Perfect simulation of point processes given noisy observations

By JENS LUND

Department of Mathematics and Physics,

The Royal Veterinary and Agricultural University,

Thorvaldsensvej 40, 1871 Frederiksberg C, Denmark

jlund@dina.kvl.dk

AND ELKE THÖNNES

Department of Statistics, University of Warwick,

Coventry CV4 7AL, United Kingdom

elke@stats.warwick.ac.uk

February 16, 2000

SUMMARY

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 process given a noisy observation. The algorithm is a perfect simulation method which applies dominated coupling from the past to a spatial birth-and-death process. Dominated coupling from the past is made amenable by augmenting the state space of the target chain. We discuss how to optimize this perfect simulation algorithm and how to use it in a statistical inference problem of pratical importance. We further describe the merits and the limitations of the method.

Some key words: Augmented state space; Bayesian inference; Disturbance; Noise; Perfect simulation; Point process.

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, which comes in the 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 & Rudemo (1998) 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 Dralle & Rudemo (1996; 1997). Given the noisy observations together with the true tree locations we may estimate the parameters of the degradation model. In Dralle & Rudemo (1996; 1997) such a noise model is used for iterative least square estimation which also yields a pairing of true and observed points. Lund & Rudemo (1999) 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 Lund et al. (1999). 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 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), 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 (2000), 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. Our paper carefully describes an implementation of dominated coupling from the past for the marked point process which samples the *joint* distribution of points and marks.

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 optimize it in Section 6. We finally present the results thus obtained in Section 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. We assume A to be a bounded measurable subset of $\mathbb{R}^d$. The degradation model is defined through the following point process operations:

1. *Thinning:* Some points from the true pattern are not observed. 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 each position suffers a displacement. 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' we may detect a location where there is actually none. These 'ghost' points are modelled by an independent superimposed Poisson process of intensity $h(\cdot|X)$.

To specify the above model we consider point processes in the plane, that is $d = 2$, and make the following distributional assumptions. (i) We assume homogeneous thinning, that is $p(X_i) \equiv p$. (ii) Furthermore, we assume that the superimposed Poisson process is homogeneous and so $h(\cdot|X) \equiv \alpha$. (iii) 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 Lund & Rudemo (1999). 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

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

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 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 (1999) the joint distribution of (Y, S) given $X = x$ and θ factorizes as

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

Here the first factor

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

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

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

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

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

is due to 'ghost' points and finally

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

is an indicator function which ensures that $s = (M_1, N_1, g)$ is a matching for x and y. Here we have used the notation $|A|_d$ for the d-dimensional volume of A.

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

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

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 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. The density of a homogeneous Poisson process of intensity β with respect to a unit rate, homogeneous Poisson process is given by

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

   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 (Strauss, 1975) 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

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

   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 $s(x)$ counts the number of pairs of neighbour points, that is pairs of points which are less than a distance R apart. Thus the interaction radius of the Strauss process 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

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

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 $H(R) = 0.5$. As H is increasing and $H(r) < 1$ this function specifies a repulsive interaction. Furthermore, the strength of repulsion 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 distribution 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). In the following we denote the state space of (X, S) by $\mathcal{X}$.

3. Markov chain Monte Carlo

3·1 *Simulation of point processes*

To produce samples of a point process with varying number of points two alternative Markov Chain Monte Carlo methods are commonly used: Metropolis-Hastings algorithms (Geyer & Møller, 1994; Møller, 1999) and spatial birth-and-death processes (Preston, 1975; Møller, 1989). 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 (2000).

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.

3·2 *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. These changes depend only on the current point pattern, and thus 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 Preston (1975). The rate of convergence is examined in Møller (1989). A description of a simulation procedure can be found in Stoyan et al. (1995).

A spatial birth-and-death process has the invariant distribution with the 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 $b(x,\cdot)$.

3·3 *The target spatial birth-and-death process*

We now determine a birth rate 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 living on the state space $\mathcal{X}$. The mark, which is given by the matching s, indicates whether a point is unmatched or matched and in the latter case it specifies the matched observation point. Points are born either as matched or 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 observed point y of index $s(\xi)$. The birth rate for matched points is given by

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

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:

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

where N_1 is the set of matched observation points. 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

4·1 *Coupling from the past for a finite Markov chain*

Our perfect simulation algorithm is based on dominated coupling from the past, see (Kendall, 1998; Kendall & Møller, 2000), an extension of the original coupling from the past algorithm in Propp & Wilson (1996). We begin with a short introduction to coupling from the past 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).

We describe the coupling from the past algorithm using a simple example; further details can be found in Propp & Wilson (1996; 1998).

Consider the Markov chain X on the integers $\{1,\dots,4\}$ 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:

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

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

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

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 coupling from the past algorithm in which coalescence occurs.

[Figure 1 about here.]

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. 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 at time -2^k. 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 the distribution π.

When developing a coupling from the past 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 suppose that complete coalescence is achieved in the kth iteration of the coupling from the past 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. 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

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

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

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

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 Kendall (1998) and further examined in Häggström & Nelander (1998), 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 & Tweedie (1998) showed that the existence of a coupling from the past algorithm as in Propp & Wilson (1996) which produces output in almost surely finite time implies that the target Markov chain is 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 coupling from the past which is called *dominated coupling from the past* or *coupling into and from the past*, see Kendall (1997), Kendall (1998), Kendall & Møller (2000), Häggström et al. (1999).

In terms of the heuristic of an infinite time simulation of X, 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 3·3.

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 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 $\lambda^* = \sup_{x,\xi} \lambda(x, \xi)$ 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

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

where l, which is defined by the last equality, is less than one. Due to this birth rate 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 $\lambda^*\, l(\xi)$. 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 $l(\xi) = p \int_{A^c} k(y|\xi)dy + (1-p)$. 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 $l(\xi)$ is close to $1-p$. As this is much smaller than one 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 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 $\mathcal{X}$ 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

$$b_D\Big(\ (z,q),\ (z\cup\{\xi\}, q\cup q(\xi))\ \Big) \quad = \quad \lambda^*\,\frac{p}{\alpha}\,k(y_{q(\xi)}|\xi), \tag{6}$$

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

$$b_D\Big(\ z,\ z\cup\{\xi\}\ \Big) \quad = \quad \lambda^*\,\frac{p}{\alpha}\sum_{j\in N} k(y_j|\xi) \quad = \quad \lambda^*\,\frac{p}{\alpha}\,g(\xi), \tag{7}$$

where g is defined by the last equality. 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 $\lambda^* p\alpha^{-1} g(\xi)$. 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. Hence 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 $[-2T, 0]$. We simply simulate D from time T forwards till time $2T$ 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 $M_t \leq \lambda(X_{t-}, \xi)/\lambda^*$, 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 $M_t \leq \lambda(X_{t-}, \xi)/\lambda^*$. 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.

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). Now suppose the set of unmatched observation points in $X_t = (x,s)$ is a superset of the unmatched observation points in $D_t = (z,q)$. If we would only permit one matched point per observation point in D, then X would have birth rate

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

However this does not coincide with the target birth rate in (5) because the restriction on $s \cup s(\xi)$ to be a matching is less stringent than the one on $q \cup q(\xi)$.

In the following we call the marked point pattern (x,s) a subset of the marked point pattern (z,q) if (i) the unmatched points in (x,s) are a subset of the unmatched points in (z,q), (ii) the matched points in (x,s) are a subset of the matched points in (z,q), (iii) 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 denote the subset relation for marked point pattern by $\subseteq$.

Note that we could have evolved X according to the above set of rules also if it had started in any state of the state space $\mathcal{X}$ which is a subset of the configuration of D at time $-T$. Suppose that at time $-T$ we start the chain X in such a state. 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 a subset of D_{-t}. Furthermore, our acceptance rules ensure that X remains in its state space $\mathcal{X}$.

5·3 *The maximal and the minimal process*

To obtain a perfect sample we may now proceed as follows. Given D on $[-T, 0]$ we start X at time $-T$ in every subset of D_{-T} that lies in $\mathcal{X}$. We then evolves each path coupled to the path of D until time 0 and then check whether all paths have coalesced. If so, their common state at time 0 is a perfect sample. If not then we need to extend D and its birth time marks backwards in time and repeat the above procedure. It would be rather expensive to monitor all these paths of X. 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 relevant paths of X.

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 $M_t \leq \lambda(X_{t-}^{\max(-T)}, \xi)/\lambda^*$, 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 $M_t \leq \lambda(X_{t-}^{\min(-T)}, \xi)/\lambda^*$. 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 $M_t \leq \lambda(X_{t-}^{\max(-T)}, \xi)/\lambda^*$ 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 $M_t \leq \lambda(X_{t-}^{\min(-T)}, \xi)/\lambda^*$ 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 and let it die 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 a 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 and thus lives in $\mathcal{X}$.

LEMMA 1. *1. For any path X^{-T} of the target chain started at time $-T$ in any subset $(x,s) \in \mathcal{X}$ of D_{-T} and evolved coupled to D we have the following sandwiching property*

$$X_t^{\min(-T)} \quad \subseteq \quad X_t^{-T} \quad \subseteq \quad X_t^{\max(-T)} \quad \subseteq \quad D_t \qquad \text{for all } t \in [-T, 0]. \tag{8}$$

2. *Given a realisation of D on $[-T-T', 0]$ the maximal and the minimal process have the following funneling property:*

$$X_t^{\min(-T)} \subseteq X_t^{\min(-T-T')} \subseteq X_t^{\max(-T-T')} \subseteq X_t^{\max(-T)} \quad \text{for all } t \in [-T, 0]. \tag{9}$$

3. *The construction of the maximal and the minimal process is such that once their paths meet they remain identical, that is if $X_t^{\min(-T)} = X_t^{\max(-T)}$ for some $t \in [-T, 0]$ then $X_u^{\min(-T)} = X_u^{\max(-T)}$ for all $u \in [t, 0]$.*

4. *The coalescence time $T_C = \inf\{T \geq 0 : \ X_0^{\min(-T)} = X_0^{\max(-T)}\}$ is almost surely finite.*

Proof First note that all the transitions of the target, the minimal and the maximal chain are a subset of transitions in D and that deaths clearly preserve the partial ordering. We now prove the individual statements.

1. The initial configuration of the tree chains are set such that we have

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

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

2. The statement clearly holds for $t = -T$ as $\emptyset \subseteq X_t^{\min(-T-T')} \subseteq X_t^{\max(-T-T')} \subseteq D_{-T}$. The statement can be shown for general $t \in [-T, 0]$ in a manner similar to the sandwiching property.

3. 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.

4. A sufficient condition for $X_t^{\min(-T)}$ and $X_t^{\max(-T)}$ to coincide is that $D_t = \emptyset$. We see that $T_C \leq T_D$ where $T_D = \inf\{t \geq 0 : D_{-t} = \emptyset\}$ and T_D is almost surely finite due to ergodicity of D. (In practice, T_C is usually much smaller than T_D.) □

Notice that because the minimal path is always in $\mathcal{X}$, the maximal path also has to be in $\mathcal{X}$ when it coalesces with the minimal path.

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 and its marks on $[-T, 0]$ to a stationary path on $[-2T, 0]$. Again, we simulate the maximal and the minimal process, now on the time interval $[-2T, 0]$, and check for coalescence. We proceed doubling the time interval 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. A complete pseudo-code description for the coupling from the past algorithm can be found in the Appendix.

The construction of a coupling from the past algorithm on an augmented state space is called *extended state space* coupling from the past in Cai & Kendall (1999). The following theorem certifies the correctness of our algorithm and is similar to the proof in Cai & Kendall (1999).

THEOREM 1. *The constructed coupling from the past algorithm stops in finite time almost surely and produces exact samples of the posterior density given in (2).*

Proof Recall that the coalescence time T_C is almost surely finite and hence the above procedure produces output in almost surely finite time.

Moreover, because T_C is almost surely finite the limit of $X_0^{\max(-T)}$ as $T \to \infty$ exists. Due to the funneling property (9) we now 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

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

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. Optimizing the algorithm

6·1 *Inefficiency of the algorithm*

The perfect simulation algorithm presented in the previous sections is rather inefficient, in particular for priors with strong repulsive interactions such as the ones used in Lund et al. (1999). This is mainly due to the fact that the dominating chain has multiple matched points for each observation point. For example for the logistic prior in Lund et al. (1999) with parameters $\beta|A|_d = 530$, $\phi = 0.196$, $R = 15$, that is $H(0) = 0.05$ and $H(15) = 0.5$, the number of points attached to each observation point in the dominating chain has a mean of approximately $\lambda^* p\alpha^{-1} = 11.1$. This causes the dominating chain and consequently the maximal chain to be much larger than the minimal and the target chain. Specifically, an observation point has to have no associated matched points in the maximal chain in order to add a matched point in the minimal process. If the mean number of matched points in the dominating chain per observation point is 11, then the probability of having no matched points is as small as 1.67×10^{-5}. As a result coalescence is not achieved within acceptable time.

6·2 *Better bounds on the Papangelou conditional intensity*

A first improvement is to use a higher lower bound on $\lambda(X_{t-}^{-T}, \xi)$ other than $\lambda(X_{t-}^{\max(-T)}, \xi)$. Let $B_t^{-T} = \{(y, r) \in \mathcal{X} : (y, r) \subseteq X_t^{\max(-T)}\}$ be the set of all subsets of $X_t^{\max(-T)}$ which lie in $\mathcal{X}$. We now may substitute $\lambda(X_{t-}^{\max(-T)}, \xi)$ by $\tilde{\lambda}(X_{t-}^{\max(-T)}, \xi) = \min\{\lambda(y, \xi) : (y, r) \in B_{t-}^{-T}\}$ without affecting any of the properties in Lemma 1. This is due to the fact that the target chain always has at most one point matched to each observation point. As this new bound on the Papangelou conditional intensity is larger, the coalescence time between minimal and maximal process will reduce. For repulsive pairwise interaction priors, we can find this new bound $\tilde{\lambda}(X_{t-}^{\max(-T)}, \xi)$ by inspecting for each $j \in N$ the cluster of points assigned to y_j. However computing this new bound is computationally more expensive and so the efficiency of the algorithm is not significantly improved.

6·3 *A coupling with faster coalescence*

To increase the efficiency of the coupling from the past algorithm it is essential to obtain coalescence of the relevant paths of the target chain as quickly as possible. For this reason we will consider a coupling between the target chain and a sligthly altered dominating process $\tilde{D}$. As a result each path of the target chain will still have the (marginal) birth and death rates as specified in (4) and (5). However, the death times of matched points between different paths are synchronised.

In the new dominating chain $\tilde{D}$, births of matched points and births of unmatched points occur in the same way as in D, see Section 5·1. 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.

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 $[-\tilde{T}, 0]$.

The target chain as well as the maximal and the minimal chain are now coupled to $\tilde{D}$ in a manner similar to our previous coupling. Whenever, a point is born in $\tilde{D}$ we check the conditions specified in Sections 5·2 and 5·3 to decide whether this point is permitted to be born in the minimal, the maximal or the target chain. Whenever an unmatched point or a cluster of matched points die, we let these points die in the minimal, the maximal and the target chain, subject to their existence in these chains.

Of course, we need to ensure that this coupling produces paths of the target chain with the correct birth and death rates. Clearly, the birth rate of matched or unmatched points as well as the death rate of unmatched points is unchanged. Thus we only need to ensure that matched points have a unit death rate in the target chain. But, due to the fact that the target chain has at most one point matched to each observation point and due to the properties of a Poisson counting process each matched point in the target chain is given a unit rate exponential lifetime. The only difference to the construction in Section 5·2 is that now the lifetimes of matched points in different paths of the target chain are no longer independent but synchronised. Lemma 1 and Theorem 1 still hold as before. However, the dominating chain $\tilde{D}$ has a higher probability than D of having no points assigned to an observation point y_j. As a consequence coalescence between the minimal and the maximal chain occurs in shorter time.

7. Results and Discussion

7·1 *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 (1999) 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 2.

[Figure 2 about here.]

To specify the conditional likelihood $L(Y|X)$ we used the parameter estimates from Lund & Rudemo (1999) which are summarized in Table 1.

[Table 1 about here.]

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.

7·2 *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 then 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 $\beta|A|_d = 225$, $R = 7$ and $\phi = 0.314$. This results in $H(0) = 0.1$ and $H(7) = 0.5$. The simulation took 16 minutes to produce output.

[Figure 3 about here.]

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 coupling from the past algorithm our Figure 4 shows the number of points in the dominating chain as well as in the maximal and in the minimal processes.

[Figure 4 about here.]

Given a perfect sample of the posterior distribution we may now continue to perform a full Bayesian analysis similar to Lund et al. (1999). Thus our algorithm is an illustration of how to use coupling from the past in a real life statistical inference problem. For other examples see Murdoch & Green (1998), Møller (1999), Møller & Nicholls (1999).

The spatial birth-and-death process construction described in Section 5·1 and 5·2 might of course also be used for conventional Markov Chain Monte Carlo as an alternative to the Metropolis-Hastings algorithm in Lund et al. (1999). In particular when the disturbances are large the present algorithm might have the advantage that it avoids the calculation of the integral $\int_{A^c} k(y|\xi)dy$.

The unoptimized version of the algorithm does not produce output within acceptable time for priors with repulsive interaction as strong as in Lund et al. (1999), see our comments in Section 6·1.

If we choose the prior with a logistic interaction function with the parameters $\beta|A|_d = 175$, $R = -3$ and $\phi = 0.462$, that is $H(-3) = 0.5$ and $H(0) = 0.8$, then the unoptimized version takes 72 minutes to produce output whereas the optimized version takes only 5 minutes. For priors whose interaction is as strong as in in Figure 3 the unoptimized version was not able to produce output in acceptable time.

Our algorithm works efficiently for realistic priors like the one in Figure 3 or for instance an almost hard core Strauss process with parameters $\beta|A|_d = 238$, $\gamma = 0.02$, and $R = 10$. However, even the optimized version of the algorithm is not efficient enough for the priors with very strong interaction like the ones in Lund et al. (1999).

Our algorithm is presently implemented in the interpreted language Python, and we would expect an increase of speed by converting the algorithm to a compiled language as for instance `C`. All quoted running times are on a PC with a 600MHz AMD Athlon processer running the Linux operating system and mounting the discs by NFS.

Acknowledgement

Elke Thönnes gratefully acknowledges 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 and Wilfrid Kendall for discussions which lead to the optimized version of the perfect simulation algorithm. We finally thank Mats Rudemo for helpful comments.

Appendix: Pseudo-code for the unoptimized algorithm

Dominating process

We give a pseudocode description of how to simulate the dominating process given the parameter $\theta = (p, \alpha, \mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$, the prior of X 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(λ*, |A|_d, θ, 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 $-T - T'$. Specifically, you may invoke 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(λ*|A|_d, θ, y, T, T', D, Event1, Event2):
  % Evolution of unmatched points
    i ← highest index in D
    event ← Event1
    while ( event > −T − T' )
        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 − T' )
        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)
```

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.

The minimal and the maximal process

Here is a description of the algorithm for the minimal and the maximal process which are evolved according to a cross-over to accomodate a repulsive prior distribution. Coalescence of the two processes implies 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 ← ∅
```

$MAX \leftarrow$ `configuration of` D `at time` $-T$

`for` $i = 1$ `to` n

`if` $\Big(t_i = D.birth(k)\Big)$ `and` $\Big(D.type(k) = u\Big)$

`if` $\Big(D.mark \leq \frac{\lambda(MAX, D.loc(k))}{\lambda^*}\Big)$

`then` $MIN \leftarrow MIN \cup D(k)$ `and` $MAX \leftarrow MAX \cup D(k)$

`else if` $\Big(D.mark \leq \frac{\lambda(MIN, D.loc(k))}{\lambda^*}\Big)$

`then` $MAX \leftarrow MAX \cup D(k)$

`if` $\Big(t_i = D.birth(k)\Big)$ `and` $\Big(D.type(k) = m\Big)$

`if` $\Big(D.mark \leq \frac{\lambda(MAX, D.loc(k))}{\lambda^*}\Big)$ `and` $\Big(D.s(k)$ `free in` $MAX\Big)$

`then` $MIN \leftarrow MIN \cup D(k)$ `and` $MAX \leftarrow MAX \cup D(k)$

`else if` $\Big(D.mark \leq \frac{\lambda(MIN, D.loc(k))}{\lambda^*}\Big)$ `and` $\Big(D.s(k)$ `free in` $MIN\Big)$

`then` $MAX \leftarrow MAX \cup D(k)$

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

`if` $\Big(D(k) \in MIN\Big)$ `then` $MIN \leftarrow MIN \backslash D(k)$

`if` $\Big(D(k) \in MAX\Big)$ `then` $MAX \leftarrow MAX \backslash D(k)$

`return(`MIN, MAX`)`

The coupling from the past algorithm

Algorithm (5)

`CFTP(`λ^*, θ, y`,seed):`

`coalescence` $\leftarrow$ `FALSE`

$T \leftarrow 0$

$\lambda^\star|A|_d \leftarrow \lambda^*|A|$

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

`Event1` $\sim$ `Exponential`$(\lambda^\star|A|_d)$

`Event2` $\sim$ `Exponential`$(n\ \lambda^\star\ p\ \alpha^{-1})$

`Event1` $\leftarrow$ `- Event1`

`Event2` $\leftarrow$ `- Event2`

```
while ( coalescence = FALSE )
    T' ← max(T, 1)
    (D, Event1, Event2) ← D-Extend(λ*|A|_d, θ, y, T, T', D, Event1, Event2)
    (MIN, MAX) ← MinMax-Evolve(D, λ*)
    T ← T + T'
    if (MIN = MAX) then coalescence ← TRUE
return(MIN)
```

References

Cai, Y. & Kendall, W. S. (1999). Perfect simulation for correlated Poisson random variables conditioned to be positive. Research Report 349, Department of Statistics, University of Warwick. Available at `http://www.warwick.ac.uk/statsdept/Staff/WSK/papers/349.ps.gz`.

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.

Foss, S. & Tweedie, R. (1998). Perfect simulation and backward coupling. *Stochastic Models* **14**, 187–203.

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

Häggström, O. & Nelander, K. (1998). Exact sampling from anti-monotone systems. *Statist. Neerlandica* **52**, 360–380.

Häggström, O., van Lieshout, M. N. M. & Møller, J. (1999). Characterisation results and Markov chain Monte Carlo algorithms including exact simulation for some spatial point processes. *Bernoulli* **5**, 641–658.

Kendall, W. (1997). Perfect simulation for spatial point processes. In *Proc. ISI* 51^{st} *session, Istanbul (August 1997)*, volume 3, pages 163–166.

KENDALL, W. (1998). Perfect simulation for the area-interaction point process. In Accardi, L. & Heyde, C., editors, *Probability Towards 2000*, pages 218–234, New York. Springer.

KENDALL, W. S. & MØLLER, J. (2000). Perfect Metropolis-Hastings simulation of locally stable point processes. Research Report 347, University of Warwick, Department of Statistics. Available from `http://www.warwick.ac.uk/statsdept/Staff/WSK/papers/347.ps.gz`.

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. Available from `ftp://ftp.dina.kvl.dk/pub/Staff/Jens.Lund/reconstruct/estx1.ps.gz`.

LUND, J. & RUDEMO, M. (1999). Models for point processes observed with noise. *Biometrika* To appear.

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

MØLLER, J. & NICHOLLS, G. K. (1999). Perfect simulation for sample-based inference. Preprint. Available from `http://www.scitec.auckland.ac.nz/~nicholls/linkfiles/papers/perfect.sim.temp1.ps.gz`.

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

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

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.

PROPP, J. & WILSON, D. (1998). Coupling from the past: a user's guide. In Aldous, D. & Propp, J., editors, *Microsurveys in Discrete Probability*, volume 41 of *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, pages 181–192. American

Mathematical Society.

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

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

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.

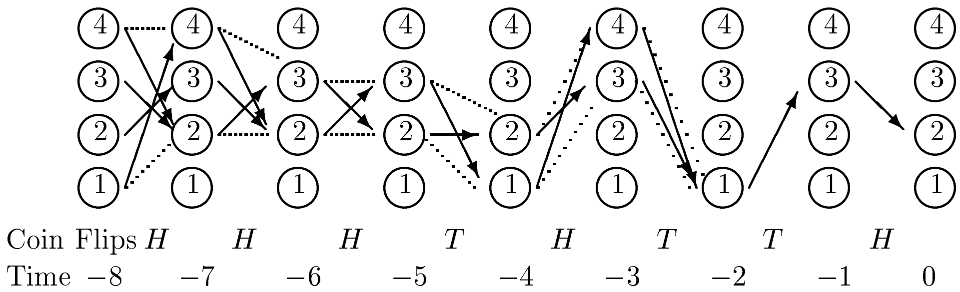

Figure 1: A coupling from the past iteration in which the coupled sample paths of X have coalesced by time 0. We use the notation H for heads and T for tails. Observe the anti-monotonicity of the transitions. The dotted lines show the minimal and the maximal path. Complete coalescence is achieved at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

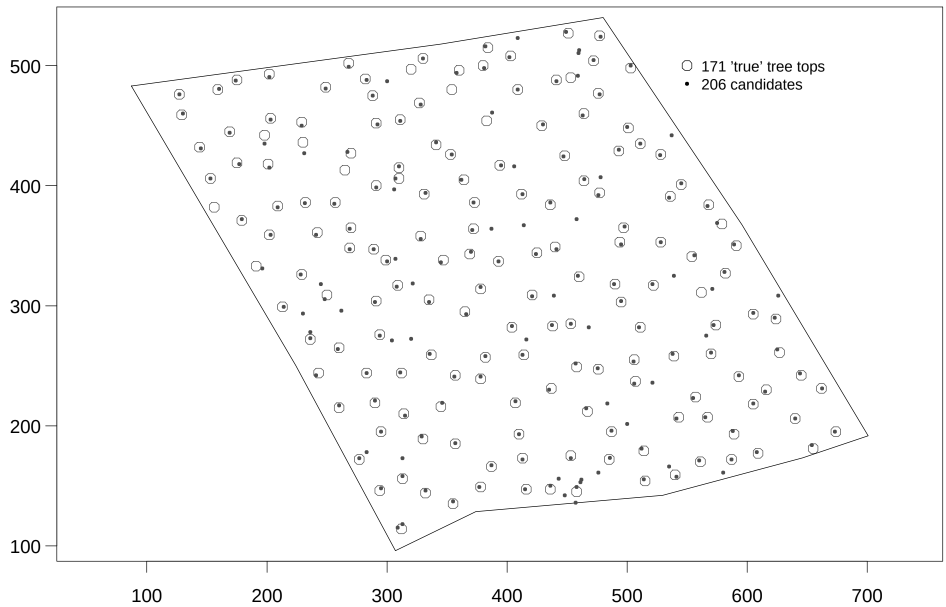

Figure 2: 171 X-points (centres of circles) corresponding to 'true' tree tops and 206 Y-points (dots) corresponding to template matching. The area of the delineated subplot is 4 454 m^2, and the unit of the axes is linear pixel size, 0.15m. In the practical calculations we have used the convex border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

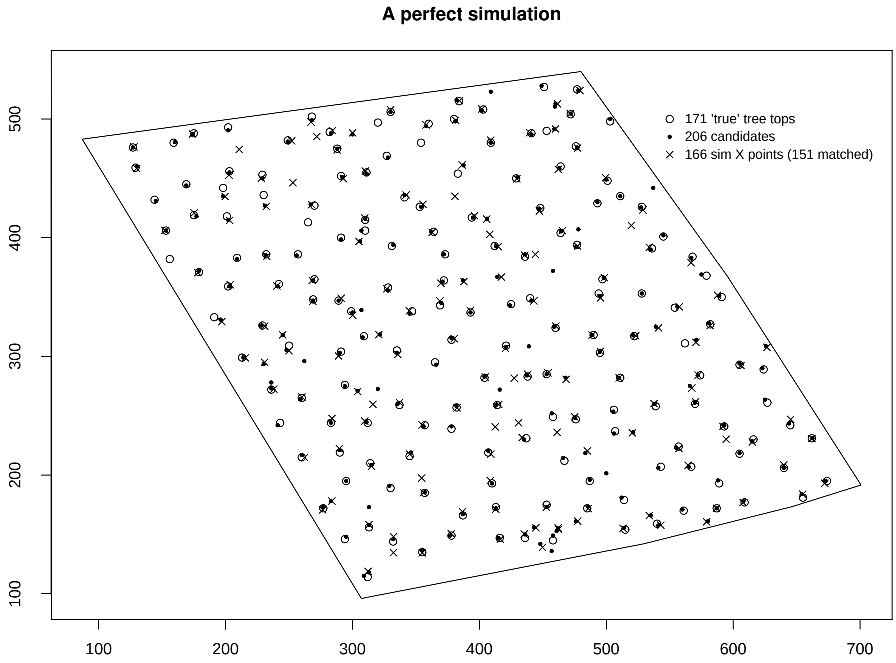

Figure 3: A perfect sample from the posterior distribution where the prior distribution has a logistic interaction function with parameters $\beta|A|_d = 225$, $\phi = 0.314$, $R = 7$, $H(7) = 0.5$ and $H(0) = 0.1$.

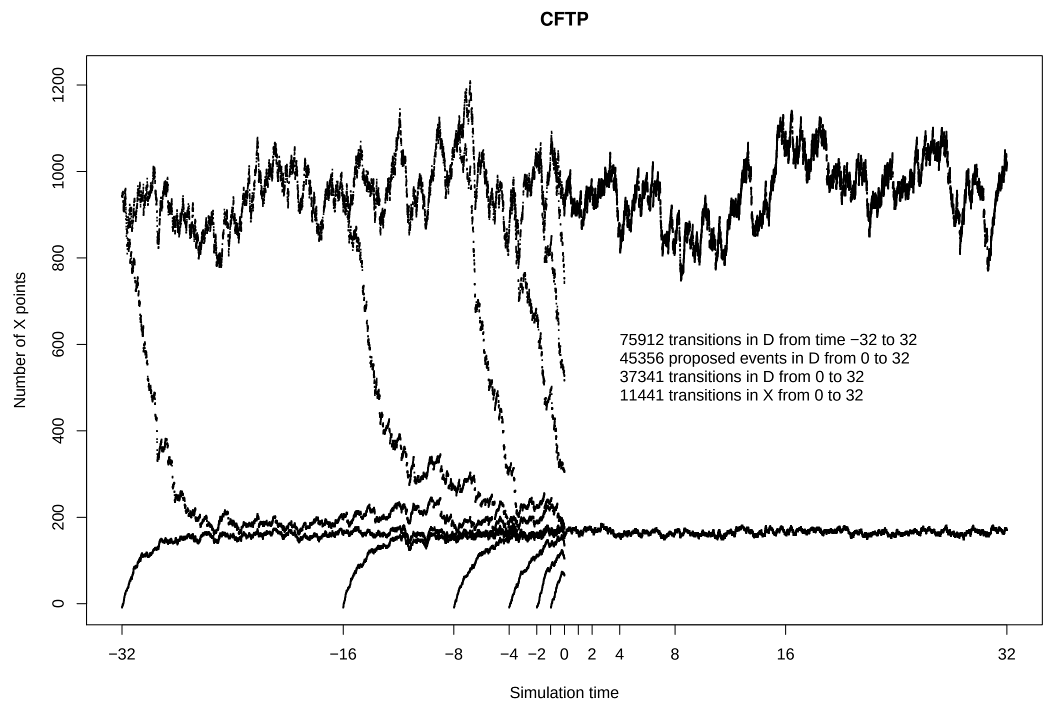

Figure 4: The number of points in the dominating chain, and in the upper and lower processes. The dominating chain and the target chain are continued forwards from time 0. We see that the coalescence time T_C is between 16 and 32. The figure specifies the number of transitions of some of the paths shown. 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.