Perfect simulation of point patterns from noisy observations

Jens Lund
Department of Mathematics and Physics
Royal Veterinary and Agricultural University, Copenhagen

Elke Thönnes
Department of Statistics,
University of Warwick, Coventry

December 10, 1999

Abstract

The paper is concerned with the Bayesian analysis of point processes which are observed with noise. It is shown how to produce exact samples from the posterior distribution of the unobserved true point pattern given a noisy observation. The algorithm is a perfect simulation method which applies dominated Coupling From The Past (CFTP) to a spatial birth-and-death process. Dominated CFTP is made amenable by simulating on an augmented state space. We discuss how to use this perfect simulation algorithm in a statistical inference problem of pratical importance and describe the merits and the limitations of the method.

1 Introduction

Point processes are versatile models for numerous applications, such as galaxies in astronomy, cell nuclei in cytology or maps of tree locations in forestry. Ideally, the data in form of a mapped point pattern is produced by exact measurements and thus the point pattern is observed without noise. For example, in forestry this is reasonable if the exact location of each tree was measured. However, exact measurements are a rather expensive method of data collection. More economical methods are usually available but, unfortunately, often lead to noisy data. For instance, aerial photographs are an economic way of mapping large areas of forest. Template matching as described in Larsen and Rudemo [13] may then be used to determine possible tree locations even for photographs obtained under off-nadir viewing angles. However, the irregular shape of tree tops, varying light conditions and motion due to wind lead to noisy observations of the tree locations. The template matching procedure therefore cannot determine the exact locations of the trees. Further noise may be introduced by not observing certain trees or by identifying a tree where there is none.

In mathematical terms, we may describe the mechanism which introduces noise into our observations as the combined effect of random and systematic displacements, censoring, thinning and/or superpositioning. Based on these point process operations a model which describes the degradation of our observations can be defined, see [1, 2]. Given the noisy observations together with the true tree locations we may estimate the parameters of the degradation model. In Dralle and Rudemo [1, 2] such a noise model is used for iterative least square estimation which also yields a pairing of true and observed points. Lund and Rudemo [15] 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 in a deterministic algorithm which performs approximate maximum likelihood estimation of the parameters for the disturbance model.

The conditional likelihood of the observations given the true locations may also be used within a Bayesian approach, see [14]. Here inference is based solely on the observed point pattern without knowledge of the true tree locations. A prior distribution for the true locations together with the likelihood of the observations given the true point pattern yields a posterior distribution for the true point process given the noisy observations. As common for many point process models, the posterior distribution of the true tree locations given the observed locations is not amenable to analytical examination. However, using Markov Chain Monte Carlo (MCMC) we may sample the posterior distribution and use these samples within Monte Carlo tests. In [14] 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 [23, 3]. In this paper we present a perfect simulation algorithm which allows for the exact sampling of the posterior distribution in [14]. We apply a method called *dominated Coupling From The Past*, see [9, 10, 11, 7] to a spatial birth-and-death process and thus produce an exact sample of the target posterior distribution.

This posterior distribution specifies 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 nor for the conditional distribution of the marks given the location of the points we cannot first sample the location of the points and then sample the marks. Thus our method is an illustration on how to produce exact samples from marked point process for which the marks cannot be sampled a posteriori.

In the following we first present the disturbance 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 (Section 2). A conventional simulation algorithm for this problem based on a spatial birth-and-death process is described in Section 3. Section 4 contains an introduction to the perfect simulation method called Coupling From The Past (CFTP). We then develop the perfect simulation algorithm for our problem in Section 5 and finally present the results thus obtained in Sections 6 and 7.

2 The posterior model

2.1 The degradation 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 sampling window A and $Y \subset A$ describe the observed point pattern. The degradation model is defined through the following point process operations:

1. *Thinning:*
   Some points of the true pattern are not observed, this is modelled using a thinning procedure. Each point $X_i \in A$ of the true point pattern is independently thinned with probability $1 - p(X_i)$ and retained with probability $p(X_i)$. If a point X_i is thinned then it does not lead to an observation point $Y_j \in Y$.

2. *Displacement:*
   We assume that the position of each tree derived from the aerial photograph through template matching suffers from a displacement. This displacement may be random

or systematic. We model the observed location Y_j of the true position X_i using a probability density $k(\cdot|X_i)$ with respect to the Lebesgue measure on $\mathbb{R}^2$.

3. *Censoring:*
   Some of the displaced locations may fall outside the observation window in which case they are not observed. These points are simply censored in our degradation model. Thus, censoring of the observed location of a true location X_i occurs with probability

$$p(X_i) \int_{A^c} k(y|X_i)dy,$$

   that is, if a point is retained and displaced outside the observation window A.

4. *Superposition:*
   Due to noise the template matching procedure may detect a tree location where there is actually no tree. These 'ghost' points are modelled by an independent superimposed Poisson process of intensity $h(\cdot|X)$.

To specify the above model we make the following distributional assumptions.

1. We assume homogeneous thinning, that is $p(X_i) \equiv p$.

2. Furthermore, we assume that the superimposed Poisson process is homogeneous and so $h(\cdot|X) \equiv \alpha$.

3. Finally, we assume that the displacement follows a 2-dimensional Normal distribution $N(X_i + \mu, \Sigma)$ where $\mu = (\mu_1, \mu_2)$ is the systematic error and

$$\Sigma \quad = \quad \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$$

   is the covariance of the error.

Under the above assumptions the degradation model is thus 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 a degradation model as described in the previous section, the conditional distribution of the observed locations given the true locations was derived in [15]. 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 specifies 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.

Let $|I|$ denote the cardinality of the set I. Furthermore, let $\mathcal{P}(M_1, N_1)$ be the set of bijections from the finite set M_1 into the set N_1 which is of the same cardinality as M_1. Let $M = \{1, \ldots, m\}$ and $N = \{1, \ldots, n\}$ and set

$$V_{m,n} \quad = \quad \Big\{(M_1, N_1, g) \;\mid\; g \in \mathcal{P}(M_1, N_1),\; M_1 \subseteq M,\; N_1 \subseteq N,\; |N_1| = |M_1|\Big\}.$$

We define $V = \cup_{m=0}^{\infty} \cup_{n=0}^{\infty} V_{m,n}$ and equip V with the σ-algebra consisting of all subsets. Then $S = (M_1, N_1, g) \in V$ is a matching and has the following interpretation:

1. The point $X_i, i \in M_1$ is observed as $Y_{g(i)}$.

2. The points $(Y_j, j \in N \backslash N_1)$ are 'ghost' points.

3. Due to thinning or censoring, the points $(X_i, i \in M \backslash M_1)$ are not observed.

Later on we 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 [15] 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 the first factor

$$L_1 \quad = \quad p^{|M_1|} \prod_{i \in M_1} k(y_{g(i)} | x_i, \mu, \Sigma),$$

corresponds to retained and displaced locations falling inside the sampling window. The second factor

$$L_2 \quad = \quad \prod_{i \in M \backslash M_1} \Big[p \int_{A^c} k(y|x_i) dy + (1-p) \Big].$$

derives from locations that were either thinned or retained and censored. The third factor

$$L_3 \quad = \quad \alpha^{|N \backslash N_1|} \exp\Big((1-\alpha) |A|_d \Big)$$

is due to 'ghost' points and finally

$$L_4 \quad = \quad \mathbb{I}_{[s \in V_{|x|,|y|}]}$$

is an indicator function which ensures that $s = (M_1, N_1, g)$ is a matching for x and y.

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

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

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 point process describing the true tree locations. We consider the following three choices.

1. *Poisson process:*
   A Poisson process describes complete spatial randomness, that is the absence of any interaction, whether repulsive or attractive. We may regard the Poisson process as a 'non-informative' prior as it only contains information about the intensity of the true tree locations. The density of a homogeneous Poisson process of intensity β with respect to a unit rate, homogeneous Poisson process is given by

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

   Here $n(x)$ is the number of points in x and the parameter β determines the intensity of the process as $\beta |A|_d$ is the expected number of points of the process.

2. *Strauss process:*
   The Strauss process [25] is a common choice for modelling repulsive point patterns. In our context it is reasonable to assume repulsive interaction as the forest in consideration was subjected to a thinning experiment. The Strauss process is defined by the density

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

   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 $n(x)$ counts the number of points in the pattern x and $s(x)$ the number of pairs of neighbour points, that is pairs of points which are less than a distance R apart. Notice that the interaction radius in this point process model is restricted to the distance R.

3. *Markov point process with logistic interaction function:*
   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

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

   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 $H(R) = 0.5$. Note that $H(r) < 1$ and so specifies a repulsive interaction. As the distance r increases $H(r)$ tends to one which means the strength of repulsion reduces with the increasing distance between two points.

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

$$\lambda(x, \xi) \quad = \quad \begin{cases} \frac{f(x \cup \{\xi\})}{f(x)} & \text{for } f(x) > 0 \\ 0 & \text{otherwise,} \end{cases}$$

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

$$\lambda(x, \xi) \quad = \quad \beta\ \gamma^{t(x,\xi)}$$

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 which is defined by the density

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

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

3 Markov Chain Monte Carlo

To produce samples of Markov point processes with varying number of points two alternative MCMC methods are commonly used: Metropolis-Hastings algorithms [5, 19] and spatial birth-and-death processes [21, 18]. A description of a Metropolis-Hastings type algorithm for the target posterior density in (2) can be found in [14]. 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 perfect Metropolis-Hastings algorithm for locally stable Markov point processes in [11].

We begin with a short introduction to spatial birth-and-death processes and then develop a spatial birth-and-death process which converges to our target distribution.

3.1 Spatial birth-and-death processes

Spatial birth-and-death processes are spatio-temporal point processes. At any given time the spatial birth-and-death process forms a point pattern in the observation window A. This pattern changes at distinct time instances. The change can be either a *birth*, that is a point is added to the current pattern, or a *death*, that is a point is deleted from the current point configuration. Since these changes depend only on the current point pattern, spatial birth-and-death processes are continuous-time Markov jump processes.

Spatial birth-and-death processes can be characterised by a birth and a death rate. The birth rate is a measurable function $b : \Omega \times A \to [0, \infty)$ such that $\int_B b(x, \xi) d\xi < \infty$ for all bounded Borel subsets B of A. Here the state space Ω is the family of finite point patterns on A. Given the current configuration x at time t, the probability of a birth in B during the short time interval $[t, t + s)$ is given by $s \int_B b(x, \xi) d\xi + o(s)$.

The death rate is a measurable function $d : \Omega \times A \to [0, \infty)$. Given the current configuration $x \cup \{\xi\}$ at time t, the probability that ξ is deleted during a time interval $[t, t + s)$ is given by $s\ d(x, \xi) + o(s)$. Conditions on b and d which ensure the existence and ergodicity of the spatial birth-and-death process can be found in [21]. The rate of convergence of ergodic spatial birth-and-death processes is examined in [18]. A description of a simulation procedure for spatial birth-and-death processes can be found in [24].

A spatial birth-and-death process has invariant distribution with density f and is time-reversible if the detailed balance condition

$$f(x)\ b(x, \xi) \quad = \quad f(x \cup \{\xi\})\ d(x, \xi) \tag{3}$$

is satisfied, where $f(x \cup \{\xi\})$ is assumed to be positive.

To sample from the target density f often the death rate is chosen at unit rate, that is $d(x, \xi) \equiv 1$, and the birth rate as $b(x, \xi) = \lambda(x, \xi)$, which is the Papangelou conditional intensity of f. The simulation evolves from $X_t = x$ as follows: generate the waiting time until the next event as an exponential variate of mean $1/(\int_A b(x, \xi)\, d\xi + n(x))$. With probability $n(x)/(\int_A b(x, \xi)\, d\xi + n(x))$ the event is the death of a point chosen uniformly among the points in x, and with probability $\int_A b(x, \xi)\, d\xi/(\int_A b(x, \xi)\, d\xi + n(x))$ it is the birth of a point ξ distributed according to the density which is proportional to $b(x, \cdot)$.

3.2 The target spatial birth-and-death process

We now determine a birth and a death rate which satisfy the detailed balance condition given in (3) for the target density in (2). Essentially, our spatial birth-and-death process will consist of two types of points. 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. Thus the states of the chain are marked point patterns. The mark indicates whether a point is unmatched or matched and in the latter case it specifies the matched observation point. These marks are given by the matching s. Points are born either as matched or unmatched points. A matched point can only become an unmatched point if it dies and is reborn as unmatched. The analogous holds for unmatched points. Changes in the current matching s occur when a matched point is born or dies.

Detailed balance holds if 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

$$b\Big(\ (x,s),\ (x\cup\{\xi\},s)\ \Big) \quad = \quad \lambda(x,\xi)\ \Big[p\int_{A^c} k(y|\xi)dy + (1-p)\Big], \tag{4}$$

where $\lambda(x,\xi)$ is the Papangelou conditional intensity of the prior model.

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

$$b\Big(\ (x,s),\ (x\cup\{\xi\},s\cup s(\xi))\ \Big) \quad = \quad \lambda(x,\xi)\ k(y_{s(\xi)}|\xi)\ \frac{p}{\alpha}\ \mathbb{I}_{[s\cup s(\xi)\in V_{\lceil x\cup\{\xi\}\rceil,\lceil y\rceil}]} \tag{5}$$

for each free y-point. 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. In particular matched points can only be born if there is an unmatched observation point. The overall birth rate of a matched point ξ can be derived by marginalisation:

$$b\Big(\ x,\ (x\cup\{\xi\})\ \Big) \quad = \quad \lambda(x,\xi)\ \frac{p}{\alpha}\ \sum_{j\in N\setminus N_1} k(y_j|\xi).$$

In the following we use $X_t = (x_t, s_t)$ to denote the spatial birth-and-death process with the above rates. Here x_t is a finite point pattern and s_t is the associated matching.

4 Coupling From The Past

Our perfect simulation algorithm is based on dominated Coupling from the Past [9, 10, 11, 7], an extension of the original Coupling From The Past (CFTP) algorithm by Propp and Wilson [23]. We begin with a short introduction to CFTP for finite Markov chains and then present briefly the ideas of dominated Coupling From The Past. This is followed by a section which develops a dominated Coupling From The Past algorithm for the posterior density given in (2).

4.1 Coupling From The Past for a finite Markov chain

We present the Coupling From The Past algorithm using a simple example; further details can be found in [23, 22].

Consider the Markov chain X on the integers $\{1, \ldots, 4\}$ with transition matrix

$$P \quad = \quad \begin{pmatrix} 0 & 0 & 1/2 & 1/2 \\ 0 & 1/2 & 1/2 & 0 \\ 1/2 & 1/2 & 0 & 0 \\ 1/2 & 1/2 & 0 & 0 \end{pmatrix}$$

We may simulate sample paths of X using a fair coin. Whenever the coin comes up heads the chain makes one of the following transitions depending on the current state of the chain:

$$1 \to 4, \qquad 2 \to 3, \qquad 3 \to 2, \qquad 4 \to 2.$$

Alternatively, when the coin comes up tails the chain makes one of the transitions:

$$1 \to 3, \qquad 2 \to 2, \qquad 3 \to 1, \qquad 4 \to 1.$$

Now, suppose we start a path in each of the four initial states and simulate them simultaneously using the same coin toss C_j for the transition at time j in all four paths. In this way we produce coupled paths of X started in all initial states. Observe that from time to time two paths meet in the same state. Then these two paths merge, we say they coalesce. We can show that in almost surely finite time all paths coalesce into one and we call this time the time of complete coalescence.

Coupling From The Past samples the equilibrium distribution π of X using the following simple, iterative procedure. In iteration k we go back to time -2^k and perform the following steps:

1. If $k = 1$ we toss a fair coin independently twice producing the realisations C_{-2} and C_{-1}. If $k > 1$ we toss a fair coin independently 2^{k-1} times yielding coin toss realisations $C_{-2^k}, C_{-2^k+1}, \ldots, C_{-2^{k-1}-1}$.

2. If $k > 1$ we prepend the new coin toss realisations to the sequence of coin toss realisations from the previous iterations, thus obtaining the sequence C_{-2^k}, C_{-2^k+1}, $\ldots, C_{-1}$.

3. We start a path at time -2^k in each initial state and simulate the paths until time 0 using the coin toss realisations $C_{-2^k}, C_{-2^k+1}, \ldots, C_{-1}$. Thus the state of each path at time $-j+1$ is determined by the state at time $-j$ and coin toss realisation C_{-j}.

4. At time 0 we check whether all paths are in the same state, that is if all paths have coalesced. If so, then their common state is a sample from π. If not, then we perform another iteration of this algorithm.

Notice the order in which we make use of the coin toss realisations. It is important that we reuse previously sampled realisations and that we *pre*pend any newly sampled ones. Only then is the output of the algorithm an exact sample from the equilibrium distribution. Figure 1 shows an iteration of the above CFTP algorithm in which coalescence occurs.

Suppose we reach complete coalescence in iteration L but nevertheless perform another iteration $L+1$. The reader may check that in iteration $L+1$ we again obtain complete coalescence at time 0 and furthermore, the common state of the paths at time 0 is the same as the common state at time 0 in the previous iteration. This occurs because of the order in which we construct our coin toss sequence. As complete coalescence is reached in almost surely finite time, L is also almost surely finite.

Here is a simple proof why the above algorithm works. Suppose we perform k iterations of the above procedure. Now consider the kth iteration. Let $Y^{(k)}$ denote the state at time 0 of the path started in state 1. Then $Y^{(k)}$ has the distribution $P^N(1,\cdot)$, where $N = 2^k$ and P^N is the N-step transition matrix. Due to the ergodicity of X we know that $P^N(1,\cdot) \to \pi$ as $N \to \infty$. At the same time we know that $Y^{(k)} = Y^{(L)}$ for $k \geq L$, where L is defined as above. It follows that $Y^{(L)}$ has distribution π.

The example illustrates the basic steps of CFTP:

1. Sample the randomness needed to simulate the target chain from time $-T$ to time 0.

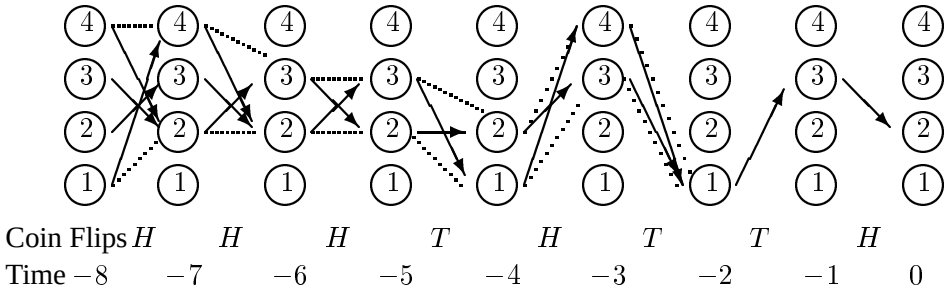

Figure 1: A CFTP iteration in which the coupled sample paths of X have coalesced by time 0. We use the notation H for heads and T for tails. Observe the anti-monotonicity of the transitions. The dotted lines show the minimal and the maximal path. Complete coalescence is achieved at time -2. However, we continue to simulate up to time 0.

2. Start a path of the target chain in each initial state and simulate them till time 0 using the sampled randomness.

3. Check whether complete coalescence has been achieved. If all paths have coalesced into one, then their common state at time 0 is a perfect sample.

4. If coalescence has not been achieved yet, go backwards in time till $-T-S$ and sample the randomness needed to simulate paths from time $-T-S$ to time $-T$. Prepend the newly sampled randomness to the previously sampled randomness and proceed with step 2.

When developing a CFTP algorithm it may be helpful to think in terms of the following heuristic. Suppose we could start the chain X at time $-\infty$ and simulate it up to time 0. Then, due to the ergodicity of X, the heuristic suggests that the state of the infinite time simulation at time 0 has the equilibrium distribution. Now consider the kth iteration of the CFTP algorithm. We assume that the randomness used in this iteration (in our example $C_{-2^k}, \ldots, C_{-1}$) is also used in the final 2^k transitions of the infinite time simulation. Suppose that coalescence is achieved in the kth iteration. As the infinite time simulation has to be in one of the initial states at time -2^k, at time 0 it has to be in the common state which the coalesced paths take at time 0.

For large state spaces monitoring the coalescence of paths started in all initial states will be prohibitively expensive. However, it may be possible to construct two paths, we call them the minimal and the maximal path, which do not neccessarily evolve according to the dynamics of the target chain but which sandwich between them the paths started in all initial states. Thus coalescence of these two paths implies complete coalescence of the paths started in all initial states. Let us return to our example on the integers $\{1,2,3,4\}$. Suppose $f(x,C)$ describes the state we reach if we make a transition from state x according to the coin toss C. The function f together with the random variable C is called a transition rule. Close inspection shows that f is montone decreasing (or anti-monotone) in its first argument. This fact can be exploited to monitor coalescence efficiently using a maximal chain and a minimal chain as follows. In the kth iteration we set the initial states as

$$X_{-2^k}^{\max} = 4 \quad \text{and} \quad X_{-2^k}^{\min} = 1$$

and then evolve the two paths till time 0 using the rule

$$X_{-j+1}^{\max} = f(X_{-j}^{\min}, C_{-j}) \quad \text{and} \quad X_{-j+1}^{\min} = f(X_{-j}^{\max}, C_{-j}).$$

Thus the minimal and the maximal chain evolve as a two-component chain in which the updates of the one component are made according to the current configuration of the other component. We call this construction, which was first used in [10] and further examined in [6], a *cross-over*. In Figure 1 the reader can see how the minimal and the maximal path sandwich between them the paths started from all initial states.

4.2 Dominated Coupling From The Past

Foss and Tweedie [4] showed that the existence of a CFTP algorithm as developed in [23] which produces output in almost surely finite time is equivalent to the Markov chain being uniformly ergodic. However, many Markov chains of interest, in particular many spatial birth-and-death processes, are not uniformly but only geometrically ergodic. We may still be able to produce perfect samples from such chains, but we will need to use an extension of CFTP which is called *dominated Coupling From The Past* or *Coupling Into And From The Past*, see [9, 10, 11, 7].

In terms of the heuristic of an infinite time simulation dominated Coupling From The Past is based on the following idea. Suppose we can find random bounds on the state which the infinite time simulation takes at any fixed time. Now, say at time -2^k we start coupled paths of the target chain in all the initial states which are within the random bounds at that time. We then evolve these paths till time 0 using the same randomness as the infinite time simulation. Suppose these paths have coalesced into one at time 0, then our heuristic suggests that their common state at time 0 is also the state of the infinite time simulation and thus has the equilibrium distribution.

For spatial birth-and-death processes we have a natural lower bound, the empty set. Upper random bounds can be found by constructing a spatial birth-and-death process D which stochastically dominates the target chain X so that given a path of D we can produce a path of X which at any time is bounded above by D. We can easily find such a process D by choosing a spatial birth-and-death process with the same death rate as X but a higher birth rate. We can then produce a path of X from a path of D by censoring some of the births in D. The following section will describe this construction in more detail for the target chain in Section 3.2.

Perfect simulation for locally stable point processes is discussed in [11]. Our method shows how to produce samples of marked point processes for which it is not possible first to produce a perfect sample of the unmarked process and then to sample the marks. Our paper carefully describes an implementation of dominated Coupling From The Past for a marked point process such that the algorithm samples the joint distribution of points and marks.

5 Perfect simulation of the posterior density

5.1 The dominating process

Recall that the target spatial birth-and-death process X has two types of points, matched and unmatched. Both types have unit death rates and their birth rates are given in (4) and (5). Similar to X the dominating process D is a spatial birth-and-death process whose states are marked point patterns (z, q). Like X it consists of matched and unmatched points. Both types of points have unit death rates in D. The birth rates of D, however, are higher than in X. Let $\lambda^* = \sup_{x,\xi} \lambda(x, \xi)$ where $\lambda(x, \xi)$ is the Papangelou conditional intensity of our prior model. For all three prior models λ^* is well defined as all processes are locally stable, that is their Papangelou conditional intensity is uniformly bounded.

Unmatched points in D have a birth rate given by

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

Thus the unmatched point component in D forms a time-reversible spatial birth-and-death process whose stationary distribution is an inhomogeneous Poisson process with intensity measure $\lambda^*\ f(\xi)$. We may sample the stationary distribution by sampling a point pattern z according to a homogeneous Poisson process of intensity λ^* and then retaining a point $\xi \in z$ with probability $p\int_{A^c} k(y|\xi)dy + (1-p)$. This can be done by marking each point ξ with a uniform $(0,1)$ mark $U(\xi)$ and a point η sampled from $k(\cdot|\xi)$ and then retaining ξ if $U(\xi) \leq 1-p$ or if $U(\xi) > 1-p$ and $\eta \notin A$. By sampling the point η we avoid the calculation of the integral $\int_{A^c} k(y|\xi)dy$. However, the integral is close to zero and thus $\lambda^*\ f(\xi)$ is close to $\lambda^*\ (1-p)$. As this is much smaller than λ^* many points are thinned. Fortunately, we found that this had only a very negligible effect on the speed of our algorithm.

In the following we use an auxiliary variable Q which assigns to each matched point in D an observational 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 observational point. Thus the dominating process lives on a state space which is an augmentation of the state space of X. 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

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

where q is an assignment. The overall birth rate of a matched point ξ in D is obtained by marginalisation:

$$b_D\Big(\ z,\ z\cup\{\xi\}\ \Big) \quad = \quad \lambda^*\ \frac{p}{\alpha}\ \sum_{j\in N} k(y_j|\xi) \quad = \quad \lambda^*\ \frac{p}{\alpha}\ h(\xi).$$

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 $\lambda^* p\alpha^{-1}h(\xi)$. 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. So to sample the stationary distribution we sample for each observation point y_j a Poisson number of daughters with mean $\lambda^* p\alpha^{-1}$ normally distributed around the observation point with mean $y_j - \mu$ and covariance Σ. We then censor any of these daughters that fall outside the observation window. Finally, the assignment then marks each point in the cluster pattern with the index of its parent point.

As the stationary distributions of both of the two components of D are easy to sample we can simulate D in equilibrium. Moreover, if we start D in equilibrium we may produce a path of D on a time interval $[-T,0]$ by simulating D on $[0,T]$ and then simply reversing the path in time. This also provides us with an easy way of extending a given path of D on a time interval $[-T,0]$ to a path on $[-T-S,0]$. We just simulate D from time T forwards till time $T+S$ and then reverse the path in time.

5.2 The target chain

We may derive a path of the target chain X from a path of D on a time interval $[-T,0]$ as follows. We first mark each birth time t of D with a mark M_t which is uniform on the unit interval. At time $-T$ we start X in the empty set. Then we evolve X from time $-T$ to time 0 in synchrony with D according to the following rules:

1. Whenever we have a birth of an unmatched point ξ in D, we check whether the mark M_t of this birth time t satisfies

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

where X_{t-} is the configuration of X immediately prior to time t. If the above is satisfied, then the unmatched point ξ is also born in X at time t. Otherwise, there is no transition in X at time t.

2. Suppose at time t we have the birth of a matched point ξ in D which is assigned to observation point $y_{q(\xi)}$. Then we check whether the mark M_t of this birth time t satisfies

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

We furthermore check whether the observation point $y_{q(\xi)}$ is not matched in X_{t-}. If both conditions are satisfied, then the matched point ξ, which is matched with $y_{q(\xi)}$, is born in X at time t. Otherwise no transition takes place in X at time t.

3. Finally, consider a death time t of D. Here we simply check whether the point ξ whose death time is t is in X_{t-}. If so, it dies in X at time t. Otherwise X does not change at time t.

Note that we also could have evolved X according to the above set of rules if it had started in any admissable subset of the configuration of D at time $-T$. We call the marked point pattern (x, s) a subset of the marked point pattern (z, q) if

1. the unmatched points in (x, s) are a subset of the unmatched points in (z, q),
2. the matched points in (x, s) are a subset of the matched points in (z, q),
3. any point in (x, s) that according to s is assigned to the observation point y_j is also assigned to y_j by q.

We call (x, s) an admissable subset of (z, q) if s is a matching, that is it assigns at most one point to any observation point. Note that any subset of an admissable subset is also admissable. We denote the subset relation for marked point pattern by $\subseteq$ and the admissable subset relation by $\preceq$.

If we evolve X as above, then marginally it behaves like a spatial birth-and-death process with unit death rates and the following birth rates. Unmatched points are born at rate

$$b\Big(\ (x, s),\ (x \cup \{\xi\}, s)\ \Big) \quad = \quad \lambda^* \left[p \int_{A^c} k(y|\xi) dy \ + (1-p)\right] \ \times \ \frac{\lambda(x, \xi)}{\lambda^*}$$

which is identical to the rate given in (4). Matched points in the above construction are born at a rate

$$b\Big(\ (x, s),\ (x \cup \{\xi\}, s \cup s(\xi))\ \Big) = \lambda^* \frac{p}{\alpha} k(y_{s(\xi)}|\xi) \times \ \frac{\lambda(x, \xi)}{\lambda^*} \ \mathbb{I}_{[s \cup s(\xi) \in V_{|x \cup \{\xi\}|, |y|}]}$$

which is the same rate as in (5). Notice that if we only permit one matched point per observation point in D and D is currently in state (z, q), then X would have birth rate

$$\begin{aligned} \frac{p}{\alpha} \ \lambda(x, \xi) \ k(y_{s(\xi)}|\xi) \ \mathbb{I}_{[q \cup s(\xi) \in V_{|z \cup \{\xi\}|, |y|}]} \ \mathbb{I}_{[s \cup s(\xi) \in V_{|x \cup \{\xi\}|, |y|}]} \\ = \frac{p}{\alpha} \ \lambda(x, \xi) \ k(y_{s(\xi)}|\xi) \ \mathbb{I}_{[q \cup s(\xi) \in V_{|z \cup \{\xi\}|, |y|}]}. \end{aligned}$$

However this does not coincide with the target birth rate in (5). This is because we assume the configuration $X_t = (x, s)$ to be an admissable subset of $D_t = (z, q)$. Thus the number of unmatched observation points in (x, s) is higher than in (z, q) and so the restriction on $s \cup s(\xi)$ to be a matching is less stringent than the one that $q \cup q(\xi)$ is a matching.

Suppose that at time $-T$ the chain X is started in an admissable subset of the configuration D_{-T}. Then points in X are only born if they are born in D and points which die in D also die in X and so the configuration of X at any time $t \geq -T$ is also an admissable subset of D_t.

In order to get the intuition on how to sample X in equilibrium using dominated Coupling From The Past consider again the heuristic of an infinite time simulation of X. Suppose the infinite time simulation of X is derived from an infinite time simulation of D. Then the configuration of the infinite time simulation of X is at any time an admissable subset of the current configuration of the infinite time simulation of D. If we sample a path of D in equilbrium on the time interval $[-T, 0]$ then we heuristically may interpret this path as a realisation of the infinite time simulation of D. But then, if we start a path of X at time $-T$ in every admissable subset of the configuration D_{-T} and they all coalesce by time 0 then their common state at time 0 is the state of the infinite time simulation of X at time 0 and thus has the target equilibrium distribution. Of course, this is just a heuristic; a formal proof follows in Section 5.4.

It would be rather expensive to start a path of X at time $-T$ in every admissable subset of D_{-T}. Instead, similar to the finite state space case, we will construct a minimal and a maximal path whose coalescence implies the complete coalescence of the paths started in every admissable subset of D at the starting time.

5.3 The maximal and the minimal process

Consider the acceptance rules for births in X. For unmatched points we check whether $\lambda^* M_t \leq \lambda(x, \xi)$. Clearly, the larger the Papangelou conditional intensity $\lambda(x, \xi)$ 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 $\lambda(x, \xi) \geq \lambda(z, \xi)$ 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 the first is a subset of the latter. Now, if $s(\xi) = q(\xi)$ then

$$\mathbb{I}_{[s \cup s(\xi) \in V_{|x \cup \{\xi\}|, |y|}]} \quad \geq \quad \mathbb{I}_{[q \cup q(\xi) \in V_{|z \cup \{\xi\}|, |y|}]}$$

because the unmatched observation points in (x, s) are a superset of the unmatched observation points in (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 can use a cross-over to define a maximal and a minimal chain. Suppose we have a stationary realisation of D on $[-T, 0]$. Then we produce a realisation of the minimal and the maximal chain as follows. Let $X^{\max(-T)}$ and $X^{\min(-T)}$ denote the maximal respectively the minimal chain started at time $-T$. We set

$$X_{-T}^{\max(-T)} \quad = \quad D_{-T} \qquad \text{and} \qquad X_{-T}^{\min(-T)} \quad = \quad \emptyset.$$

We then evolve the two chains as follows.

1. Whenever we have a birth of an unmatched point ξ in D, we check whether the mark M_t of this birth time t satisfies

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

in which case ξ is born at time t in $X^{\max(-T)}$ and $X^{\min(-T)}$. If the above condition is not satisfied we check whether

$$M_t \quad \leq \quad \frac{\lambda(X_{t-}^{\min(-T)}, \xi)}{\lambda^*}.$$

Note that this is a less stringent condition. If it holds then the unmatched point is born at time t only in $X^{\max(-T)}$. The minimal process does not have a transition. If the latter condition is not satisfied then neither the maximal process nor the minimal process change at time t.

2. Suppose that at time t we have the birth of a matched point ξ in D which is assigned to the observation point $y_{q(\xi)}$. We first check whether

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

and furthermore whether the observation point $y_{q(\xi)}$ is not matched in $X_{t-}^{\max(-T)}$. If both conditions are satisfied then the matched point ξ is born at time t both in the maximal and the minimal chain where it is matched to $y_{q(\xi)}$. If not both of the conditions are satisfied then we check whether

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

and, moreover, whether the observation point $y_{q(\xi)}$ is not matched in $X_{t-}^{\min(-T)}$. If these two conditions are satisfied, then the unmatched point ξ is born in the maximal path at time t but not in the minimal path. Note that this procedure allows the maximal 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 maximal or the minimal chain.

3. Finally, consider a death time t of D. Here we simply check whether the point ξ whose death time is t is in the maximal respectively the minimal path. Then the point ξ dies in either path subject to its existence.

As mentioned before, the maximal process like the dominating chain can have more than one matched point for a given observation point. This is motivated by the following. We would like the configuration of the maximal process to be a superset of the configurations of all paths of X started in an admissable subset of D and evolved according to D. Depending on their initial configuration these paths will have a point matched to a particular observation point at different times and these times may overlap. Thus the union of all the configurations of all these paths at a given time may contain more than one matched point per observation point. This is why we also allow such configurations in the maximal process. The minimal process, however, has only ever one matched point for any observation point.

The above construction of the minimal and the maximal process ensures that

$$X_t^{\min(-T)} \quad \subseteq \quad X_t^{\max(-T)} \quad \subseteq \quad D_t \qquad \text{for all } t \in [-T, 0]. \tag{6}$$

This follows from the fact that transitions in the minimal and the maximal chain are a subset of transitions in D, and, moreover, that births in D are more often censored in $X_t^{\min(-T)}$ than in $X_t^{\max(-T)}$.

Given a realisation of D on $[-T-S, 0]$ the maximal and the minimal process have the following funneling property [8]:

$$X_t^{\min(-T)} \subseteq X_t^{\min(-T-S)} \subseteq X_t^{\max(-T-S)} \subseteq X_t^{\max(-T)} \quad \text{for all } t \in [-T, 0]. \qquad (7)$$

The above statement clearly holds for $t = -T$ as

$$\emptyset \quad \subseteq \quad X_t^{\min(-T-S)} \quad \subseteq \quad X_t^{\max(-T-S)} \quad \subseteq \quad D_{-T}.$$

Moreover, if $X_{t-}^{\min(-T)} \subseteq X_{t-}^{\min(-T-S)}$ then, due to the anti-monotonicity of the acceptance rule for births, a birth is more often accepted in $X_t^{\max(-T)}$ than in $X_t^{\max(-T-S)}$. Similarly if $X_{t-}^{\max(-T)} \supseteq X_{t-}^{\max(-T-S)}$ then a birth is less often accepted in $X_t^{\min(-T)}$ than in $X_t^{\min(-T-S)}$ and so births preserve the ordering between the four processes. As a death of a point in D induces the death of the point in any of the four processes subject to its existence it follows that deaths also preserve the partial ordering and so statement (7) holds.

The construction of the maximal and the minimal process is such that once their paths meet they remain identical, that is if

$$X_t^{\min(-T)} \quad = \quad X_t^{\max(-T)} \qquad \text{for some } t \in [-T, 0]$$

then

$$X_u^{\min(-T)} \quad = \quad X_u^{\max(-T)} \qquad \text{for all } u \in [t, 0].$$

Notice that because the minimal path is always in an admissable subset of the dominating chain, the maximal path also has to be in an admissable subset when it coalesces with the minimal path. Once the two paths coalesce the evolution construction for the minimal and the maximal process reduces to the same construction as in Section 5.2 and so the two processes then have the same birth and death rates as the target chain. A sufficient condition for $X_t^{\min(-T)}$ and $X_t^{\max(-T)}$ to coincide is that $D_t = \emptyset$. Define the coalescence time T_C as

$$T_C \quad = \quad \min\{T \geq 0 : \ X_0^{\min(-T)} = X_0^{\max(-T)}\}$$

then it follows that $T_C \leq T_D$ where $T_D = \min\{t \geq 0 : \ D_{-t} = \emptyset\}$. As T_D is almost surely finite it follows that T_C is also almost surely finite. In practice, T_C is usually much smaller than T_D.

Our aim in constructing the minimal and the maximal process was to enable the efficient monitoring of the paths of the target chain started at time $-T$ in all admissable subsets of D_{-T} and evolved in accordance to D as described in Section 5.2. Thus we have to make sure that the paths of the minimal and the maximal process started at time $-T$ sandwich between them all the relevant paths of the target chain. Let X^{-T} be the path the target chain started in state $(x, s) \preceq D_{-T}$. Then

$$\emptyset \quad = \quad X_{-T}^{\min(-T)} \quad \subseteq \quad X_{-T}^{-T} \quad \subseteq \quad X_{-T}^{\max(-T)} \quad = \quad D_{-T}.$$

Now observe that if

$$X_{t-}^{\min(-T)} \quad \subseteq \quad X_{t-}^{-T} \quad \subseteq \quad X_{t-}^{\max(-T)}$$

then

$$\lambda(X_{t-}^{\min(-T)}, \xi) \quad \geq \quad \lambda(X_{t-}^{-T}, \xi) \quad \geq \quad \lambda(X_{t-}^{\max(-T)}, \xi).$$

Moreover, the free observation points in $X_{t-}^{\min(-T)}$ are a superset of the free observation points in X_{t-}^{-T} which themselves are a superset of those in $X_{t-}^{\max(-T)}$. By examining the acceptance rule for births we deduce that any point that is born in the path of the target chain is also born in the path of the maximal process and any point that is born in the path of the minimal process is also born in the path of the target chain. Thus the partial ordering between the three processes is preserved by any transition and the following sandwiching property holds

$$X_t^{\min(-T)} \quad \subseteq \quad X_t^{-T} \quad \subseteq \quad X_t^{\max(-T)}. \tag{8}$$

5.4 The perfect simulation algorithm

The perfect simulation algorithm now works as follows. We simulate a stationary version of D on $[-T, 0]$. Then we simulate the maximal and the minimal process as described in Section 5.3 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 D on $[-T, 0]$ to a stationary path on $[-T - S, 0]$. Again, we simulate the maximal and the minimal process, now on the time interval $[-T - S, 0]$, and check for coalescence. We proceed in this iterative manner until the minimal and the maximal path have coalesced at time 0 and then return their common state as a sample from the target posterior density. Recall that

$$T_C \quad = \quad \min\{T \geq 0 : \; X_0^{\min(-T)} = X_0^{\max(-T)}\}$$

is almost surely finite and so we deduce that the above procedure produces output in almost surely finite time when S is uniformly (across iterations) bounded away from zero.

Usually a doubling strategy as described in [23] is used. This means that the time S by which we extend the dominating chain backwards in time in each iteration is equal to the time T already gone back. We have also adopted the doubling strategy in our algorithm. A complete pseudo-code description for the CFTP algorithm can be found in Appendix A.

Theorem 5.1 *The constructed CFTP algorithm produces exact samples of the posterior density given in (2).*

The following proof is a special case of the proof in [11, Theorem 3.1].

Proof: We know that T_C is almost surely finite and so the limit of $X_0^{\max(-T)}$ as $T \to \infty$ exists. Note that due to the funneling property (7) we know that

$$\lim_{T\to\infty} X_0^{\min(-T)} \quad = \quad \lim_{T\to\infty} X_0^{\max(-T)} \quad = \quad X_0^{\max(-T_C)}.$$

Suppose X^{-T} denotes a path of the target chain started at time $-T$ in the empty set and evolved coupled to D. Then the sandwiching property (8) ensures that

$$\lim_{T\to\infty} X_0^{-T} \quad = \quad \lim_{T\to\infty} X_0^{\max(-T)} \quad = \quad X_0^{\max(-T_C)}.$$

Now suppose we start the target chain X at time 0 in the empty set and evolve it forwards till time T when it is in state X_T. Then $\mathcal{L}(X_T) = \mathcal{L}(X_0^{-T})$, where $\mathcal{L}(Y)$ denotes the distribution of Y. The ergodicity of X implies that

$$\lim_{T\to\infty} \mathcal{L}(X_0^{-T}) \quad = \quad \lim_{T\to\infty} \mathcal{L}(X_T) \quad = \quad \pi,$$

where π is the target posterior distribution as given in (2). It follows that $X_0^{\max(-T_C)} = \lim_{T\to\infty} X_0^{-T}$ has the equilibrium distribution. □

6 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 [1, 2] 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 [13]. This data set together with the 'true' tree top positions $X = (X_1, \ldots, X_m)$ with $m = 171$ was analysed in [15] with focus on the conditional likelihood $L(Y|X)$. In the present paper we study the posterior distribution of X given the observation of Y. We assume the knowledge of the prior distribution of X and of the conditional likelihood $L(Y|X)$ but not of the true positions of the X-points. The X- and Y-points are shown in Figure 2.

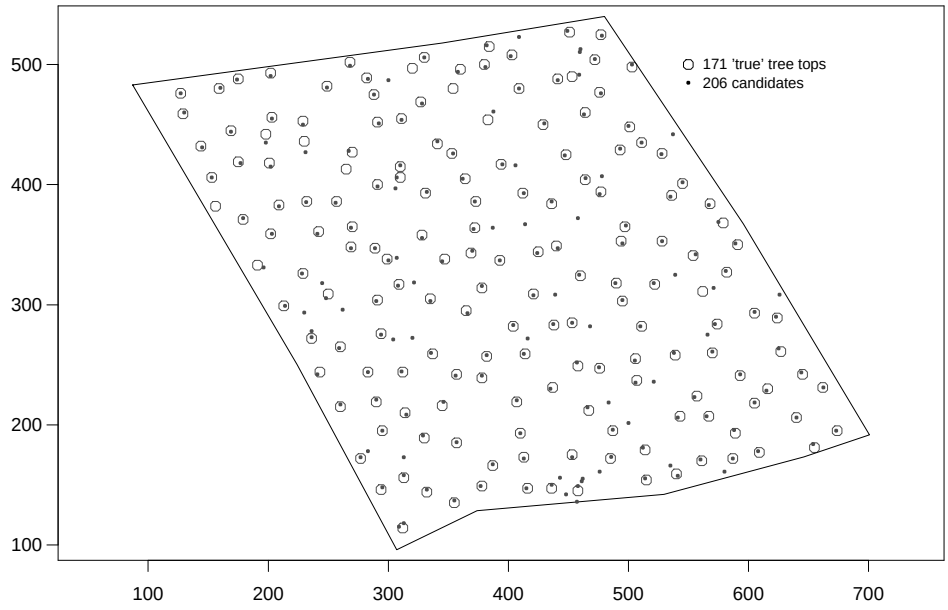

Figure 2: 171 X-points (centres of circles) corresponding to 'true' tree tops and 206 Y-points (dots) corresponding to template matching. The area of the delineated subplot is 4 454 m^2, and the unit of the axes is linear pixel size, 0.15 m. In the practical calculations we have used the convex border for convenience, and have thus increased the area with 2.12% to 164314.8 pixels from 160898.5 pixels. This increase is mainly in the lower left corner.

The analysis in [15] yielded the parameter estimates summarized in Table 1 which we used to specify the conditional likelihood $L(Y|X)$ in this paper.

Lund *et al.* [14] 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.

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.

7 Discussion

In the present paper we have shown an example of how to produce exact samples of a marked Markov point process whose marks cannot be sampled a posteriori. This means that it is not possible to sample first the unmarked point process and subsequently the marks.

As an example, Figure 3 shows a perfect sample from the posterior distribution where the prior has a logistic interaction function with the parameters $\beta|A|_d = 175$, $R = -3$ and $\phi = 0.462$. This results in $H(-3) = 0.5$ and $H(0) = 0.8$. This simulation took 72 minutes to produce output.

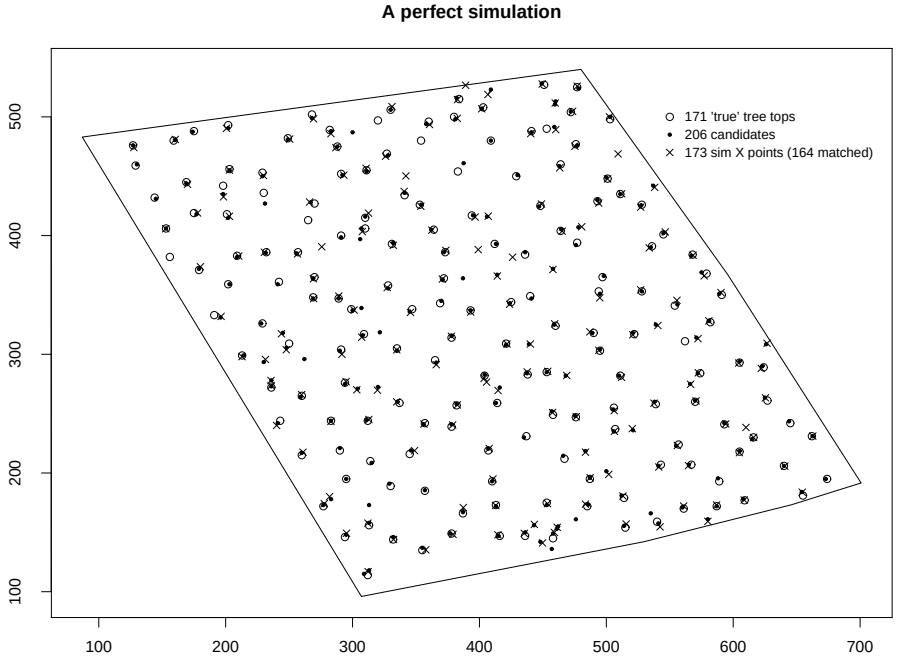

Figure 3: A perfect sample from the posterior distribution where the prior distribution has a logistic interaction function with parameters $\beta|A|_d = 175$, $H(-3) = 0.5$, $R = -3$, and $H(0) = 0.8$.

Our algorithm uses dominated Coupling From The Past applied to a spatial birth-and-death process. A dominating process is constructed by augmenting the state space of the target chain. As an illustration of the CFTP algorithm our Figure 4 shows the number of points in the dominating chain as well as in the maximal and in the minimal processes.

Given a perfect sample of the posterior distribution we may now continue to perform a full Bayesian analysis similar to Lund *et al.* [14]. Thus our algorithm is an illustration of how to use CFTP in a real life statistical inference problem. For other examples see [20, 16, 17].

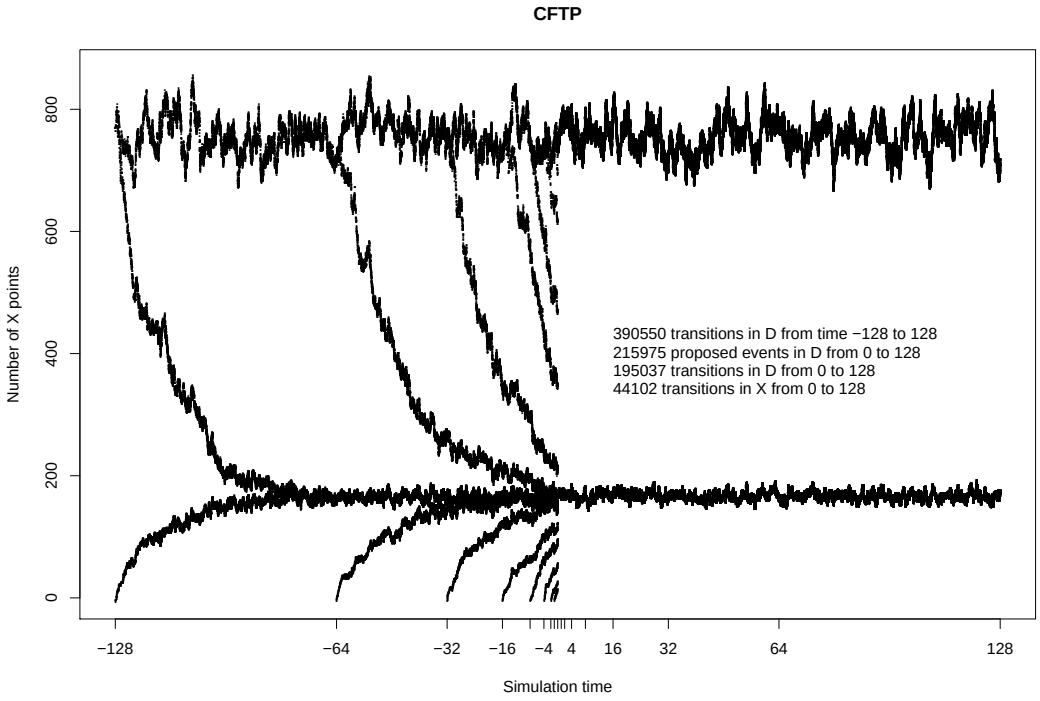

Figure 4: The number of points in the dominating chain, and in the upper and lower processes. The dominating chain and the target chain are continued forwards from time 0. We see that the coalescence time T_C is between 64 and 128. The figure specifies the number of transitions of some of the paths shown. 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 it configuration.

We would have liked to use our CFTP algorithm for priors with the same parameters as in [14]. Unfortunately, our algorithm turned out to be too slow to be useful for the priors in [14] as the current implementation produces samples in reasonable time only for less repulsive priors. For instance, a logistic prior with parameters $\beta|A|_d = 530$, $H(0) = 0.05$, and $R = 15$, which has much stronger repulsion than the one used for Figure 3, was used in [14]. The reason why our algorithm does not produce output in reasonable time for such a prior is as follows. With this more repulsive prior the number of X-points attached to each Y-point in the dominating chain has a mean of $\approx \lambda^* p \alpha^{-1} = 11.1$. This causes the difference between the dominating chain and the target chain to become so large that coalescence of the CFTP algorithm is not achieved in reasonable time. In order for the upper and lower processes to coalesce we have to add points in the minimal process and delete (and not re-add) points in the maximal process. However, as long as the lower process is in the empty set most of the points proposed by the dominating chain are added to the upper process. On the other hand, it is very difficult to add points to the lower process because the upper process (still similar to the dominating process) contains far too many points points compared to the desired target density. Specifically, a Y-point has to have no associated X-points in the upper chain in order to add a matched point in the lower process. If the mean number of matched points in the dominating chain per Y-point is 11, then the probability of having 0 matched X-points is as small as 1.67×10^{-5}.

Typically, if we use a prior that is more repulsive than the one used for Figure 3 and Figure 4 we observe that the minimal and the maximal process approach each other quickly. However, after an initial "burn in" period they do not get any closer but rather behave like stationary processes at certain distinct levels.

Van Zwet [26] in collaboration with Van Lieshout developed an iterative CFTP algorithm which samples a conditional Boolean model and is more efficient than the analogous algorithm developed in [12]. We have investigated whether similar ideas could be used in our setting. Unfortunately, we found that the method was not amenable to Markov point processes.

The main cause of the inefficiency of our algorithm is that we allow the dominating process to have multiple matched points for each observation point. We suspect that the number of matched points per observation point may be reduced by reusing previously sampled matched points. We are planning to investigate this in future research.

Acknowledgement:
Elke Thönnes gratefully acknowledges the support by the EU TMR network ERB-FMRX-CT96-0096 and the Stochastic Centre at Chalmers University.

A Appendix: Pseudo-code for the algorithm

This appendix describes the perfect simulation algorithm in detailed pseudo-code.

A.1 Dominating process

We give a pseudocode description of how to simulate the dominating process given the parameter $\theta = (p, \alpha, \mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$ and the noisy observation $Y = y$. We start simulating at time 0 and simulate backwards in time till time $-T$. The first step is to simulate the initial pattern and the associated birth times. The type m is a matched point, the type u stands for an unmatched point. Furthermore we write $|A|$ for the area of A and set $N = \{1, \ldots, n\}$.

Comments within the program are marked with the sign %. In the algorithm the realisation of the dominating process is specified by a multi-dimensional variable composed of $D.type, D.loc, D.life, D.birth$ coding the type, location, life time and birth time of each point. The variable $D.s$ specifies the matched observation point, where appropriate. Each birth time is marked with $D.mark$. `Poisson(`a`)` produces a Poisson variable of mean a and `Exponential(`a`)` produces an exponential variable of rate a that is of mean a^{-1}. Finally, the procedure `Normal(`z`)` produces a point which is normal with mean $z - \mu$ and covariance Σ.

Algorithm (1)

`Initial(`$\lambda^\star, |A|_d, \theta, y$`,seed):`

```
    set.seed(seed)
    % Initial matched and total number of points
        M_1(0) ~ Poisson(n λ* p α^{-1})
        M(0) ~ Poisson(λ* |A|_d)
        M(0) ← M(0) + M_1(0)
    % Marks, locations, life- and birth-times of initial points.
        for i = 1 to M_1(0)
            D.type(i) ← m
            D.s(i) ~ Uniform(N)
            D.loc(i) ~ Normal(y_{D.s(i)})
            D.life(i) ~ Exponential(1)
            D.birth(i) ← −D.life
            D.mark(i) ~ Uniform([0,1])
            if D.loc(i) ∉ A, then delete this point.
        for i = M_1(0) + 1 to M(0)
            ξ ~ Uniform(A)
            q ~ Uniform([0,1])
            z ~ Normal(ξ) + 2μ
            if (q ≤ (1 − p)) or ((q > 1 − p) and (z ∉ A))
                then D.type(i) ← u
                D.loc(i) ← ξ
                D.life(i) ~ Exponential(1)
                D.birth(i) ← −D.life
                D.mark(i) ~ Uniform([0,1])
        return(D)
```

The algorithm `D-Extend` may be used to extend an existing simulation of D on $[0, -T]$ backwards until some time $-T - S$. Specifically, you may call it with $T = 0$ and the initial patterns produced as above. The variable `Event1` records the next death time of an unmatched point, the variable `Event2` records the next death time of a matched point. The birth time is then the death time minus the exponential lifetime. The variable `nextEvent` is the waiting time between two successive death events.

Algorithm (2)

`D-Extend(`$\lambda^\star|A|_d, \theta, y, T, S, D$`,Event1,Event2):`

```
% Evolution of unmatched points
    i ← highest index in D
    event ← Event1
    while ( event > −T − S )
        i ← i + 1
        ξ ∼ Uniform(A)
        q ∼ Uniform([0, 1])
        z ∼ Normal(ξ) + 2μ
        if (q ≤ (1 − p)) or ((q > 1 − p) and (z ∉ A))
            then D.type(i) ← u
            D.loc(i) ← ξ
            D.life(i) ∼ Exponential(1)
            D.birth(i) ← event−D.life(i)
            D.mark(i) ∼ Uniform([0.1])
        nextEvent ∼ Exponential(λ⋆|A|_d)
        event ← event - nextEvent
    Event1 ← event
% Evolution of matched points
    event ← Event2
    while ( event > −T − S )
        i ← i + 1
        D.type(i) ← m
        D.life(i) ∼ Exponential(1)
        D.birth(i) ← event −D.life(i)
        D.s(i) ∼ Uniform(N)
        D.loc(i) ∼ Normal(y_{D.s(i)})
        D.mark(i) ∼ Uniform([0, 1])
        if D.loc(i) ∉ A, then delete this point.
        nextEvent ∼ Exponential(n λ⋆ p α^{−1})
        event ← event - nextEvent
    Event2 ← event
    return(D,Event1,Event2)
```

A.2 The target chain

In the following we present an algorithm which produces a realisation of the target chain from a realisation of the dominating process. Thus the target chain is an adapted functional of the dominating process. For the following algorithm we assume that we have a realisation of the dominating process D on the time-interval $[-T, 0]$. The realisation is coded in an ordered list of events $-T < t_1 < t_2 < \ldots < t_n < 0$. For each event t_i we know whether it is a birth or a death and which point k of D it concerns, that is we know

$$D(k) \quad = \quad \Big(D.type(k), D.loc(k), D.life(k), D.birth(k), D.mark(k), D.s(k)\Big).$$

The target chain is started in the empty pattern.

Algorithm (3)

```
X-Evolve(D):
    % Derives a realisation of X from a realisation of D
    X ← ∅
    for i = 1 to n
        if (t_i = D.birth(k))
            if (D.type(k) = u) and (D.mark ≤ λ(X,D.loc(k))/λ*)
                then X ← X ∪ D(k)
            if (D.type(k) = m) and (D.mark ≤ λ(X,D.loc(k))/λ*)
                and (D.s(k) is index of unmatched y-point in X )
                    then X ← X ∪ D(k)
        if (t_i = D.birth(k) + D.life(k)) and (D(k) ∈ X)
            then X ← X\D(k)
    return(X)
```

The algorithm `X-Evolve` is not directly used in the perfect simulation algorithm, but is described here for completeness.

A.3 The minimal and the maximal process

Here are the algorithm for the minimal and the maximal process which are evolved according to a cross-over because we have a repulsive prior distribution. Coalescence of the two processes means that we can accept their common state at time 0 as an exact sample.

Algorithm (4)

```
MinMax-Evolve(D, λ*):
    % Produces a realisation of the minimal and maximal
    % process from a realisation of D
    MIN ← ∅
    MAX ← configuration of D at time −T
    for i = 1 to n
        if (t_i = D.birth(k)) and (D.type(k) = u)
            if (D.mark ≤ λ(MAX,D.loc(k))/λ*)
                then MIN ← MIN ∪ D(k) and MAX ← MAX ∪ D(k)
            else if (D.mark ≤ λ(MIN,D.loc(k))/λ*)
                then MAX ← MAX ∪ D(k)

        if (t_i = D.birth(k)) and (D.type(k) = m)
            if (D.mark ≤ λ(MAX,D.loc(k))/λ*) and (D.s(k) free in MAX )
                then MIN ← MIN ∪ D(k) and MAX ← MAX ∪ D(k)
            else if (D.mark ≤ λ(MIN,D.loc(k))/λ*) and (D.s(k) free in MIN )
                then MAX ← MAX ∪ D(k)

        if (t_i = D.birth(k) + D.life(k))
```

```
            if (D(k) ∈ MIN) then MIN ← MIN\D(k)
            if (D(k) ∈ MAX) then MAX ← MAX\D(k)
    return(MIN, MAX)
```

A.4 The CFTP algorithm

Algorithm (5)

```
CFTP(λ*, θ, y, seed):
    coalescence ← FALSE
    T ← 0
    λ*|A|_d ← λ*|A|
    D ← Initial(λ*|A|_d, θ, y, seed)
    Event1 ~ Exponential(λ*|A|_d)
    Event2 ~ Exponential(n λ* p α^{-1})
    Event1 ← - Event1
    Event2 ← - Event2
    while ( coalescence = FALSE )
        S ← max(T, 1)
        (D, Event1, Event2) ← D-Extend(λ*|A|_d, θ, y, T, S, D, Event1, Event2)
        (MIN, MAX) ← MinMax-Evolve(D, λ*)
        T ← T + S
        if (MIN = MAX) then coalescence ← TRUE
    return(MIN)
```

References

[1] Kim Dralle and Mats Rudemo, *Stem number estimation by kernel smoothing of aerial photos*, Canadian Journal of Forest Research **26** (1996), 1228–1236.

[2] ———, *Automatic estimation of individual tree positions from aerial photos*, Canadian Journal of Forest Research **27** (1997), 1728–1736.

[3] J. A. Fill, *An interruptible algorithm for exact sampling via Markov Chains*, Annals of Applied Probability **8** (1998), 131–162.

[4] S.G. Foss and R.L. Tweedie, *Perfect simulation and backward coupling*, Stochastic Models **14** (1998), 187–203.

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

[6] Olle Häggström and Karin Nelander, *Exact sampling from anti-monotone systems*, Statistica Neerlandica **52** (1998), 360–380.

[7] O. Häggström, M. N. M. van Lieshout, and J. Møller, *Characterisation results and Markov chain Monte Carlo algorithms including exact simulation for some spatial point*

processes, Report R-96-2040, Institute for Electronic Systems, Department of Mathematics and Computer Science, Aalborg University, Denmark, November 1996.

[8] Wilfrid S. Kendall and Jesper Møller, *Perfect Metropolis-Hastings simulation of locally stable point processes*, Preprint 347, University of Warwick, Department of Statistics, August 3 1999.

[9] W.S. Kendall, *Perfect simulation for spatial point processes*, Proc. ISI 51^{st} session, Istanbul (August 1997), vol. 3, 1997, Research report 308, University of Warwick, pp. 163–166.

[10] ———, *Perfect simulation for the area-interaction point process*, Probability Towards 2000 (New York) (L. Accardi and C.C. Heyde, eds.), Springer, 1998, pp. 218–234.

[11] W.S. Kendall and J. Møller, *Perfect Metropolis-Hastings simulation of locally stable point processes*, Research report 347, University of Warwick, 1999.

[12] W.S. Kendall and E. Thönnes, *Perfect simulation in stochastic geometry*, Pattern Recognition **32** (1999), 1569 – 1586.

[13] Morten Larsen and Mats Rudemo, *Optimizing templates for finding trees in aerial photographs*, Pattern Recognition Letters **19** (1998), 1153–1162.

[14] Jens Lund, Antti Penttinen, and Mats Rudemo, *Bayesian analysis of spatial point patters from noisy observations*, Preprint, Department of Mathematics and Physics, The Royal Veterinary and Agricultural University, 1999, Submitted.

[15] Jens Lund and Mats Rudemo, *Models for point processes observed with noise*, Report 10, Department of Mathematics and Physics, The Royal Veterinary and Agricultural University, November 4 1998, Submitted.

[16] Jesper Møller, *Perfect simulation of conditionally specified models*, Research Report 7, MaPhySto, May 1998, to appear in JRSS B.

[17] Jesper Møller and Geoff K. Nicholls, *Perfect simulation for sample-based inference*, preprint, August 4 1999.

[18] J. Møller, *On the rate of convergence of spatial birth-and-death processes*, Ann. Inst. Statist. Math. **41(3)** (1989), 565–581.

[19] ———, *Markov Chain Monte Carlo and Spatial Point Processes*, Stochastic Geometry: likelihood and computation (O.E. Barndorff-Nielsen, W.S. Kendall, and M.N.M. van Lieshout, eds.), Chapman & Hall/CRC, Boca Raton, 1999, pp. 141–172.

[20] D. J. Murdoch and P. J. Green, *Exact sampling from a continuous state space*, Scandinavian Journal of Statistics (1998?), to appear.

[21] C.J. Preston, *Spatial birth-and-death processes*, Bulletin of the International Statistical Institute **46** (1975), 371–391.

[22] James Propp and David Wilson, *Coupling from the past: a user's guide*, Microsurveys in Discrete Probability (D. Aldous and J. Propp, eds.), DIMACS Series in Discrete Mathematics and Theoretical Computer Science, vol. 41, American Mathematical Society, 1998, pp. 181–192.

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

[24] D. Stoyan, W.S. Kendall, and J. Mecke, *Stochastic Geometry and its applications*, second ed., Wiley Series in Probability and Statistics, John Wiley & Sons, Chichester, 1995.

[25] D.J. Strauss, *A model for clustering*, Biometrika **63** (1975), 467–475.

[26] Erik van Zwet, *Likelihood devices in spatial statistics*, Phd thesis, Universiteit Utrecht, The Netherlands, 1999.