Perfect simulation and inference for point processes given noisy observations

Jens Lund[1] and Elke Thönnes[2]

[1] Department of Mathematics and Physics, The Royal Veterinary and Agricultural University, Thorvaldsensvej 40, DK-1871 Frederiksberg C, Denmark
[2] Department of Statistics, University of Warwick, Coventry, CV4 7AL, UK

Summary

The paper is concerned with the exact simulation of an unobserved true point process conditional on a noisy observation. We use dominated coupling from the past (CFTP) on an augmented state space to produce perfect samples of the target marked point process. An optimized coupling of the target chains makes the algorithm considerable faster than with the standard coupling used in dominated CFTP for point processes. The perfect simulations are used for inference and the results are compared to an ordinary Metropolis-Hastings sampler.

Keywords: Noisy point process, marked point process, dominated coupling from the past, spatial birth-and-death process

1 Introduction

Data in form of a point pattern is very common in a variety of areas, for example in forestry. Often the data is a noisy observation of the underlying true point pattern. For instance, template matching as described in Larsen & Rudemo (1998) may be used to determine possible tree locations in aerial photographs. However, the irregular shape of tree tops, varying light conditions and motion due to wind lead to noisy observations of the tree locations. Further noise may be introduced by not observing certain trees or by identifying a tree where there is none.

Generally, we may describe the mechanism introducing noise into our observations as the combined effect of random or systematic displacements, censoring, thinning and superpositioning. Based on these point process operations a model which describes the degradation of the observations can be defined. For the example of forest photographs, Dralle & Rudemo (1996, 1997) use such a noise model for iterative least square estimation of its parameters. Lund & Rudemo (2000) derive the conditional likelihood for the observed process given the true process. A training set of true tree locations and their noisy observations is then used to perform approximate maximum likelihood estimation of the parameters of the noise model. The conditional likelihood of the observations given the true locations may also be used within a Bayesian framework, see Lund, Penttinen & Rudemo (1999). A prior distribution for the true locations together with the likelihood of the observations given the true point pattern yields a posterior distribution. As common for many point process models, the posterior distribution is not amenable to analytical examination. However, using MCMC we may sample the posterior distribution and use these samples within Monte Carlo tests. In Lund et al. (1999) a Metropolis-Hastings algorithm is suggested whose convergence is monitored by time series methods. Although these convergence diagnostics may warn when convergence has not been reached yet they do not guarantee convergence. The problem of verifying convergence can be solved using perfect simulation methods, see Propp & Wilson (1996) and Fill (1998). In this paper we present a perfect simulation algorithm which allows for the exact sampling of the posterior distribution in Lund et al. (1999). We apply a method called *dominated coupling from the past*, see for example Kendall & Møller (2001), to a spatial birth-and-death process and thus produce an exact sample of the target posterior distribution. This posterior distribution is specified as the marginal of a marked point process where the mark of each point influences the distribution of its location. As we do not have a closed form expression for the marginal distribution of the unmarked point process we cannot first sample the location of the points and then sample the marks. New aspects of the developed algorithm are the sampling of a marked point process, the use of an augmented state space and the problem-specific optimization of the coupling of the target chains.

In Section 2 we first present the noise model which specifies the relation between the true and the observed point pattern. This is followed by a description of the posterior distribution of the true point process. We then introduce dominated coupling from the past for marked point processes in Section 3 and devolop an algorithm for the forestry example in Section 4. Finally we present the results thus obtained in Section 5.

2 The posterior model

2.1 The noise model

We assume that the observations are a degradation of the true point pattern. Let $X \subset A$ denote the true point pattern in the bounded sampling window $A \subset \mathbb{R}^d$ and $Y \subset A$ describe the observed point pattern. The noise model is defined through the following point process operations:

1. *Thinning:* Each point $X_i \in A$ of the true point pattern is independently removed with probability $1 - p$.

2. *Displacement:* Each position X_i experiences a displacement which follows a two-dimensional Normal distribution of mean vector μ, variances σ_1^2 and σ_2^2 and correlation ρ. We denote the Normal density of the displaced location Y_j by $k(y|X_i)$.

3. *Censoring:* Some of the displaced locations may fall outside the observation window in which case they are not observed. This occurs with probability $p \int_{A^c} k(y|X_i)dy$, that is, if a point at location X_i is retained and displaced outside of A.

4. *Superposition:* We may detect a tree location where there is actually none. These 'ghost' points are modelled by an independent superimposed Poisson process of intensity α.

Under the above assumptions the noise model is completely specified by the parameter vector $\theta = (p, \alpha, \mu_1, \mu_2, \sigma_1, \sigma_2, \rho)$.

2.2 The likelihood of the data

For the above noise model, the conditional distribution of the observed locations given the true locations is derived in Lund & Rudemo (2000). Consider the process $(X_i, i \in M)$ of true locations and the process $(Y_j, j \in N)$ of observed positions. To simplify expressions we introduce an auxiliary variable S, a 'matching' which for each true location X_i, if observed, specifies

the index j of the corresponding observed location Y_j. Thus S implicitly determines which observation Y_j is caused by which true location X_i, which observations are 'ghost' points and which true locations are not observed due to thinning or censoring. Later on we will use the shorthand notation $s(\xi)$ for the index of the Y-point which is matched to the X-point ξ by the matching s.

As derived in Lund & Rudemo (2000) the joint distribution of (Y, S) given $X = x$ and θ factorizes as

$$T(y, s \mid x, \theta) \quad = \quad L_1 \, L_2 \, L_3 \, L_4. \tag{1}$$

Here $L_1 = p^{|M_1|} \prod_{i \in M_1} k(y_{s(X_i)}|X_i)$ where M_1 denotes the set of indices of the true locations which lead to an observation in A. Thus L_1 corresponds to retained and displaced locations falling inside the sampling window. The second factor $L_2 = \prod_{i \in M \setminus M_1} \left\{ p \int_{A^c} k(y|x_i)dy + (1-p) \right\}$ derives from locations that were either thinned or retained and censored. Let N_1 denote the set of indices of observational points which are caused by a true location. The third factor $L_3 = \alpha^{|N \setminus N_1|} \exp\left\{(1-\alpha)|A|_d\right\}$ is due to 'ghost' points, where we have used the notation $|A|_d$ for the d-dimensional volume of A. Finally $L_4 = \mathbb{I}_{[s \text{ is a matching}]}$ is an indicator function which ensures that s is a coherent matching, that is it assigns to each observation point at most one point in x.

The likelihood of Y given (x, θ) may be determined by marginalisation:

$$L(y|x, \theta) \quad = \quad \sum_{s \in V} T(y, s \mid x, \theta).$$

Here V is the set of all possible matchings. Unfortunately this sum does not simplify to a more compact algebraic expression. In the following we will skip any reference to θ in our notation, but assume that it is fixed and known.

2.3 The prior model

The prior model describes our beliefs about the stochastic properties of the underlying 'true' point process. We consider the following three choices. Our first prior model is a homogeneous Poisson process of intensity $\beta > 0$. Its density with respect to a unit rate, homogeneous Poisson process is given by

$$f(x) \quad = \quad \exp\left\{(1-\beta)|A|_d\right\} \beta^{n(x)} \qquad\qquad x \subset A.$$

Here $n(x)$ is the number of points in x.

The forest considered in this paper was subjected to a thinning experiment which motivates the choice of the following two priors modelling repulsive

interaction. The Strauss process (Strauss 1975) is defined by the density

$$f(x) \quad = \quad c\,\beta^{n(x)}\,\gamma^{b(x)}, \qquad\qquad x \subset A$$

with respect to a unit rate homogeneous Poisson process on A. Here β is positive, $0 < \gamma < 1$, and c is the normalizing constant. The exponent $b(x)$ counts the number of pairs of neighbour points, that is pairs of points which are less than a distance R apart.

We may assume that the interaction is not restricted to a certain radius but reduces continuously with increasing distance. This can be modelled using a pairwise interaction Markov point process with a continuous interaction function H. The density of such a point process with respect to a unit rate, homogeneous Poisson process is given by

$$f(x) \quad = \quad c\,\beta^{n(x)} \prod_{i,j:i<j} H(||x_i - x_j||),$$

where $||\cdot||$ is the Euclidean distance in $\mathbb{R}^2$. We consider the logistic interaction function

$$H(r) \quad = \quad \frac{1}{1+\exp\{-\phi(r-R)\}}$$

with slope $\phi > 0$ and $H(R) = 0.5$. Thus H specifies a repulsive interaction whose strength reduces with increasing distance between two points.

In the following we will make use of the Papangelou conditional intensity of the prior model. For a point process with density f the Papangelou conditional intensity is defined as

$$\lambda(x,\xi) \quad = \quad \frac{f(x \cup \{\xi\})}{f(x)} \qquad \text{for } f(x) > 0 \text{ and } \xi \notin x,$$

where x is a point pattern and ξ is a single point in A. Thus, for a Poisson prior of intensity β we have $\lambda(x,\xi) = \beta$. For the Strauss process we have $\lambda(x,\xi) = \beta\,\gamma^{t(x,\xi)}$ where $t(x,\xi)$ is the number of neighbours of ξ in x. Finally, for a Markov point process with logistic interaction function we have

$$\lambda(x,\xi) = \beta \prod_{x_j \in x} H(||x_j - \xi||).$$

2.4 The posterior density

Our aim now is to sample the joint posterior distribution of the true point process X and the matching S given the observation $Y = y$. This posterior density is given by

$$\pi(x,s) \quad = \quad L(x,s|y,\theta) \quad \propto \quad L(x)\,T(y,s \mid x,\theta) \qquad (2)$$

Here $L(x)$ is the prior density for the true point process and $T(y, s|x, \theta)$ is defined as in (1).

3 Perfect simulation for Markov point processes

3.1 Simulation of point processes

To produce samples of point processes with varying number of points we commonly use spatial birth-and-death processes, see Preston (1975), or Metropolis-Hastings algorithms, see Geyer & Møller (1994) and Møller (1999). A description of a Metropolis-Hastings type algorithm for the target posterior density in (2) can be found in Lund et al. (1999). In the following we use a spatial birth-and-death process because its implementation in a perfect simulation algorithm is less intricate than for a Metropolis-Hastings algorithm. The interested reader may find a description of a general perfect Metropolis-Hastings algorithm for locally stable Markov point processes in Kendall & Møller (2001).

3.2 Dominated coupling from the past for point processes

This section provides the reader with a short review on how to use dominated coupling from the past to produce perfect samples for a locally stable Markov point processes. The exposition follows Kendall & Møller (2001), where further details and formal proofs may be found; see also Møller (2001).

Suppose the point process of interest has the density $f(x)$ with respect to a unit rate, homogeneous Poisson point process. We assume that f is hereditary and that its Papangelou conditional intensity $\lambda(x, \xi)$ is uniformly bounded by a constant λ^*. We may produce a spatial birth-and-death process X whose distribution converges towards the desired point process distribution by choosing its death rate $d(x, \xi)$ constant to one for each $\xi \in x$ and its birth rate

$$b(x, \xi) \quad = \quad \lambda(x, \xi).$$

The key ingredients of dominated coupling from the past as presented in Kendall & Møller (2001) are a dominating process and an upper and a lower sandwiching process constructed as follows. The dominating process D is a stationary spatial birth-and-death process with a unit death rate and a birth rate which is constant to λ^*. Each of the birth times t in D is assigned a mark V_t which is uniform on the unit interval. This allows the coupling of paths of the target spatial birth-and-death process X to D as follows. We

start X in a subset of the current configuration D. Now, whenever at time t a point ξ is born in D, we check whether

$$V_t \quad \leq \quad \frac{\lambda(X_{t-}, \xi)}{\lambda^*},$$

where X_{t-} is the configuration of X immediately prior to time t. If the above inequality is satisfied, then the point ξ is also born in X at time t; otherwise no transition takes place in X at this time. Whenever a point dies in D we remove it from X if it is in the current configuration of X. Notice that due to this construction $X_t \subseteq D_t$ at all times t.

The lower and upper sandwiching processes L and U are coupled to D in a similar manner. We start the lower process in the empty set and the upper process in the same configuration as D. Now, whenever a point ξ is born at time t in D we check whether

$$V_t \leq \frac{\min\{\lambda(x, \xi), L_t \subseteq x \subseteq U_t\}}{\lambda^*} \quad \text{or} \quad V_t \leq \frac{\max\{\lambda(x, \xi), L_t \subseteq x \subseteq U_t\}}{\lambda^*}.$$

If the first and thus also the second condition is satisfied, then ξ is born at time t in both L and U. If only the second condition is satisfied, then ξ is born at time t in U, but L does not experience a birth at time t. If none of the two conditions are satisfied, then neither process changes at time t. If a point ξ dies at time t in the dominating process D, then it dies in L respectively U subject to its existence.

Dominated coupling from the past now works as follows. We simulate a stationary version of D on the time interval $[-T, 0]$ and mark all birth times. Then we simulate the lower process $L^{(-T)}$ and upper process $U^{(-T)}$ started at time $-T$. As mentioned earlier, we set the initial states of these two processes at time $-T$ to $L_{-T}^{(-T)} = \emptyset$ and $U_{-T}^{(-T)} = D_{-T}$. We then evolve the upper and lower process from time $-T$ to time 0 coupled to the realisation of D as described above. If the state of the two processes at time 0 coincides, then we return this state as a perfect sample. If not, then we extend the realisation of D to a stationary path on $[-2T, 0]$ and mark all new birth times in $[-2T, 0]$. (Notice that in the realisation of D on $[-2T, 0]$ we reuse the previous realisation of D on $[-T, 0]$ including its birth time marks). We again start a path of the two sandwiching processes, now at time $-2T$, by setting $L_{-2T}^{(-2T)} = \emptyset$ and $U_{-2T}^{(-2T)} = D_{-2T}$. As before we evolve these two processes coupled to D up to time 0 and check for coalescence. We repeat the above procedure, iteratively going further and further back into the past until we produce output.

Notice that if D hits the empty set, then this is sufficient for the corresponding upper and lower process to coalesce. As the empty set is an ergodic atom of D we are sure to produce output in finite time. Further details and a formal proof for the correctness of the algorithm can be found in Kendall & Møller

(2001). Nevertheless, we would like to give the reader a heuristic motivation. Suppose we could have started the target chain X and the dominating chain D at time $-\infty$, where the initial pattern of X would have been the empty set. Then at time $-T$ the process X would be in a state x, such that $L_{-T}^{(-S)} \subseteq x \subseteq U_{-T}^{(-S)}$ for any $-S \leq -T$. Furthermore, X would remain sandwiched between the lower and upper process up to time 0. Thus, if $L_0^{-S} = U_0^{-S} = y$, then X would be in y at time 0. Due to the ergodicity of X, our heuristic suggests that, as X has run for an infinite time, the state y is a sample from the equilibrium distribution.

We may extend the algorithm by Kendall & Moller as follows. If we can bound the Papangelou conditional intensity of the process of interest using a tighter bound $\lambda(x, \xi) = \lambda^* \, l(\xi) \leq \lambda^*$, where $l(\cdot)$ is an intensity function on A, then we may set the birth rate of the dominating process to $\lambda^* \, l(\cdot)$ instead of λ^*. It can still be easily simulated in stationarity, as its equilibrium distribution is an inhomogeneous Poisson process.

3.3 Dominated coupling from the past for marked point processes

We may adapt the above algorithm to produce a perfect sample of a marked point process. Suppose the mark distribution is such that we cannot sample the marks given the location of the points. This means that we need to produce a perfect sample for the joint distribution of the marks and the locations. Similar to the unmarked case, we may produce a spatial birth-and-death process of marked point pattern whose distribution converges under regularity conditions to the target distribution. We may achieve this by setting its death rate to one and its birth rate to

$$b\Big((x, m), (\xi, n)\Big) \quad = \quad \lambda\Big((x, m), (\xi, n)\Big),$$

where (x, m) is a marked point pattern and n is the mark of point ξ. Suppose

$$\lambda((x, m), (\xi, n)) \quad \leq \quad \lambda^* M(dn),$$

where M is a probability measure on the mark space. Then we may produce a dominating process D as a spatial birth-and-death process of unit death rate and birth rate λ^*. Each birth time t is then marked with a uniform mark V_t and each point ξ that is born in D is marked with n sampled from $M(dn)$. The lower and the upper process are simulated analogous to before but replacing $\lambda(x, \xi)$ with $\lambda((x, m), (\xi, n))$.

Alternatively, we may set the above approach into the framework of Kendall & Møller (2001) by expressing the density of the marked point process with respect to a Poisson process living on the product space of mark space and

the space of finite point configurations on A. The correctness of the algorithm then follows from the proofs in Kendall & Møller (2001). For the algorithm developed in the next section such a description may be found in Møller (2001).

In the following we will simulate a marked point whose mark space is finite and whose Papangelou conditional intensity satisfies

$$\lambda\Big((x,m),(\xi,n)\Big) \quad \leq \quad \begin{cases} \lambda_1 l_1(\xi)\frac{1}{|L|} & \text{if } n \in L \\ \lambda_2 l_2(\xi)\frac{1}{|L^c|} & \text{if } n \in L^c \end{cases}$$

for a subset L of the mark space. Here $|L|$ denotes the cardinality of L. In this case we may produce a dominating process D by superimposing two spatial birth-and-death processes D_k, where each D_k has a unit death rate and a birth rate which is $\lambda_k l_k(\cdot)$ for $k = 1, 2$. Each point ξ born in D_1 is assigned a mark chosen uniformly from L, each point ξ born in D_2 is assigned a mark chosen uniformly from L^c.

4 Perfect simulation of the posterior density

4.1 The target spatial birth-and-death process

In the following we specify a spatial birth-and-death process X which fits into the framework of the previous section and whose stationary distribution is the target distribution in (2). Essentially each configuration of X is a multi-type point pattern. The first type of points are so-called *matched* points. These are locations x_i to which the current matching s assigns an observed location y_j. The second type of points are *unmatched points*, these are points which did not lead to an observation due to thinning or displacement outside of the window. Recall that the matching s specifies for each matched point the index of the observation point that it is matched to. Thus the marks of the point pattern, which indicate whether a point is unmatched or matched and in the latter case its matched observation point, are implicitly given by s.

To ensure the correct stationary distribution we choose a unit death rate for all points and birth rates as follows. If the current configuration is (x, s), then the birth rate for unmatched points is given by

$$b\Big(\ (x,s),\ (x \cup \{\xi\}, s)\ \Big) \quad = \quad \gamma(x,\xi)\left\{p\int_{A^c} k(y|\xi)dy + (1-p)\right\}, \qquad (3)$$

where $\gamma(x,\xi)$ is the Papangelou conditional intensity of the prior model. Note that we assume that s does not change as ξ is an unmatched point.

Let $s \cup s(\xi)$ be the matching generated by adding to the existing matching s a matching of the point ξ with an observed point y of index $s(\xi)$. The birth rate for matched points is given by

$$b\Big(\ (x,s),\ (x \cup \{\xi\}, s \cup s(\xi))\ \Big) = \gamma(x,\xi)\ k(y_{s(\xi)}|\xi)\ \frac{p}{\alpha}\ \mathbb{I}_{[s \cup s(\xi) \text{ is a matching}]}. \tag{4}$$

The indicator function in the last expression ensures that a matched point ξ is only born if $s(\xi)$ matches the point to an unmatched observation point y_j.

4.2 The dominating process

Similar to X the dominating process D is a spatial birth-and-death process whose states are marked point patterns (z,q). Like X the process D consists of matched and unmatched points which both have unit death rates. The birth rates of D, however, are higher than in X. Let $\lambda^* = \sup_{x,\xi} \gamma(x,\xi)$ where $\gamma(x,\xi)$ is the Papangelou conditional intensity of our prior model.

Unmatched points in D have a birth rate given by

$$b_D\Big(\ (z,q),\ (z \cup \{\xi\}, q)\ \Big) \quad = \quad \lambda^* \left\{ p \int_{A^c} k(y|\xi)dy\ + (1-p) \right\} \quad = \quad \lambda^*\ l(\xi)$$

where l, which is defined by the last equality, is less than one.

In the following we use an auxiliary variable Q which assigns to each matched point in D an observation point $y_j \in Y$. However, we do not require that Q is a matching, that is Q may assign more than one matched point to the same observation point. Thus the dominating process lives on a state space which is an augmentation of the state space of X. The construction of a dominating chain on an augmented state space is a useful tool for dominated coupling from the past and was also used in Kendall & Thönnes (1999) and Cai & Kendall (2002). We will motivate later why we do not restrict Q to be a matching. We call Q an 'assignment' to contrast it from a matching. Matched points in D have a birth rate which is defined by

$$b_D\Big(\ (z,q),\ (z \cup \{\xi\}, q \cup q(\xi))\ \Big) \quad = \quad |N|\lambda^*\ \frac{p}{\alpha}\ k(y_{q(\xi)}|\xi)\frac{1}{|N|}, \tag{5}$$

where q is an assignment and $|N|$ the number of observation points. Thus, the matched point component is a time-reversible spatial birth-and-death process whose stationary distribution is an inhomogeneous Poisson process with intensity measure $\lambda^* p \alpha^{-1} k(y_{q(\xi)}|\xi)$. The stationary distribution of the matched points thus may be interpreted as a cluster process with the observation points as the cluster centres. The cluster around y_j has a Poisson

number of daughters with mean $\lambda^* p \alpha^{-1} \int_A k(y_j|\xi)d\xi$ and the daughters are scattered according to a Normal distribution around the cluster centers.

The reason we allow q to be an assignment and not a matching is due to the fact that the birth rate of matched points in the target chain contains an indicator function $\mathbb{I}_{[s \cup s(\xi) \text{ is a matching }]}$ ensuring that the target chain has at most one point matched to each observation point y_j. However, if for example the set of matched points specified in s is empty then this indicator function will be one and thus needs to be bounded by one. Hence we are forced to allow the dominating process D to have more than one point matched to an observation point. For a similar reason, the upper process U described in the next section also lives on the augmented state space.

As its birth rates are higher, the chosen process D dominates the target chain X. However, it will have many more matched points than X and so be unnecessarily large. This slows down the CFTP algorithm considerably. We may reduce the number of matched points per observation point by synchronising the death times of points which are matched to the same observation point. This is justified by the fact that we know that the target chain has only one matched point per observation point. In this altered dominating process $\tilde{D}$ births of matched points and births of unmatched points occur in the same way as in D. However, deaths of matched points do not occur at unit rate any longer. Instead we assign to each observation point y_j a unit rate Poisson counting process P_j. Whenever there is an incident time in P_j all the matched points in $\tilde{D}$ which are currently assigned to y_j die. Thus all the matched points which are currently alive and assigned to the same observation point have the same death time. However, due to the properties of the Poisson counting process each matched point ξ has an exponential lifetime of mean one. Thus if we derive a version of the target chain from $\tilde{D}$, its death rate for matched points will still be one, as it has at most one matched point per observation point.

The reader may have noticed that $\tilde{D}$, in contrast to D, is no longer time-reversible. However, we can produce a stationary path of $\tilde{D}$ on the time interval $[-T, 0]$ as follows. Let $-T_j$ be the most recent incident time prior to $-T$ of the Poisson process P_j. Simulate a stationary path of D and paths of $(P_j)_{j \in N}$ on $[-\tilde{T}, 0]$ where $-\tilde{T} = \min\{-T_j : j \in N\}$. We can now derive a stationary path of $\tilde{D}$ on $[-T, 0]$ from these paths. The unmatched points of $\tilde{D}$ coincide at any time in $[-T, 0]$ with the unmatched points of D. Moreover, the matched points in $\tilde{D}$ are born at the same time as the matched points in D. However, we ignore the death times of matched points in D. Instead, the matched points assigned to y_j die in $\tilde{D}$ whenever there is an incident in P_j. Thus the matched points of $\tilde{D}_{-T}$ which are assigned to the observation point y_j are the matched points of D born in the interval $[-T_j, T]$ and assigned to y_j. As all $(P_j)_{j \in N}$ suffer at least one incident in $[-\tilde{T}, -T]$ we can construct

a stationary path of $\tilde{D}$ on $[-T, 0]$ given a stationary path of D and paths of $(P_j)_{j \in N}$ on $[-\tilde{T}, 0]$. We found in simulation experiments that $-\tilde{T}$ is usually close to $-T$ and thus this approach does not influence much the speed of the algorithm. A more elegant approach was suggested to us by Jesper Møller after he read an earlier manuscript of this paper. For the given case, we actually can compute the distribution of the number of matched points per observation point in stationarity. It is geometric with parameter $\alpha/(\alpha + \lambda^* p)$ and thus easy to sample, see Møller (2001).

4.3 The upper and the lower process

In this section we construct the lower and the upper process whose coalescence implies the complete coalescence of the relevant paths of X. To determine the dynamics of the upper and the lower chain we first consider the acceptance rules for births in X. For unmatched points we check whether $V_t \leq \gamma(x, \xi)/\lambda^*$. Clearly, the larger the Papangelou conditional intensity $\gamma(x, \xi)$ of the prior model the easier it is to satisfy this condition. But for a Poisson prior or a repulsive process like our other two prior models we have $\gamma(x, \xi) \geq \gamma(z, \xi)$ if $x \subseteq z$. Thus the acceptance rule for unmatched points is anti-monotone, that is the smaller x with respect to the subset relation the more likely it is that a birth of an unmatched point ξ is accepted.

For matched points we check two conditions when deciding whether to accept a birth in X. The first condition is the same as the condition for unmatched points and therefore anti-monotone. For the second condition consider two configurations (x, s) and (z, q) of X where $x \subset y$ and s is such that any point in x that is matched to a y-point is matched to the same y-point by q. Now, if $s(\xi) = q(\xi)$ then

$$\mathbb{I}_{[s \cup s(\xi) \text{ is a matching}]} \quad \geq \quad \mathbb{I}_{[q \cup q(\xi) \text{ is a matching}]}$$

because the observation points not matched by (x, s) are a superset of the observation points not matched by (z, q). It follows that the acceptance rule for births of matched points is also anti-monotone.

Because the acceptance rules for births are anti-monotone, we may produce a realisation of the lower and the upper chain as follows. Suppose we have a stationary realisation of $\tilde{D}$ on $[-T, 0]$. Let $U^{(-T)}$ and $L^{(-T)}$ denote the upper respectively the lower chain started at time $-T$. We set $U^{(-T)}_{-T} = \tilde{D}_{-T}$ and $L^{(-T)}_{-T} = \emptyset$. We then evolve the two chains as follows.

1. Whenever we have a birth of an unmatched point ξ in $\tilde{D}$, we check whether the mark V_t of this birth time t satisfies $V_t \leq \gamma(U^{(-T)}_{t-}, \xi)/\lambda^*$, in which case ξ is born at time t in $U^{(-T)}$ and $L^{(-T)}$. If the above

condition is not satisfied we check whether $V_t \leq \gamma(L_{t-}^{(-T)}, \xi)/\lambda^*$. Note that this is a less stringent condition. If it holds then the unmatched point is born at time t only in $U^{(-T)}$. The lower process does not have a transition. If the latter condition is not satisfied then neither the upper process nor the lower process change at time t.

2. Suppose that at time t we have the birth of a matched point ξ in $\tilde{D}$ which is assigned to the observation point $y_{q(\xi)}$. We first check whether $V_t \leq \gamma(U_{t-}^{(-T)}, \xi)/\gamma^*$ and furthermore whether the observation point $y_{q(\xi)}$ is not matched in $U_{t-}^{(-T)}$. If both conditions are satisfied then the matched point ξ is born at time t both in the upper and the lower chain where it is matched to $y_{q(\xi)}$. If not both of the conditions are satisfied then we check whether $V_t \leq \gamma(L_{t-}^{(-T)}, \xi)/\lambda^*$ and, moreover, whether the observation point $y_{q(\xi)}$ is not matched in $L_{t-}^{(-T)}$. If these two conditions are satisfied, then the unmatched point ξ is born in the upper path at time t but not in the lower path. Note that this procedure allows the upper path to have several matched points which are assigned to the same observation point. If one or two of the latter two conditions are not satisfied no transition takes place at time t in either the upper or the lower chain.

3. Now consider an incident time t of the Poisson process P_j. Here, the points matched to y_j, if any, die both in the lower and upper path at time t.

4. Finally, consider a death time t of an unmatched point in $\tilde{D}$. Here we simply check whether the point whose death time is t is in the upper respectively the lower path and let it die in either path subject to its existence.

The upper process and the dominating chain live on a larger state space than the target chain. However, as the lower process almost surely remains in the state space of X, the upper process will be in the state space of X once it coalesces with the lower process.

4.4 The perfect simulation algorithm

The perfect simulation algorithm now works as follows. We simulate a stationary version of $\tilde{D}$ on $[-T, 0]$. Then we simulate the upper and the lower process as described in Section 4.3 and check whether their states at time 0 coincide. If so, we return their common state as a sample from the equilibrium distribution. If not we extend the stationary path of $\tilde{D}$ and its marks on $[-T, 0]$ to a stationary path on $[-2T, 0]$. Again, we simulate the upper

and the lower process, now on the time interval $[-2T, 0]$, and check for coalescence. We proceed doubling the time interval until the lower and the upper path have coalesced at time 0 and then return their common state as a sample from the target posterior density.

The dominating process and the sandwiching processes satisfy the conditions specified in Kendall & Møller (2001). For brevity we omitted any proofs here.

5 Results and Discussion

5.1 The observed point process obtained from template matching

The data analysed originate from aerial photographs of a thinning experiment of Norway spruce (*Picea abies* (L.) Karst.), see Dralle & Rudemo (1996, 1997) for detailed descriptions of the experiment and image acquisition. The observed point process $Y = (Y_1, \ldots, Y_n)$ with $n = 206$ points was obtained by template matching (Larsen & Rudemo 1998). This data set together with the 'true' tree top positions $X = (X_1, \ldots, X_m)$ with $m = 171$ was analysed in Lund & Rudemo (2000) with focus on the conditional likelihood $L(Y|X)$. The true X positions are obtained from ground measurements projected onto the image. In the present paper we study the posterior distribution of X given the observation of Y. We assume that we know the prior distribution of X and the conditional likelihood $L(Y|X)$ but not the true positions of the X-points. The 'true' X-points and the observation points Y are shown in Figure 1.

To specify the conditional likelihood $L(Y|X)$ we used the parameter estimates from Lund & Rudemo (2000) which are summarized in Table 1.

θ	p	α	μ_1	μ_2	σ_1^2	σ_2^2	$\sigma_1\sigma_2\rho$
Estimate	0.941	0.0002745	-0.342	0.0815	1.047	2.028	-0.0489

Table 1: Parameter vector $\theta = (p, \alpha, \mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$ used in the analysis. The unit for α is points per pixel, and the units for μ_1 and μ_2 are pixels.

Lund et al. (1999) analyse the unobserved point pattern X using a conventional forward Metropolis-Hastings sampler. Besides a thorough Bayesian analysis the authors also provide a discussion of the choice of the prior.

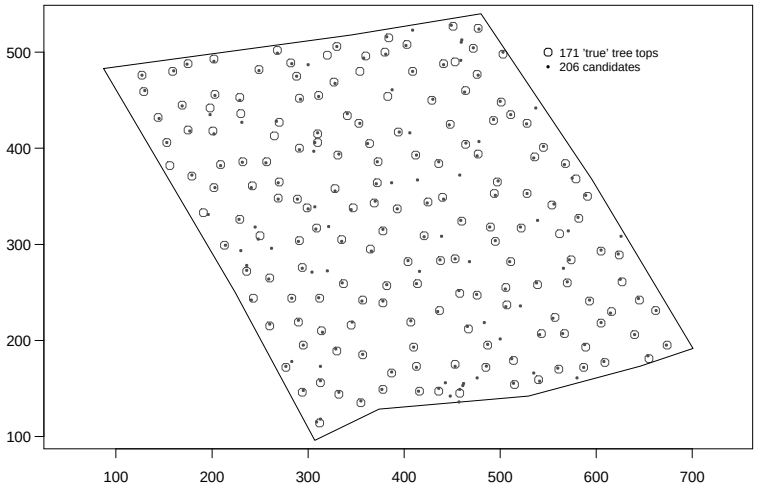

Figure 1: 171 X-points (centres of circles) corresponding to 'true' tree tops and 206 Y-points (dots) corresponding to template matching. The area of the delineated subplot is 4 454 m^2, and the unit of the axes is linear pixel size, 0.15m. In the practical calculations we have used the convex border for convenience, and have thus increased the area with 2.12%.

5.2 Discussion

As an example, Figure 2 shows a perfect sample from the posterior distribution where the prior is a strongly repulsive Strauss process with parameters $\beta|A|_d = 238$, $\gamma = 0.02$, and $R = 10$.

Figure 3 shows the number of points in the dominating chain as well as in the upper and lower processes. Note that the number of points in the dominating chain is much larger than in the target chain.

Given perfect samples from the posterior distribution we may now continue to perform a full Bayesian analysis similar to Lund et al. (1999). The statistical analysis of point pattern is usually based on summary functions like the reduced second-order moment function K. In practice the L-function is used

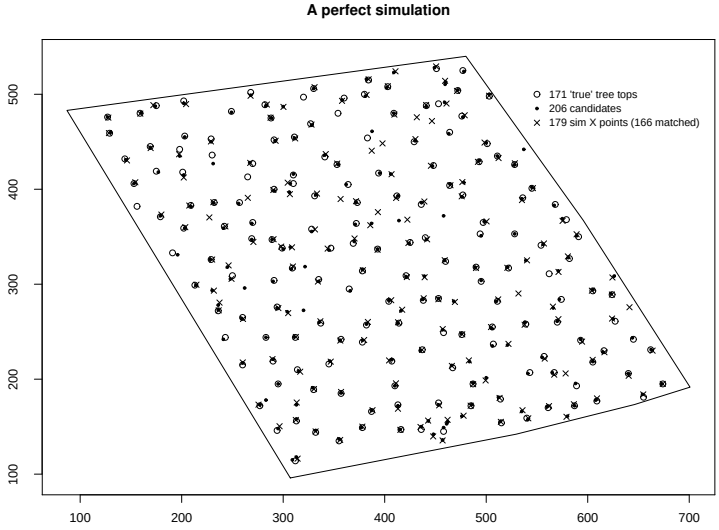

Figure 2: A perfect sample from the posterior distribution where the prior distribution is a Strauss process with parameters $\beta|A|_d = 238$, $\gamma = 0.02$, and $R = 10$.

rather than K, where

$$L(r) \quad = \quad \sqrt{\frac{K(r)}{r}} \qquad \text{for } r \geq 0.$$

Figure 4 shows the estimated L-function derived from 50 independent perfect samples from the posterior distribution for the above Strauss prior. For comparison the graph also displays the L-function estimated from the noisy Y points as well as the one estimated from the true X points. Our posterior based estimate is better than the Y-based estimate in the distance range 0–12, but does not differ much at larger distances.

Our algorithm works efficiently for realistic priors like the one in Figure 2 or for instance a logistic interaction function with the parameters $\beta|A|_d = 225$, $R = 7$ and $\phi = 0.314$ (equivalent to $H(0) = 0.1$). The use of the smaller dominating chain $\tilde{D}$ introduced in Section 4.2 on page 11 and the corresponding coupling of the target chains increased the speed of our algorithm for this choice of parameters by a factor of 14. For a weak repulsive prior with logistic interaction function with parameters $\beta|A|_d = 175$, $R = -3$ and $\phi = 0.462$ (equivalent to $H(0) = 0.8$) the unoptimised algorithm took 72 minutes to run, whereas the optimised algorithm took only 5 minutes. Unfortunately,

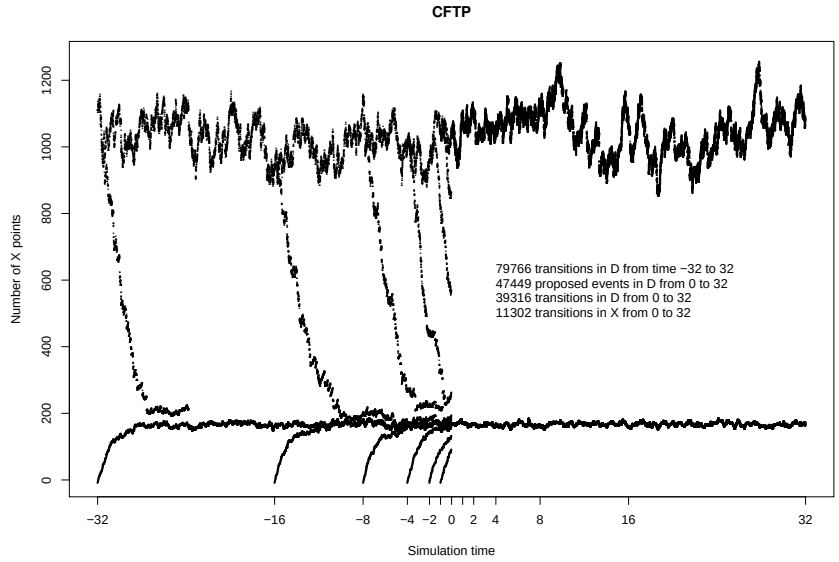

Figure 3: The number of points in the dominating chain, and in the upper and lower processes. The dominating chain and the target chain are continued forwards from time 0. We see that the coalescence time T_C is between 16 and 32. The figure specifies the number of transitions of some of the paths shown. The number of "proposed events" in D includes for instance proposals of unmatched X-points that are not added to the dominating chain. The number of transitions counts the number of times a path changes its configuration.

the algorithm is not efficient enough for the priors with very strong interaction like the ones in Lund et al. (1999). For these priors the dominating chain becomes too large and our perfect algorithm takes too long time to terminate. Nevertheless, we may use the perfect simulation algorithm to see whether the convergence diagnostics used in Lund et al. (1999) fail to detect non-convergence. We used the non-perfect Metropolis-Hastings algorithm in Lund et al. (1999) to produce (approximate) samples from the posterior chain. The Metropolis-Hastings chain was run until the time-series plots indicated convergence. We then used these samples to estimate the L-function of the posterior distribution. Figure 4 shows the L-function derived from the non-perfect Metropolis-Hastings algorithm both for the Strauss prior used here and for a logistic interaction function prior. (Unfortunately, for the logistic prior we were not able to produce an L-function estimate based on perfect samples, because for a repulsion of this strength the CFTP algorithm

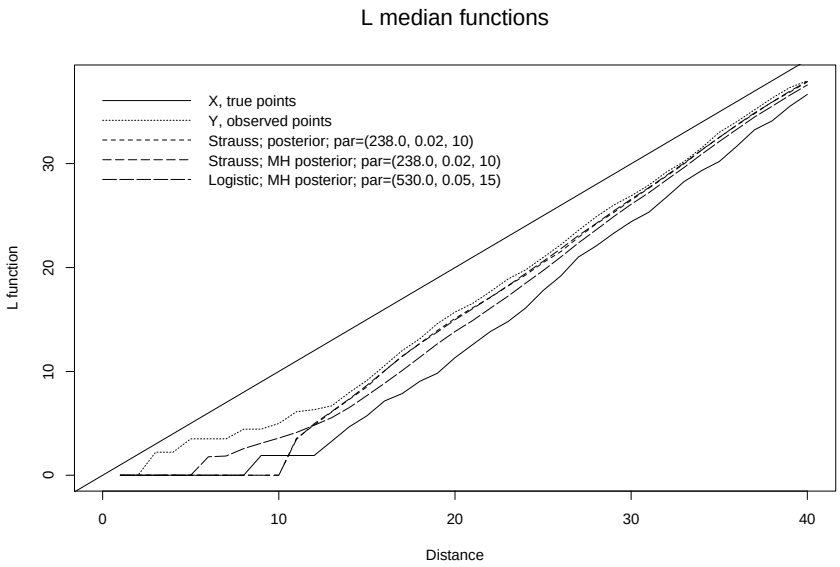

Figure 4: Estimated L function based on the true points, the observed points, and simulated points from the posterior distribution. The simulation based estimates are formed by taking the pointwise median of the functions estimated on the basis of each simulated point pattern. Simulation based estimates are shown for 50 perfect simulations from the posterior distribution with a Strauss prior, and for the non-perfect Metropolis-Hastings based simulation algorithm described in Lund et al. (1999) (for two priors). The two L-functions for Strauss priors are nearly identical, no matter the simulation method.

is too inefficient). The L-functions with the Strauss prior based on i.i.d. perfect samples and the Metropolis-Hastings samples are almost identical, indicating that the L-function estimated from the Metropolis-Hastings algorithm is actually an adequate estimate of the L-function from the posterior distribution. Although these results allow only limited generalization to the case of posterior distributions with stronger repulsive prior interaction, they do increase our confidence that the results of the convergence diagnostics used in Lund et al. (1999) are de facto indicating convergence.

Thus, our algorithm may be used for the statistical analysis of an observed point pattern, if the assumed prior does not have too strong a repulsive interaction. It may also be used to examine whether convergence diagnostics may falsely indicate convergence for the non-perfect algorithm. Thus we hope to

have illustrated how to use coupling from the past in a real life statistical inference problem whilst pointing out some of the present limitations of perfect sampling. For other examples see Murdoch & Green (1998), Møller (1999), and Møller & Nicholls (1999).

Acknowledgements

Elke Thönnes thanks for the support by the EU TMR network ERB-FMRX-CT96-0096 and the Stochastic Centre at Chalmers University. Both authors acknowledge the support by the ESF program "Highly structured stochastic systems". We are grateful to Ib Skovgaard, Wilfrid Kendall and Jesper Møller for inspiring discussions. We also thank Mats Rudemo for helpful comments, and the Danish Forest and Landscape Research Institute for access to the data.

References

Cai, Y. & Kendall, W. S. (2002), 'Perfect simulation for correlated Poisson random variables conditioned to be positive', *Statistics and Computing* **To appear**.

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

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

Fill, J. A. (1998), 'An interruptible algorithm for perfect sampling via Markov chains', *Ann. Appl. Probab.* **8**, 131–162.

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

Kendall, W. S. & Møller, J. (2001), 'Perfect simulation using dominating processes on ordered spaces, with application to locally stable point processes', *Adv.in Appl. Probab.* **32**, 844 – 865.

Kendall, W. & Thönnes, E. (1999), 'Perfect simulation in stochastic geometry', *Pattern Recognition* **32**, 1569 – 1586.

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

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. (2000), 'Models for point processes observed with noise', *Biometrika* **87**(2), 235–249.

Møller, J. (1999), 'Perfect simulation of conditionally specified models', *J. Roy. Statist. Soc. Ser. B* **61**(1), 251–264.

Møller, J. (2001), A review on perfect simulation in stochastic geometry, *in* C. C. H. I. V. Basawa & R. L. Taylor, eds, 'Selected Proceedings of the Symposium on Inference for Stochastic Processes', Vol. 37 of *IMS Lecture Notes & Monographs*, pp. 333 – 355.

Møller, J. & Nicholls, G. K. (1999), Perfect simulation for sample-based inference, Research report R-99-2011, Aalborg University, Department of Mathematics.

Møller, J. (1999), Markov Chain Monte Carlo and Spatial Point Processes, *in* O. Barndorff-Nielsen, W. Kendall & M. van Lieshout, eds, 'Stochastic Geometry: likelihood and computation', Chapman & Hall/CRC, Boca Raton, pp. 141–172.

Murdoch, D. J. & Green, P. J. (1998), 'Exact sampling from a continuous state space', *Scand. J. of Statist.* **25**, 483–502.

Preston, C. (1975), 'Spatial birth-and-death processes', *Bull. Internat. Statist. Inst.* **46**, 371–391.

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

Strauss, D. (1975), 'A model for clustering', *Biometrika* **63**, 467–475.