indices that enter the model as covariates. It is shown that small trees have a higher risk of dying than large tress and the area of the experiment is inhomogeneous. In addition, Hegyi's competition index based on basal area is a significant covariate.

Dansk resumé

Hovedemnerne for denne afhandling er rumlig statistik og simuleringsalgoritmer. Afhandlingen består af fem artikler, der kan læses uafhængigt. Alle fem artikler omhandler rumlig statistik. Lund & Rudemo (1999), Lund et al. (1999) og Lund & Thönnes (1999*b*) omhandler den samme nye model for punktprocesser observeret med støj, mens Lund et al. (1999), Lund & Thönnes (1999*b*), samt Lund & Thönnes (1999*a*) indeholder simuleringsaspekter.

Lund & Rudemo (1999), Lund et al. (1999) og Lund & Thönnes (1999*b*) udvikler og analyserer en ny model for punktprocesser observeret med støj. Normalt antages i analysen af rumlige punktprocesser at de observerede punkter (de sande punkter) er en realisation fra en bestemt model. Vi antager, at de observerede punkter er fremkommet ved tynding og flytning af sande punkter, samt at punkter der ikke tilhører det sande punktmønster kan være tilføjet.

Lund & Rudemo (1999) beskriver modellen for punktprocesser observeret med støj, og likelihoodfunktionen for en støjfyldt observation, givet det sande punktmønster, udledes. Data for analysen er en realisation af et sandt punktmønster og en tilhørende støjfyldt observation, hvorfor en model for de sande punkter ikke er nødvendig. Parametrene i modellen beskriver sandsynligheden for at miste et sandt punkt, størrelsen af flytningerne og antallet af ekstra punkter. Til estimation af parametrene i støjmodellen udvikles en deterministisk, iterativ, og approksimativ maksimum likelihood estimationsalgoritme. Likelihoodfunktionen er en sum af et meget stort antal led og algoritmen finder de største af disse. Alternative estimationsmetoder diskuteres.

Lund et al. (1999) analyserer modellen udviklet i Lund & Rudemo (1999) mht. til de nu uobserverede sande punkter. Vi har en støjfyldt observation af det sande punktmønster og kendskab til parametrene i modellen. Fra en Bayesiansk synsvinkel specificeres nu en *a priori* fordeling for den underliggende sande process. Med modellen, *a priori* fordelingen, og den støjfyldte observation får vi *a posteriori* fordelingen for de sande punkter. Denne *a posteriori* fordeling beskrives med simulerede udfald fra fordelingen. Udfaldene fås fra en Markovkæde med *a posteriori* fordelingen som stationær fordeling og den er designet som en udvidelse af Metropolis-Hastings algoritmen for punktprocesser. Der er en grundig diskussion af valget af *a priori* fordeling og hvordan udfaldene fra *a posteriori* fordelingen bedst præsenteres. Udfaldene bruges bl.a. til estimation af K-funktionen for det uobserverede sande punktmønster. Disse estimater er klart bedre end estimater baseret alene på den støjfyldte observation.

Markovkæde-algoritmen i Lund et al. (1999) benytter at en Markovkæde udviklet i mange trin nærmer sig sin stationære fordeling. Lund & Thönnes (1999*b*) bruger den nye teknik *kobling fra fortiden* (Coupling From The Past, CFTP) til at generere udfald fra *a posteriori* fordelingen af de uobserverede punkter uden approksimationer, en såkaldt perfekt simulering. Denne perfekte simuleringsalgoritme er baseret på rumlige fødsels- og dødsprocesser for punktprocesser. For at bruge kobling fra fortiden i vores problem udføres simulationen på et udvidet tilstandsrum. Algoritmen viser sig at være for langsom i praksis og demonstrerer således mulige nuværende begrænsninger ved kobling fra fortiden.

Lund & Thönnes (1999*a*) udvikler en ny perfekt simuleringsalgoritme for generelle lokalt stabile punktprocesser. Algoritmen er baseret på kobling fra fortiden af Metropolis-Hastings simulering af punktprocesser og er simplere end den tidligere kendte tilsvarende algoritme. For nuværende er algoritmen alt for langsom i praksis og den kan have teoretiske fejl. Disse problemer er diskuteret i den indledende del af afhandlingen.

Lund (1998) udvikler en model for levetider af træer der tager højde for de rumlige positioner af træerne. På et endeligt antal tidspunkter er det observeret, om et givent træ er i live eller ej, og vi har således intervalcensur af de ensaldrende træer. Modellen er en diskret-tids version af Cox's model for proportionale intensiteter. Positionerne af træerne opfattes som givne og de bruges til beregning af konkurrenceindex, der indgår som kovariater i modellen. Det vises at små træer har en højere dødelighed end store træer, og at forsøgsområdet er inhomogent. Hegyi's konkurrenceindex baseret på grundfladen viser sig at være en signifikant kovariat.

Contents

1 Introduction

This thesis contains the introductory sections 1–3, the four papers Lund & Rudemo (1999), Lund et al. (1999), Lund & Thönnes (1999*b*), Lund (1998), as well as the preliminary manuscript Lund & Thönnes (1999*a*). The papers consitute the main part of the thesis. In this Section 1 we introduce the background for the thesis and give brief comments on the content of the papers and their relation to Section 2 and Section 3.

1.1 A bit of history

This project started as a continuation of the Ph.D. thesis Dralle (1997) (especially Dralle & Rudemo (1996) and Dralle & Rudemo (1997)). The fundamental idea in the project and the work of Larsen (1997) and Larsen & Rudemo (1998) is to identify tree tops in aerial photos of Norway spruce.

In forestry there is a wish to get information about the number and volume of trees in the forest, as well as the positions of the trees. This will help management and planning of the forest. So far, the expenses in terms of working hours and money to obtain this information has prevented a detailed inventory. Analysis of aerial photos seem to promise a less expensive way to get detailed information on the individual tree. Further background information can be found in Dralle (1997). The data from the forest experiments are introduced in Section 2.1.1.

The image analysis leads to an observation of the tree top positions. However, the observation is not exact as some trees tops are for example lost or moved. The work in this thesis is concentrated on the analysis of such point processes with noise. Two of the four papers (Lund & Rudemo (1999), Lund et al. (1999)) deal with a model for point processes with noise. The model is introduced in Section 2.2. The third paper Lund & Thönnes (1999*b*) does also consider this model, but is primarily interested in perfect simulation from the posterior model used in Lund et al. (1999).

Because these methods seem to promise cheap and detailed data from the individual tree it is natural to collect data from the same area over time. In order to use these data we need models for the development of trees that take their positions into account. Lund (1998) is one step in this direction as it describes a model for survival times of trees. The analysis uses the positions of trees through a competition index.

1.2 Overview of thesis

We now comment briefly on the papers, their status, their relation to each other, and the connection with the introductory part of this thesis that may contain further remarks. Section 2 and 3 contain introductory material for the papers and elaborate on possible extensions etc. In order to understand the comments on possible extensions it is probably necessary to have read the papers first. While reading Section 2 and Section 3 it should be quite clear which parts are ment to be read after the papers.

Paper 1 Jens Lund and Mats Rudemo (1999), *Models for point processes observed with noise*, accepted for publication in Biometrika.

This paper develops a model for point processes observed with noise. The focus is on the likelihood function for the observation and maximum likelihood estimation of the parameters in the model when both the true points and the noisy points are observed. Section 2 contains an introduction to the data set considered in this article and describes the basic model.

Paper 2 Jens Lund, Antti Penttinen, and Mats Rudemo (1999), *Bayesian analysis of spatial point patterns from noisy observations*, submitted.

This paper is a natural extension of Lund & Rudemo (1999), and we describe a method to analyse the unobserved true points given a noisy observation. We adopt a Bayesian point of view and put a prior distribution on the true points. The prior distribution expresses knowledge about the regularity of the true points. The posterior distribution is then analysed by dependent samples originating from a Markov chain Monte Carlo algorithm.

Paper 3 Jens Lund and Elke Thönnes (1999), *Perfect simulation of point patterns from noisy observations*, manuscript.

This manuscript considers perfect simulation of the posterior distribution in Lund et al. (1999). The simulation algorithm used in Paper 2 is approximative in the sense that we are not assured of the convergence of the Markov chain. The algorithm considered in this paper is based on the new idea of perfect simulation (Propp & Wilson, 1996), and guarantees that the output sample has exactly the correct distribution. Section 3 on page 9 contains further references to perfect simulation.

Unfortunately, the perfect algorithm turns out to be too slow with the prior distributions used in Lund et al. (1999). It is our hope that further investigation will lead to possible ways to speed up the algorithm.

Paper 4 Jens Lund (1999), *Survival of the Fattest? Self-thinning among Trees*, course report.

A model for survival times of Sitka spruce based on a discrete time version of a Cox model is suggested. The positions af all the trees are known, and a competition index based on neighbours of a tree is used as a covariate to take the spatial distribution of trees into account. This paper assumes detailed data on a tree level, that may be obtained by the same procedure that leads to the data considered in the previous three papers.

This paper is a little different from the three previous in style. The paper is originally a course report from one of my Ph.D. courses, but the plan is to rewrite it into an article.

Preliminary manuscript Jens Lund and Elke Thönnes (1999), *Perfect adaptive Metropolis-Hastings Simulation for Point Processes.*

This manuscript is on perfect simulation of locally stable point processes based on the Metropolis-Hastings algorithm for simulation of point processes. Very shortly before I turned in the thesis some problems arised. These problems are discussed in Section 3.4 of this introductory part and at present it seems to be an open question whether the idea presented in this paper does in fact work or not. Even if the idea as presented does turn out not to work, then parts of it might still be used in other contexts.

2 A model for point patterns observed with noise

This section introduces the model for point processes observed with noise considered in the papers Lund & Rudemo (1999), Lund et al. (1999), and Lund & Thönnes (1999*b*).

2.1 Examples of point processes observed with noise

We introduce examples of point processes observed with noise. Section 2.1.1 describes the dataset used in Lund & Rudemo (1999), Lund et al. (1999), and Lund & Thönnes (1999*b*), whereas Section 2.1.2 briefly describes other examples.

2.1.1 Aerial photos of Norway spruce

The upper part of Figure 1 shows an aerial photo from a flight 560m above a thinning experiment in Norway spruce. The goal is to identify the positions of tree tops in the image.

Larsen (1997) and Larsen & Rudemo (1998) developed a template model for one tree taking into account the positions of the camera and light sources. The resulting template, shown in the right part of Figure 2 bounded by an ellipse, was moved pixelwise over the image. Local maxima of the correlation between template and image pixel grey levels were considered as candidate positions of tree tops. The left part of Figure 2 sketches the model for light reflection within a tree.

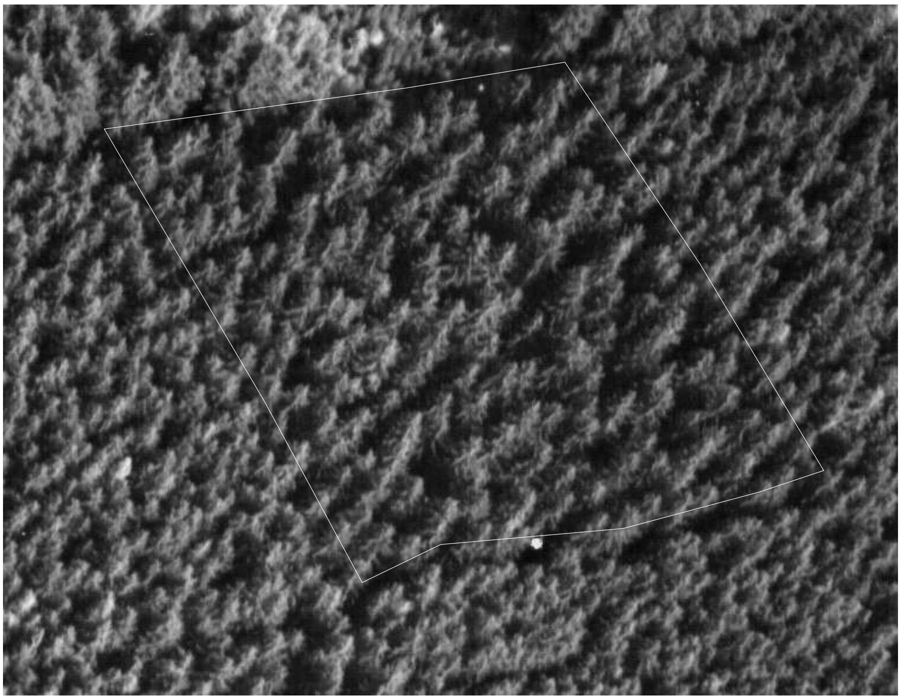
The lower part of Figure 1 shows a map of the image in the upper part. Both the true positions of tree tops (denoted by circles) and the estimated positions as found by the template matching method (denoted by dots) are shown. Let $X = \{X_i : i \in M\}$, $M = \{1, \ldots, m\}$, denote the positions of the true points and let $Y = \{Y_j : j \in N\}$, $N = \{1, \ldots, n\}$ denote the positions found by the template matching method. We want to estimate the true positions X but do instead observe Y. The template method in Larsen & Rudemo (1998) gives 570 candidate positions for tree tops, but only the 206 best candidates are used here. See Lund & Rudemo (1999) for more details.

2.1.2 Other examples

Point processes with noisy observations arise from many other applications. For instance, any time the observation is made in an indirect way it must be assumed to be noisy. This can be in image analysis applications as above. Another example could be the following: We want to observe the positions of the nest of a certain animal. Instead of the nests themselves we do observe the positions of the animal coming and going from the nest, or traces of the animal like feathers or stools. The Global Positioning System (GPS) is also known to have noise in its observations — this noise is even deliberately larger than technical possible because the military uses this system too.

2.2 The basic model

As noted above, the X and Y point sets are not identical. We can consider Y as a disturbed observation of the true points X, and will now propose a model for the observation of Y (Lund & Rudemo, 1999). Consider X and Y as point processes on a bounded subset A of $\mathbb{R}^d$.



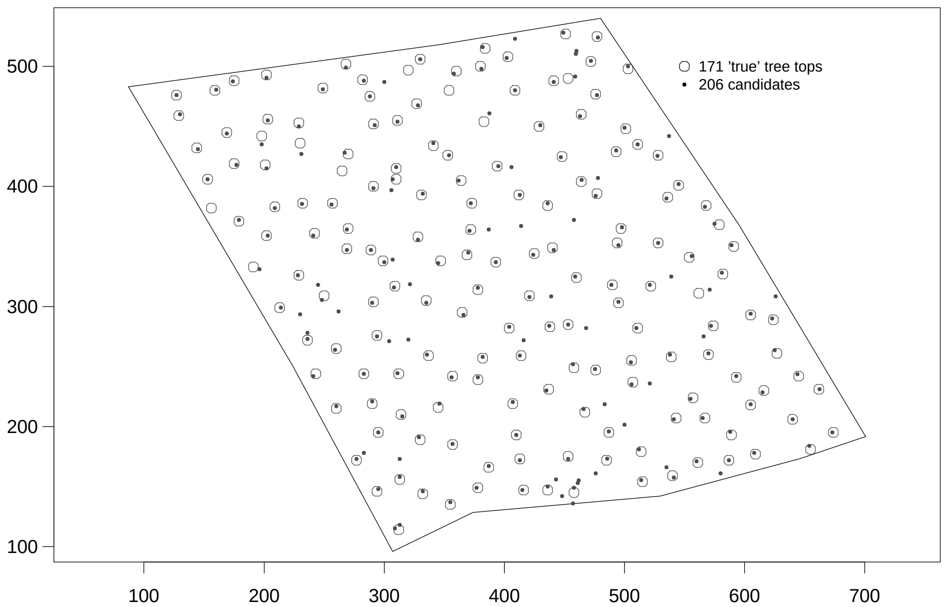


Figure 1: Image with sidelighted trees, and, in the lower part, 171 X-points (centres of circles) corresponding to 'true' tree tops and 206 Y-points (dots) corresponding to template matching. The area of the delineated subplot is 4 454 m^2, and the unit of the axes in the lower part is linear pixel size, 0.15 m.

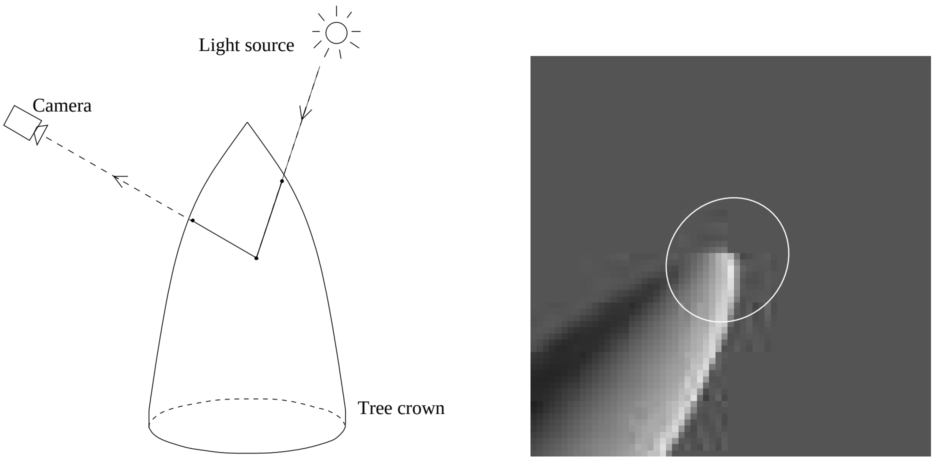


Figure 2: Model tree and, in the right part, template with optimal bounding ellipse.

Suppose that Y is generated from the X-process by the following disturbance mechanisms:

(i) *Thinning.* Each point $X_i, i \in M$, is thinned with probability $1 - p(X_i)$ and retained with probability $p(X_i)$. If an X-point is thinned, then there will not be any corresponding Y-point.

(ii) *Displacement.* For each remaining point X_i a corresponding Y-point is generated by displacement to a position with probability density $k(\,\cdot\,|X_i)$ with respect to the Lebesgue measure on $\mathbb{R}^d$.

(iii) *Censoring.* The displaced points are observed if they are within an observation area A; otherwise they are censored and not observed. Thus censoring of an unthinned point generated by X_i occurs with probability $\int_{A^c} k(y|X_i)\,dy$. (Here A^c denotes the complement $\mathbb{R}^d \setminus A$ of the set A.) Censoring is mainly a technical problem as $\int_{A^c} k(y|X_i)\,dy \approx 0$ in most applications.

(iv) *Superposition of ghost points.* In addition to the points generated as described above we have superposition of extra 'ghost' points. These points are assumed to arise from a Poisson process on A with intensity $g(\,\cdot\,|X)$, where X, as above, denotes the entire X-process including thinned points.

We assume suitable independence relations between the stochastic elements. The points generated from X by the combination of thinning, displacement, censoring and superposition form the Y-process, which is thus restricted to the set A.

So far I have used a homogeneous thinning probability, $p(x) = p$, a normal distribution with mean $X_i + \mu$ and covariance Σ as the displacement distribution, and homogeneous Poisson noise $g(\,\cdot\,|X) = \lambda$. Thus, the parameter vector is $\theta = (p, \lambda, \mu, \Sigma)$.

Let $|B|$ denote the number of elements in a finite set B and let $|B|_d$ denote the d-dimensional volume of B when B is a measurable subset of $\mathbb{R}^d$. For two finite sets M_1 and N_1 with $|M_1| = |N_1|$ we let $\mathcal{P}(M_1, N_1)$ denote the set of all bijections π from M_1 to N_1. Introduce $S_{m,n} = \{(M_1, N_1, \pi) : M_1 \subseteq M = \{1, \ldots, m\},\ N_1 \subseteq N = \{1, \ldots, n\},\ |M_1| = |N_1|,\ \pi \in \mathcal{P}(M_1, N_1)\}$ for $m, n \geq 0$. In case $m = 0$ $(n = 0)$

we let $M = \emptyset$ ($N = \emptyset$) in the definition. Further, define $S = \cup_{m=0}^{\infty} \cup_{n=0}^{\infty} S_{m,n}$, and provide S with the σ-algebra consisting of all subsets. We let $s = (M_1, N_1, \pi) \in S$ specify the correspondence between X- and Y-points and we call s a 'matching' with the interpretation that $(X_i, i \in M_1)$ and $(Y_j, j \in N_1)$ consist of matched points. More precisely, $Y_{\pi(i)}$ is matched to X_i for $i \in M_1$, that is, $Y_{\pi(i)}$ is obtained from X_i by displacement.

Define

$$T(X, Y, \theta, s) = L_1\, L_2\, L_3\, L_4$$

for $s = (M_1, N_1, \pi)$ where

$$\begin{aligned} L_1 &= p^{|M_1|} \prod_{i \in M_1} k(Y_{\pi(i)} | X_i, \mu, \Sigma), \\ L_2 &= \prod_{i \in M \setminus M_1} \left(p \int_{A^c} k(y | X_i, \mu, \Sigma)\, dy + 1 - p \right), \\ L_3 &= \lambda^{|N \setminus N_1|} \exp((1 - \lambda)|A|_d), \\ L_4 &= 1\big(s \in S_{|X|,|Y|}\big)\,. \end{aligned}$$

Here L_1 corresponds to unthinned points displaced within A, L_2 to thinned points and unthinned points displaced out of A, and L_3 to superpositioned 'ghost' points. We have used the notation $1(\cdot)$ for the indicator function.

Theorem 1 in Lund & Rudemo (1999) states that the likelihood for (X, θ) based on the observation Y is

$$L(Y|X, \theta) = \sum_{s \in S} T(X, Y, \theta, s), \tag{1}$$

with the reference measure taken as the Poisson process with intensity 1 on the set A. One should note that the sum typically contains a very large number of non-zero terms but the number of leading terms may be small and thus approximations are possible. We also consider the joint likelihood of the point processes Y and the matching s given by

$$L(Y, s|X, \theta) = T(X, Y, \theta, s), \tag{2}$$

with the reference measure taken as the product of the measure of a Poisson process on A with intensity 1 and the counting measure on S. The relation between (1) and (2) is that (1) is the marginal distribution of the point process Y in (2). The matching s is unobserved and thus (2) could be considered as a missing data model (Smith & Roberts, 1993, Sec. 6). This approach is used in Lund et al. (1999) and could be used in Lund & Rudemo (1999) in alternative estimation algorithms.

A natural extension of the above setup is to associate a mark with each observed point Y_j. The mark should express knowledge about the belief in the point at that mark. In the analysis of the image data considered in Lund & Rudemo (1999) the mark could be the correlation between the template and the image. Currently this information is ignored, but the analysis could be improved when such a mark is available.

2.3 Interesting questions

The basic questions of estimation of the parameter θ and reconstruction of X are treated in Lund & Rudemo (1999) and Lund et al. (1999), respectively, and are introduced in the following sections 2.3.1 and 2.3.2.

2.3.1 Estimation of parameters

Estimation of the parameter θ in the model (1) is difficult due to the very large number of terms in in the sum. The sum is over all possible ways to pair X and Y points. We estimate the parameters in the model in case both X and Y is observed. Further, we suggest an approximate likelihood analysis based on the concept of "neighbours" to a matching s. A neighbour of a matching s is a matching s' which is very similar to s. The crucial issue in the approximate likelihood computation is to find matches $s = (M_1, N_1, \pi)$ such that the corresponding terms give large contributions to (1), and then focus on only a small number of terms. This is achieved by a deterministically, iterative algorithm consisting of a starting procedure for finding an initial set of matches and local maximizations over suitably chosen neighbourhoods of matches until no further improvement is obtained.

2.3.2 Reconstruction of a disturbed point pattern

The idea is to use the model (1) to reconstruct X when the parameter $\theta = (p, \lambda, \mu, \Sigma)$ is known and we have a noisy observation Y. The knowledge of the parameter can e.g. be obtained through training data sets observed under similar conditions as the current data.

Similar problems are found in Baddeley & van Lieshout (1993) which consider a point process description of an image and cluster center estimation in a clustered point pattern. Cressie & Lawson (1998) and Dasgupta & Raftery (1998) are about detection of mines in minefields and are also similar.

We use a Bayesian approach and have a prior distribution $L(X)$ on X. The prior distribution gives regular point patterns, and it expresses that trees in planted and managed forests (as our images) tend to be placed regularly in the area.

The posterior distribution for the true points is $L(X|Y, \theta) \propto L(Y|X, \theta)L(X)$. We explore this posterior distribution by Markov chain Monte Carlo (MCMC) samples from the distribution. We make a sampler of Metropolis-Hasting type along the lines in Geyer & Møller (1994), Geyer (1999), and Green (1995). In order to avoid the huge sum in (1) we consider the matching s as missing information here and we simulate both the true point positions X and the unobserved matching s. The expression for the posterior distribution is $L(X, s|Y, \theta) \propto L(Y, s|X, \theta)L(X)$ which is easy to compute.

2.4 Extensions to several observations of a point process

An extension of Section 2.3.2 and Lund et al. (1999) is the situation with several independent noisy observations of the same point pattern. The likelihood is easily found by multiplying the likelihoods for the individual observations, but the MCMC algorithm needs to be extended compared to Lund et al. (1999). A further complication is that points belonging to different observations should also somehow be matched.

We could improve the analysis of the forest data described in Section 2.1.1 by including several images taken from different postions compared to the ground and the sun. Then it is natural to estimate the true positions $X = \{X_1, \ldots, X_m\}$ in three-dimensional real world coordinates, and a problem is that each image just has a two-dimensional observation of the tree tops. Instead of extending the above model for point processes observed with noise to this setup I outline another way to do this analysis. It combines the image analysis and point process analysis steps into one analysis. This has the advantage of utilizing the information in the images better, but will on the other hand be more tailored to this specific problem. The analysis combines ingredients from the above analysis as well as Baddeley & van Lieshout (1993) and van Lieshout (1994).

Let $X = \{X_1, \ldots, X_m\}$ be the three-dimensional coordinates of the tree tops to be estimated and assume that we have observed q images $Z_1, \ldots, Z_q$ with independent noise. For each image Z_k there is a function f_k that transforms a three-dimensional point X_i into two-dimensional image coordinates $f_k(X_i)$.

For each of the q images we make an artificial image given the X points. On top of each of the positions $f_k(X_1), \ldots, f_k(X_m)$ we put a template for the image like the one in Figure 2. The template models the mean image around a tree top. At this stage we must decide how to model the mean image outside the templates and what to do if the templates overlap because the tree tops are close together. One solution could be to use a constant mean image outside the templates and favour the template for the tree closest to the camera.

Based on these "mean images" we can write a likelihood $L(Z_k|X)$ for each image. It would be preferable if this likelihood allowed for the possibility of loosing a tree in the image by a certain probability, say $1 - p_k$, and allowed for a small movement of the template at position $f_k(X_i)$. The movement is necessary because the f_k functions are fixed and the X_i positions are common for all the images so some disturbance should be allowed. We multiply the likelihoods for each image to obtain a likelihood $L(Z|X) = \prod_{k=1}^{q} L(Z_k|X)$ for the sequence of images.

We could now proceed similar to the analysis in Section 2.3.2. Put a prior $L(X)$ on the three-dimensional X-coordinates and investigate the posterior distribution $L(X|Z) \propto L(Z|X)L(X)$ of the positions given the images. Some kind of Markov chain sampling from the positions will probably be involved here.

I would expect this idea to work reasonable well with 3–4 images. However, the above description is just an outline and quite some details must be filled in for this idea to work smoothly.

2.5 Extensions to development in time

The data collection method described in the previous sections is easily applied to the same area over time. This give us detailed information on the development of each individual tree. If we have obtained the positions of the tree tops in real world three-dimensional coordinates as described in Section 2.4 we get information on the ground position of each tree and the growth of the tree between measurements. This increases the use of models for the competition between trees that take their positions into account. Lund (1998) describe one such model for the survival times of Sitka spruce.

3 Perfect simulation of point processes

3.1 Introduction

Propp & Wilson (1996) started a new area of probability theory called *perfect simulation*. The basic goal is to simulate from a target distribution π by designing a Markov chain X_t that has π as its stationary distribution and then monitor the convergence of the chain. Methods for designing Markov chains with π as its stationary distribution have existed for a long time, see e.g. Gilks et al. (1996) for an overview. Propp & Wilson (1996) contributed a new method of monitoring convergence of the Markov chain such that the output is known to come from π exactly. Previously the Markov chain X_t was run for a long time until it was believed to be sufficiently close to stationarity. Even though theoretical results on convergence are possible, for instance for the Metropolis-Hastings sampler for points processes (Geyer & Møller, 1994), they may be of little practical use (Geyer, 1999, Section 3.8). An introduction to this new method called Coupling From The Past (CFTP) can be found in Lund & Thönnes (1999*b*, Section 4) in this thesis. The recent overview papers Thönnes (1999) and Dimakos (1999) give a more thorough introduction than contained in Lund & Thönnes (1999*b*). A full annotated and updated bibliography on the area of perfect simulation can be found on David Wilsons homepage, `http://dimacs.rutgers.edu/~dbwilson/exact.html/`.

3.2 Perfect simulation papers in this thesis

This thesis contains two manuscripts on perfect simulation. Lund & Thönnes (1999*a*) tries to develop a new perfect simulation algorithm for locally stable point processes based on Metropolis-Hastings simulation of point processes (Geyer & Møller, 1994). This algorithm would be conceptually simpler than the algorithm described in Kendall & Møller (1999). However, some open questions still remain unsolved as discussed in the following Section 3.4. The paper Lund & Thönnes (1999*b*) develops a perfect simulation algorithm for the posterior distribution of the true unobserved points in the reconstruction problem in Section 2.3.2.

Perfect simulation of locally stable point processes is considered in Kendall & Møller (1999). They describe two methods based on spatial birth-and-death processes and Metropolis-Hastings simulation, respectively — the former being much simpler than the latter.

The perfect simulation algorithm in Lund & Thönnes (1999*b*) is based on spatial birth-and-death processes whereas the non-perfect algorithm used in Lund et al. (1999) is based on Metropolis-Hastings simulation. One difference between the two approaches, apart from the perfect simulation aspect, is the following. To use the Metropolis-Hastings simulation algorithm we must be able to compute $\int_{A^c} k(y|X_i,\mu,\Sigma)\,dy$ whereas the spatial birth-and-death algorithm just requires us to be able to make a coin flip that has probability $\int_{A^c} k(y|X_i,\mu,\Sigma)\,dy$ of heads. In the Metropolis-Hastings setting the value of $\int_{A^c} k(y|X_i,\mu,\Sigma)\,dy$ is required to make a coin flip with probability $\alpha = \min(1, \int_{A^c} k(y|X_i,\mu,\Sigma)\,dy \times \text{a factor})$ of heads and it is impossible to do the α-coin flip without knowing the value of $\int_{A^c} k(y|X_i,\mu,\Sigma)\,dy$. The approximation $\int_{A^c} k(y|X_i,\mu,\Sigma)\,dy = 0$ is used in Lund et al. (1999) to avoid the calculation of the integral.

3.3 Rejection sampling

The rejection sampling method is a well-known method to generate samples from a target distribution π with unnormalized density f on a space Ω equipped with a suitable σ-algebra and a reference measure γ. We will repeat it here because it is simple, because we can demonstrate how to do rejection sampling for the problem in Lund & Thönnes (1999*b*), and because some of the perfect simulation algorithms are clever applications of rejection sampling (Fill, 1998; Fill et al., 1999).

Assume that we have another distribution μ with unnormalized density g on the same space, and that we are able to simulate easily from μ. Assume further that g dominates f such that $f(x) \leq g(x)$ for all $x \in \Omega$. The distribution μ is called the proposal distribution.

Rejection sampling:
Simulate X according to μ. Accept X with probability $f(X)/g(X)$. If X is not accepted, then simulate a new X and continue until the simulated X is accepted. The returned sample is distributed according to π.

It can easily be seen from the following simple lemma that this procedure returns a sample from the target distribution π.

Lemma *If X is distributed according to μ and U is distributed uniformly on $[0, g(X)]$ given X then (X, U) is uniformly distributed on $\{(x, u) \in \Omega \times [0, \infty[: 0 \leq u \leq g(x)\}$.*

If (X, U) is uniformly distributed on $\{(x, u) \in \Omega \times [0, \infty[: 0 \leq u \leq f(x)\}$ then X has marginal distribution π.

The uniform distribution on a subset of $\Omega \times [0, \infty[$ means the distribution with density equal to a constant times the indicator function of the set with respect to the product measure $\gamma \otimes l$ where l is the Lebesgue measue on $\mathbb{R}$.

With the notation from Lund & Thönnes (1999*b*) we can now describe how to generate a perfect sample from the posterior distribution of the true points given the noisy observation in the problem of Section 2.3.2. The proposal distribution is the stationary distribution of the dominating chain, described in Section 5.1 of the paper, conditioned to have at most one X-point matched to each Y-point. It is easy to sample from this distribution as it is basically a Poisson process and the result of the conditioning can be stated explicitly. This sample is then accepted by probability $L(X)/\,(\lambda^\star)^{n(X)}$.

The problem with rejection sampling in this particular example is that it is too slow to be useful, that is the number of rejections before acceptance is too high. In general another complication that prevents us from using rejection sampling is that we cannot find a suitable proposal distribution with a dominating density.

3.4 Convergence of the target process?

This section is intended to be read after Section 1 up to and including Section 5.2 in Lund & Thönnes (1999*a*) have been read. Shortly before submission of this thesis Jesper Møller pointed out to me our construction in Section 5 of Lund & Thönnes (1999*a*) had to be carefully checked. To be more precise, the convergence of X to the correct equilibrium distribution had to be proven rigorously. In order for the algorithm in Lund & Thönnes (1999*a*) to work, we need that the target chain X, coupled to the dominating chain Z, marginally converges in distribution to the target distribution π (Kendall & Møller, 1999, Theorem 3.1). At present there is no rigorous prove for the convergence of X. What follows is a discussion on a more abstract level than in Lund & Thönnes (1999*a*) of the requirements needed to ensure convergence of the target chain X defined in the paper. The reader should be able to follow this section without having to read the paper.

Let Ω be the state space of Z and X. Assume that Ω has a partial order $\preceq$, and has a state called $\mathbf{0}$ and a reference measure γ. The special state $\mathbf{0}$ is a minimal element in Ω, that is $\mathbf{0} \preceq z$ for any $z \in \Omega$.

In our setting Ω is the point process state-space, all finite subsets of a set $C \subseteq \mathbb{R}^2$, where C is bounded and Borel. The minimal state is the empty point configuration, $\mathbf{0} = \emptyset$. The reference measure γ is the Poisson process at unit rate and the partial ordering is the set inclusion. The transition kernels for Markov chains we use are Metropolis-Hastings kernels for the sampling of point processes (Geyer & Møller, 1994; Geyer, 1999).

Assume that $Z = (Z_1, Z_2, \dots)$ is a stationary, time-reversible, Markov chain with invariant distribution μ. Suppose Z_1 is distributed according to μ. We assume that Z evolves according to a transition kernel $Q_z(A) = \int_A q(z,y)\, d\gamma(y)$ which is aperiodic, irreducible, and positive Harris recurrent. The state $\mathbf{0} \in \Omega$ is an ergodic atom for Z, that is the probability of Z visiting $\mathbf{0}$ in the time interval $[0,t]$ tends to 1 as $t \to \infty$.

Let P^z be a family of Markov kernels on Ω. The family is indexed by a value $z \in \Omega$ and has density p, $P^z_x(B) = \int_B p^z(x,y)\, d\gamma(y)$. Suppose that for each fixed $z \in \Omega$ the kernel P^z defines an aperiodic, irreducible, and positive Harris recurrent Markov chain with invariant distribution π. Thus the n-step transition kernel based on P^z converges to π as $n \to \infty$ (Geyer, 1999).

Let $X = (X_1, X_2, \dots)$ be a process defined in the following way. Let $X_1 = x$ be a determistic starting state. In order to get from $X_t = x_t$ to X_{t+1} choose X_{t+1} according to the distribution $P^{Z_t}_{x_t}(\,\cdot\,)$.

In our specific setting we couple the update of X_t according to $P^{Z_t}_{x_t}(\,\cdot\,)$ to Z such that $X_{t+1} \preceq Z_{t+1}$.

Conjecture *The marginal distribution of X_t converges weakly to π as $t \to \infty$.*

We now give an outline of a proof with comments on the parts still to be shown.

Proof: Due to the dependence on Z the X chain is not a Markov chain. Instead we consider the joint distribution of Z and X and (Z_t, X_t) is a Markov chain.

Consider the Markov kernel $R_{z,x}(A \times B) = Q_z(A) P^z_x(B)$ which has the invariant distribution $\mu \otimes \pi$. In order for this to be the only invariant distribution we need irreducibility of $R_{z,x}$. However, the (Z,X) chain described above does *not* evolve according to $R_{z,x}$. We have a coupling of Z and X such that (Z,X) lives on $\tilde{\Omega}_2 = \{(z,x) \in \Omega \times \Omega : x \preceq z\}$. This (Z,X) chain evolves according to a kernel $\tilde{R}_{z,x}$ and the relation between $\tilde{R}_{z,x}$ and $R_{z,x}$ is that they agree on the marginal distributions. That is $\tilde{R}_{z,x}(A \times \Omega) = Q_z(A) = R_{z,x}(A \times \Omega)$ and $\tilde{R}_{z,x}(\Omega \times B) = P^z_x(B) = R_{z,x}(\Omega \times B)$.

$\tilde{R}_{z,x}$ defines a irreducible, aperiodic, positive Harris recurrent chain such that an invariant distribution with density f exists. Furthermore, the (Z,X) chain evolving according to $\tilde{R}_{z,x}$ converges towards this distribution which must have support $\tilde{\Omega}_2 \subset \Omega \times \Omega$.

Now, the idea is to show that the marginal stationary distribution of X, $f(x) = \int_\Omega f(z,x)\, d\gamma(z)$, is π. Further, let $f(z|x)$ denote the conditional density of Z given $X = x$ for the stationary distribution.

Because f is invariant for $\tilde{R}_{z,x}$ we know that

$$\int_{\Omega\times\Omega} f(z,x)\tilde{R}_{z,x}(A\times B)\, d(\gamma\times\gamma)(z,x) = \int_{A\times B} f(z,x)\, d(\gamma\times\gamma)(z,x).$$

With $A\times B = \Omega\times B$ we get

$$\int_{\Omega\times\Omega} f(z,x)P_x^z(B)\, d(\gamma\times\gamma)(z,x) = \int_B f(x)\, d\gamma(x),$$

and if we write $f(z,x) = f(x)f(z|x)$ we get

$$\int_\Omega f(x)\int_\Omega f(z|x)P_x^z(B)\, d\gamma(z)\, d\gamma(x) = \int_B f(x)\, d\gamma(x).$$

It is seen that the stationary distribution f of $\tilde{R}_{z,x}$ has the marginal distribution $\int_\Omega f(z,x)\, d\gamma(z) = \pi(x)$ if and only if π is invariant for the transition kernel $\tilde{P}_x$ defined by

$$\tilde{P}_x(B) = \int_\Omega f(z|x)P_x^z(B)\, d\gamma(z).$$

Whether π is invariant for $\tilde{P}_x$ or not is still an open question. □

It is obvious that π is invariant for $\tilde{P}_x$ if (i) Z and X are independent in f or if (ii) P_x^z does not depend on z. The first case means that we have no coupling between Z and X and they evolve according to $R_{z,x}$, and the second case means that X evolves as a Markov chain in itself. Neither of these cases applies here.

The kernel $\tilde{P}_x$ is a form of what Geyer (1999) call "state dependent mixing". In state dependent mixing a transition kernel is chosen according to a distribution which depends on the current state of the chain. However, if the target stationary distribution is invariant for each transition kernel this does not imply the invariance for the mixture transition kernel as pointed out in Geyer (1999, Section 3.4.2). Properties specific to our problem must be used to prove the conjecture or adjust $\tilde{P}_x$ to ensure that it has π as the stationary distribution.

References

Baddeley, A. J. & van Lieshout, M. N. M. (1993), 'Stochastic geometry models in high-level vision', *Advances in Applied Statistics: Statistics and Images* **1**, 231–255. Supplement to Journal of Applied Statistics vol. 20 Nos. 5/6 1993, *Statistics and images*, eds. K. V. Mardia and G. K. Kanji.

Cressie, N. & Lawson, A. B. (1998), Bayesian hierarchical analysis of minefield data, *in* 'Proc. SPIE Vol. 3392, Detection and Remediation Technologies for Mines and Minelike Targets III, Abinash C. Dubey; James F. Harvey; J. Thomas Broach; Eds.', pp. 930–940.

Dasgupta, A. & Raftery, A. E. (1998), 'Detecting features in spatial point processes with clutter via model-based clustering', *Journal of the American Statistical Association* **93**(441), 294–302.

Dimakos, X. K. (1999), 'A guide to exact simulation', *International Statistical Review* **unknown volume**. To appear.

Dralle, K. (1997), Locating trees by digital image processing of aerial photos, Ph. D. Thesis, Dina Research Report No. 58, The Royal Veterinary and Agricultural University.

Dralle, K. & Rudemo, M. (1996), 'Stem number estimation by kernel smoothing of aerial photos', *Canadian Journal of Forest Research* **26**, 1228–1236.

Dralle, K. & Rudemo, M. (1997), 'Automatic estimation of individual tree positions from aerial photos', *Canadian Journal of Forest Research* **27**, 1728–1736.

Fill, J. A. (1998), 'An interruptible algorithm for perfect sampling via Markov chains', *The Annals of Applied Probability* **8**, 131–162. preprint available from MCMC preprint server.

Fill, J. A., Machida, M., Murdoch, D. J. & Rosenthal, J. S. (1999), Extensions of Fill's perfect rejection sampling algorithm to general chains, Technical Report 592, Department of Mathematical Sciences, The Johns Hopkins University.

Geyer, C. J. (1999), *Likelihood Inference for Spatial Point Processes* in *Stochastic Geometry, Likelihood and Computation*, Chapman & Hall/CRC, chapter 3, pp. 79–140.

Geyer, C. J. & Møller, J. (1994), 'Simulation procedures and likelihood inference for spatial point processes', *Scandinavian Journal of Statistics* **21**(4), 359–373.

Gilks, W. R., Richardson, S. & Spiegelhalter, D. J. (1996), *Markov Chain Monte Carlo in Practice*, Chapman & Hall.

Green, P. J. (1995), 'Reversible jump Markov chain Monte Carlo computation and Bayesian model determination', *Biometrika* **82**(4), 711–732.

Kendall, W. S. & Møller, J. (1999), Perfect Metropolis-Hastings simulation of locally stable point processes, Preprint 347, University of Warwick, Department of Statistics.

Larsen, M. (1997), Crown modelling to find tree top positions in aerial photographs, In the third international airborne remote sensing conference and exhibition, Vol. II, Copenhagen, Denmark, pp. 428–435.

Larsen, M. & Rudemo, M. (1998), 'Optimizing templates for finding trees in aerial photographs', *Pattern Recognition Letters* **19**, 1153–1162.

Lund, J. (1998), Survival of the fattest? Self-thinning among trees, Report 3, The Royal Veterinary and Agricultural University, Department of Mathematics and Physics. Course report from PhD course in forest biometrics.

Lund, J., Penttinen, A. & Rudemo, M. (1999), Bayesian analysis of spatial point patters from noisy observations, Preprint, Department of Mathematics and Physics, The Royal Veterinary and Agricultural University. Submitted.

Lund, J. & Rudemo, M. (1999), 'Models for point processes observed with noise', *Biometrika* . To appear.

Lund, J. & Thönnes, E. (1999*a*), Perfect adaptive Metropolis-Hastings simulation for point processes, preliminary manuscript.

Lund, J. & Thönnes, E. (1999*b*), Perfect sampling from point patterns observed with noise, manuscript.

Propp, J. G. & Wilson, D. B. (1996), 'Exact sampling with coupled Markov chains and applications to statistical mechanics', *Random Structures and Algorithms* **9**(1 and 2), 223–252.

Smith, A. F. M. & Roberts, G. O. (1993), 'Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods', *Journal of the Royal Statistical Society, Series B, Statistical Methodology* **55**(1), 3–23.

Thönnes, E. (1999), A primer on perfect simulation, Preprint, Mathematical Statistics, Chalmers University of Technology.

van Lieshout, M. N. M. (1994), 'Stochastic annealing for nearest-neighbour point processes with application to object recognition', *Advances in Applied Probability* **26**, 281–300.

Paper 1

Jens Lund and Mats Rudemo

Models for point processes observed with noise

Revised version November 1, 1999

Report 10, Department of Mathematics and Physics
The Royal Veterinary and Agricultural University

Accepted for publication in Biometrika.

Paper 2

Jens Lund, Antti Penttinen, and Mats Rudemo

Bayesian analysis of spatial point patterns from noisy observations

October 27, 1999

Department of Mathematics and Physics
The Royal Veterinary and Agricultural University

Submitted

Paper 3

Jens Lund and Elke Thönnes

Perfect simulation of point patterns from noisy observations

December 10, 1999

Department of Mathematics and Physics
The Royal Veterinary and Agricultural University

Manuscript

Paper 4

Jens Lund

Survival of the Fattest? Self-thinning among Trees

April 29, 1998

Department of Mathematics and Physics
The Royal Veterinary and Agricultural University

Course report

Preliminary Manuscript

Jens Lund and Elke Thönnes

Perfect adaptive Metropolis-Hastings Simulation for Point Processes

December 13, 1999

Department of Mathematics and Physics
The Royal Veterinary and Agricultural University