Centrum voor Wiskunde en Informatica

REPORT*RAPPORT*

PNA

Probability, Networks and Algorithms

Probability, Networks and Algorithms

Perfect simulation for marked point processes

M.N.M. van Lieshout, R.S. Stoica

REPORT PNA-R0306 MAY 31, 2003

CWI is the National Research Institute for Mathematics and Computer Science. It is sponsored by the Netherlands Organization for Scientific Research (NWO).
CWI is a founding member of ERCIM, the European Research Consortium for Informatics and Mathematics.

CWI's research has a theme-oriented structure and is grouped into four clusters. Listed below are the names of the clusters and in parentheses their acronyms.

Probability, Networks and Algorithms (PNA)

Software Engineering (SEN)

Modelling, Analysis and Simulation (MAS)

Information Systems (INS)

P.O. Box 94079, 1090 GB Amsterdam (NL)
Kruislaan 413, 1098 SJ Amsterdam (NL)
Telephone +31 20 592 9333
Telefax +31 20 592 4199

ISSN 1386-3711

Perfect Simulation for Marked Point Processes

M.N.M. van Lieshout[1] and R.S. Stoica[2]

[1] *CWI*
P.O. Box 94079, 1090 GB Amsterdam, The Netherlands

[2] *Universitat Jaume I*
Campus Riu Sec, E-12071 Castelló de la Plana, Spain

ABSTRACT
This paper extends some recently proposed exact simulation algorithms for point processes to marked patterns and reports on a simulation study into the relative efficiency for a range of Markov marked point processes.

2000 Mathematics Subject Classification: 60G55, 62M40.

Keywords and Phrases: Exact simulation, Markov chain Monte Carlo simulation, Markov marked point process.

Note: Work carried out under project PNA4.3 'Stochastic Geometry'. This research was supported by NWO grant 'Inference for random sets' (613-03-045).

1. Introduction

One of the most exciting developments in computational statistics in recent years has been the introduction of exact (or perfect) simulation methods following the ground breaking paper by Propp and Wilson [30]. During the last two decades, statistical inference for complex models tended to be based on Markov Chain Monte Carlo (MCMC) techniques, as exemplified by the review papers in [2]. The difficulty with such methods is the need for careful burn-in and convergence diagnostics to assess whether the underlying Markov chain has reached its stationary distribution. Exact simulation methods in contrast determine by themselves during run time if and when equilibrium is reached, hence their appeal.

In their 1996 paper, Propp and Wilson presented exact samplers for a range of discrete distributions. Their method, coupling from the past, requires an order relation on the state space which has to be preserved by the transition kernel to be practical. Modifications that do not require such an order were studied in [17, 18] and further extended in [26]. All papers cited above use the Gibbs sampler dynamics. For specific models, faster convergence may be obtained by exploiting salient properties of the model. Examples include Fill and Huber's randomness recycler [10, 11].

Murdoch and Green [27] proposed coupling from the past algorithms for continuous state spaces. They do not rely on order relations nor on the Gibbs sampler. Instead, bounds and rejection sampling ideas are exploited. Rejection sampling is also the driving principle behind the FMMR method [8, 9], but, as the method is closely related to coupling from the past, monotonicity properties are helpful.

For point process models, MCMC methods are typically based on spatial birth-and-death processes [25, 29, 31] or the Metropolis-Hastings paradigm [13, 15, 28]. One may think of the first approach as the natural analogue of the Gibbs sampler, as both methods change a

point or component at a time based on the local characteristics of the model. Indeed, just as the Gibbs dynamics are amenable to coupling from the past in the discrete set-up, so are birth-and-death processes for spatial point patterns as demonstrated by Kendall [19]. In this context, the requirements [20] are that the target point process distribution is locally stable and, for practical purposes, that an order relation exists between point configurations that is preserved by the transition kernel. An alternative that does not rely on monotonicity properties is the clan of ancestors algorithm of Ferrari et al. [7]; instead its efficiency relies on a small interaction radius compared to the point intensity. It should be noted that both methods may be placed in the context of spatial jump processes, see [4].

As in the discrete case, exploiting model characteristics may help to improve the mixing properties of the process. For example, auxiliary variables were used in [16, 24, 34]. The interested reader is referred to the excellent web site

`http://dimacs.rutgers.edu/~dbwilson/exact`

maintained by D. Wilson for an annotated bibliography and up to date information including references to review papers and tutorials.

Here, we generalise the local characteristics based exact simulation algorithms proposed for point processes to marked patterns and carry out a simulation study into their relative efficiency for a range of Markov models with both discrete and continuous marks. A follow-up paper will deal with proposal-rejection alternatives, including Metropolis–Hastings sampling.

The plan of this paper is as follows. We shortly review basic facts on marked point processes in Section 2. Then, in Section 3 we discuss coupling from the past, turn to the clan of ancestors algorithm in Section 4, and next consider Gibbs sampling of a lattice approximation in Section 5. Results on speed of convergence are reported in Section 6. The paper closes with a critical evaluation.

2. Marked point processes

Let $K \subset \mathbb{R}^d$ be a compact subset of strictly positive Lebesgue measure $0 < \nu(K) < \infty$ and M a complete separable metric space. A *marked point process* Y with positions in K and marks in M is a point process on $K \times M$ such that the process of unmarked points is (locally) finite [6]. The realisations are of the form $\mathbf{y} = \{(k_1, m_1), \ldots (k_n, m_n)\}$ where $n \in \mathbb{N}_0$, $k_i \in K$ and $m_i \in M$ for all $i = 1, \ldots, n$.

Let ν_M be a probability measure on the Borel $\sigma-$algebra $\mathcal{B}(M)$. In this paper, we shall restrict attention to marked point processes that are absolutely continuous with respect to the distribution of a Poisson process on $K \times M$ with intensity measure $\nu \times \nu_M$. Thus, under the reference measure, points in a realisation of a unit rate Poisson process on K are given i.i.d. marks distributed according to ν_M.

Example 1. Widom–Rowlinson mixture model
The Widom–Rowlinson mixture model for penetrable spheres [35] has mark space $M = \{1, 2\}$ *and density*

$$f(\mathbf{y}) = \alpha \prod_{(k,m)\in\mathbf{y}} \beta_m \prod_{(u,1),(v,2)\in\mathbf{y}} \mathbf{1}\{\| u - v \| > R\} \tag{2.1}$$

with respect to a unit rate Poisson process labelled according to the symmetric Bernoulli

distribution. Hence, particles with a different label keep at least a distance R away from each other.

Example 2. Multi-type pairwise interaction process
Consider the mark space $M = \{1, \dots, I\}$ for some $I \in \mathbb{N}$ equipped with the uniform probability distribution ν_M. Multi-type pairwise interaction processes [1, 32] are defined by a density f with respect to the dominating Poisson process that is of the form

$$f(\mathbf{y}) = \alpha \prod_{(k,m)\in\mathbf{y}} \beta_m \prod_{(u,i)\neq(v,j)\in\mathbf{y}} \gamma_{ij}(\| u - v \|) \tag{2.2}$$

with the second product ranging over all distinct pairs of marked points. Here $\alpha > 0$ is the normalising constant, the scalars $\beta_m > 0$, $m \in M$, are intensity parameters, and for each pair of labels $i, j \in M$, $\gamma_{ij} : [0, \infty) \to [0, 1]$ is a measurable interaction function. We shall assume that $\gamma_{ij} \equiv \gamma_{ji}$ for all $i, j \in M$.

Example 3. Candy model
The Candy model [33, 23] is a line segment process. The segments are characterized by the position of their centre, their length $l \in [l_{\min}, l_{\max}]$ for some $0 < l_{\min} < l_{\max} < \infty$, and orientation $\theta \in [0, \pi)$. The orientation space is equipped with the complete metric $\rho(\theta, \theta') = \min\{|\theta - \theta'|, \pi - |\theta - \theta'|\}$ that identifies 0 and π. Hence the Candy model may be seen as a marked point process with marks in $M = [l_{\min}, l_{\max}] \times [0, \pi)$. It has density

$$f(\mathbf{y}) = \alpha\, \beta^{n(\mathbf{y})} \prod_{i=1}^{n(\mathbf{y})} \exp\left[\frac{l_i - l_{\max}}{l_{\max}}\right] \times \gamma_r^{n_r(\mathbf{y})}\, \gamma_o^{n_o(\mathbf{y})} \tag{2.3}$$

with respect to a unit rate Poisson process marked uniformly and independently. The model parameters are γ_r, $\gamma_o \in (0, 1)$ and $\beta > 0$. The sufficient statistics $n(\mathbf{y})$, $n_r(\mathbf{y})$ and $n_o(\mathbf{y})$ represent the total number of segments in $\mathbf{y}$, the number of pairs of segments crossing at too sharp an angle, and the number of pairs of segments that are disoriented. More formally, for a given $\delta \in (0, \pi/2)$, define the relation $\sim_r$ on $K \times M$ by

$$y = (k, l, \theta) \sim_r y' = (k', l', \theta') \Leftrightarrow \|k - k'\| \leq \max\{l, l'\}/2 \text{ and } \left|\rho(\theta, \theta') - \pi/2\right| > \delta.$$

Then $n_r(\mathbf{y})$ is the number of pairs of different points in $\mathbf{y}$ that are $\sim_r$-related. Moreover, let the influence zone $Z(y)$ of a marked point $y = (k, l, \theta)$ be the union of balls with radius $l/4$ around the endpoints, and define the relation $\sim_o$ on $K \times M$ as follows: $y \sim_o y' \Leftrightarrow \|k - k'\| > \max\{l, l'\}/2$ and either exactly one endpoint of y is a member of $Z(y')$ or exactly one endpoint of y' is a member of $Z(y)$. Then $n_o(\mathbf{y})$ is the number of $\sim_o$ neighbour pairs in $\mathbf{y}$ with the property that $\rho(\theta, \theta') > \tau$. Generalisations of (2.3) may be obtained by distinguishing several types of connection between the segments.

The *Papangelou conditional intensity* of a marked point process with density f is defined as

$$\lambda((k, m); \mathbf{y}) := \frac{f(\mathbf{y} \cup \{(k, m)\})}{f(\mathbf{y})}$$

whenever $f(\mathbf{y}) > 0$ and arbitrarily (say 0) on the null set $\{\mathbf{y} : f(\mathbf{y}) = 0\}$.

Henceforth, we shall assume the following properties to hold:

- the density is *hereditary*, that is, $f(\mathbf{y}) > 0$ implies $f(\mathbf{y}') > 0$ for all $\mathbf{y}' \subseteq \mathbf{y}$;
- the density is *locally stable*, that is the Papangelou conditional intensity is bounded from above by some positive, finite constant Λ.

It is left to the reader to verify that the models introduced in Examples 1–3 satisfy these assumptions.

3. Coupling from the past for marked point processes

The study of exact simulation algorithms was pioneered in the seminal paper [30] by Propp and Wilson. Their coupling from the past (CFTP) algorithm was developed with discrete probability distributions on graphs in mind, for example the critical Ising model.

Analogous techniques for point processes were first described in [19] for a special model, the general method for locally stable point processes can be found in [20], while [22] present an adaptive variation on the theme. The dynamics are those of a spatial birth-and-death process [29] with constant death rate and birth rate proportional to the conditional intensity. In our marked point process context, we suggest the following algorithm.

Algorithm 1. *Let $V_{t,(k,m)}$, $t \leq 0$, $(k,m) \in K \times M$ be a family of independent, uniformly distributed random variables on $(0,1)$. Initialise $T = 1$, and let $D(0)$ be a realisation of a marked point process with Poisson locations in K of rate Λ marked i.i.d. according to ν_M.*

1. *Extend $D(\cdot)$ backwards to time $-T$ by means of a spatial birth-and-death process with birth rate $\Lambda\,\nu \times \nu_M$ and unit death rate.*
2. *Generate $L^{-T}(\cdot)$ (lower process) and $U^{-T}(\cdot)$ (upper process) forwards in time as follows:*
   - *set $L^{-T}(-T) = \emptyset$ and $U^{-T}(-T) = D(-T)$;*
   - *if $D(\cdot)$ experiences a backward birth, i.e. $D(t-) = D(t) \cup \{(k,m)\}$ where $D(t-)$ denotes the state just prior to time t, delete (k,m) from $L^{-T}(t-)$ and $U^{-T}(t-)$;*
   - *if $D(\cdot)$ experiences a backward death, i.e. $D(t-) = D(t) \setminus \{(k,m)\}$, the marked point (k,m) is added to $L^{-T}(t-)$ if $V_{t,(k,m)} \leq \alpha_{\min}(U^{-T}(t-), L^{-T}(t-), (k,m))$ and to $U^{-T}(t-)$ if $V_{t,(k,m)} \leq \alpha_{\max}(U^{-T}(t-), L^{-T}(t-), (k,m))$.*
3. *If $U^{-T}(0) = L^{-T}(0)$ stop. Else set $T = 2T$ and repeat.*
4. *Return $U^{-T}(0)$.*

The birth acceptance probabilities are given by

$$\begin{aligned} \alpha_{\min}(U, L, (k,m)) &= \min\{\lambda((k,m);\mathbf{y})/\Lambda : L \subseteq \mathbf{y} \subseteq U\} \\ \alpha_{\max}(U, L, (k,m)) &= \max\{\lambda((k,m);\mathbf{y})/\Lambda : L \subseteq \mathbf{y} \subseteq U\} \end{aligned} \tag{3.1}$$

On top of the assumptions outlined at the end of Section 2 it is convenient, though not strictly necessary, to assume either attraction or repulsion, that is $\lambda((k,m);\mathbf{y}') \leq \lambda((k,m);\mathbf{y})$ respectively $\lambda((k,m);\mathbf{y}') \geq \lambda((k,m);\mathbf{y})$ for all $\mathbf{y}' \subseteq \mathbf{y}$ and all $(k,m) \in K \times M$.

Doing so, the birth acceptance probabilities (3.1) simplify considerably. Indeed, if f defines an attractive marked point process, $\alpha_{\min}(U, L, (k,m)) = \lambda((k,m); L)$ and $\alpha_{\max}(U, L, (k,m)) =$

$\lambda((k,m);U)$; in the repulsive case, $\alpha_{\min}(U,L,(k,m)) = \lambda((k,m);U)$ and $\alpha_{\max}(U,L,(k,m)) = \lambda((k,m);L)$.

We defer the proof of correctness of Algorithm 1 to the Appendix.

Example 1. Widom–Rowlinson mixture model (continued)
The conditional intensity for adding a point of type 1 *at* $k \in K$ *to the multi-type point pattern* $\mathbf{y}$ *is*

$$\lambda((k,1);\mathbf{y}) = \beta_1 \, \mathbf{1}\{\| u-k \| > R \text{ for all } (u,2) \in \mathbf{y}\};$$

a similar expression holds for addition of a point of type 2. *Hence, the Widom–Rowlinson model is hereditary and locally stable with* $\Lambda = \max\{\beta_1, \beta_2\}$. *Furthermore,* $\lambda((k,m);\mathbf{y}') \geq \lambda((k,m);\mathbf{y})$ *for all* $\mathbf{y}' \subseteq \mathbf{y}$ *and all* $(k,m) \in K \times M$.

Example 2. Multi-type pairwise interaction process (continued)
Model (2.2) is hereditary and locally stable with upper bound $\Lambda = \max_{m \in M} \beta_m$. *The Papangelou conditional intensity*

$$\lambda((k,m);\mathbf{y}) = \beta_m \prod_{(u,i)\in\mathbf{y}} \gamma_{im}(\| u-k \|)$$

is decreasing in $\mathbf{y}$ *with respect to the inclusion order, so Algorithm 1 is directly applicable.*

Example 3. Candy model (continued)
The Candy model density (2.3) is strictly positive, and locally stable with $\Lambda = \beta$. *The conditional intensity at* $\xi = (k,l,\theta)$ *is given by*

$$\lambda(\xi;\mathbf{y}) = \beta \, \exp\left[\frac{l - l_{\max}}{l_{\max}}\right] \gamma_r^{n_r(\xi;\mathbf{y})} \, \gamma_o^{n_o(\xi;\mathbf{y})}$$

and is decreasing with respect to the inclusion order. Here we use the notation $n_r(\xi;\mathbf{y})$ *for the number of* $\sim_r$ *neighbours of* ξ *in the marked point pattern* $\mathbf{y}$, *and similarly* $n_o(\xi;\mathbf{y})$ *denotes the number of* $\sim_o$ *neighbours of* ξ *in* $\mathbf{y}$ *for which the angle condition holds. Thus, the model is repulsive, and Algorithm 1 may be used to obtain samples.*

4. Clan of ancestors for marked point processes

The coupling from the past algorithm is particularly good when the distribution to sample from possesses some partial order structure. However, it generates a lot of marked points that have no effect on the final outcome. This observation motivated [7] to build an algorithm that avoids births of such points, or more generally individuals in some class of objects. Note however that even in the earliest days, Kendall had similar ideas in mind, especially the concept of space time cylinders (personal communication, 1995).

Write R for the range of interaction of a marked point process density f in the sense that

$$\lambda((k,m);\mathbf{y}) = \lambda((k,m);\mathbf{y} \cap B(k,R))$$

where $B(k,R) = \{(u,i) \in K \times M, \| u-k \| \leq R\}$. The equation is clearly satisfied for R equal to half the diameter of K, but the algorithm to be described below is efficient only for a low density of cylinders, that is for R that are small compared to Λ.

Algorithm 2. *Let $D(0)$ be a realisation of a marked point process with Poisson locations in K of rate Λ marked i.i.d. according to $\nu_M(\cdot)$. Initialise the clan of ancestors as $A = D(0)$.*

1. *Extend $D(\cdot)$ backwards by means of a spatial birth-and-death process with birth rate $\Lambda\,\nu\times\nu_M$ and unit death rate. At each (backward) death incident $D(t-) = D(t)\backslash\{(k,m)\}$ for some $t < 0$ and $(k,m) \in A \cap D(t)$, add the marked points in $B(k,R) \cap D(t-)$ to A. The backwards sweep ends when $A_t = A \cap D(t) = \emptyset$. The stopping time is denoted by $-T$.*

2. *Generate $Y(\cdot)$ forwards in time as follows:*

   - *set $Y(-T) = \emptyset$;*
   - *if $D(\cdot)$ experiences a backward birth, i.e. $D(t-) = D(t) \cup \{(k,m)\}$, delete (k,m) from $Y(t-)$;*
   - *if $D(\cdot)$ experiences a backward death, i.e. $D(t-) = D(t) \setminus \{(k,m)\}$ for some $(k,m) \in A_t$, the marked point (k,m) is added to $Y(t-)$ with probability $\lambda((k,m); Y(t-))/\,\Lambda$; if $(k,m) \notin A_t$ then $Y(t) = Y(t-)$ remains unchanged.*

Notice that Algorithm 2 does not require monotonicity properties nor a check on coalescence as in Algorithm 1. Rather, there are single backward and forward runs to respectively obtain the clan of ancestors and to thin the dominating process. On the other hand, the interaction structure becomes of prime importance. Here, we used a fixed range assumption, but alternatively the *incompatibility index*

$$I((k,m),(l,n)) := \mathbf{1}\{\sup_{\mathbf{y}} |\lambda((k,m);\mathbf{y}) - \lambda((k,m);\mathbf{y}\cup\{(l,n)\})| > 0\}$$

might be used. Also, one might generate the ancestors themselves directly, see [7]. The formulation above in terms of space-time cylinders is an adaptation to our marked point process context of the presentation in [12], see also [7, p. 80] and [3]. A proof of correctness may be found in the Appendix.

Example 1. Widom–Rowlinson mixture model (continued)
For the Widom–Rowlinson model, the clan of ancestors algorithm can be implemented using cylinders of radius R, cf. (2.1). In terms of the incompatibility index, note that $\lambda((k,m);\mathbf{y})$ depends only on those marked points in $\mathbf{y} \cap B(k,R)$ that have a label other than m.

Example 2. Multi-type pairwise interaction process (continued)
The range of interaction of (2.2) depends on the interaction function. If $\gamma_{ij}(t) = 1$ for $t > r_{ij}$, the range is $R = \max\{r_{ij} : i,j \in M\}$.

Example 3. Candy model (continued)
For the Candy model (2.3), consider a fixed segment ξ. Now, any $\sim_r$-neighbour of ξ must have its midpoint closer than $l_{\max}/2$ to the midpoint of ξ. Moreover, for a $\sim_o$-neighbour, the distance between its midpoint and that of ξ is at most the sum of the half lengths of each of the segments and the distance between their endpoints; the latter by definition of $\sim_o$ does not exceed $l_{\max}/4$, so that for this relation the interaction range is $5l_{\max}/4$.

5. Gibbs sampler for marked point processes

The Gibbs sampler is a simulation technique for multivariate data that repeatedly draws from the conditional distribution of one of the components given the others. Marked point processes, however, with the exception of the Widom–Rowlinson mixture model [16], do not in general split themselves naturally into components. Even for multi-type point processes, sampling from the marginal distribution of a single component is just as complicated as sampling from the joint distribution [21]. Lattice approximations of marked point processes on the other hand are quite common, especially in the context of pseudo-likelihood estimation [5, 14], and for such processes, Gibbs sampling is the natural way to proceed.

5.1 Approximation by a lattice process

In this section, we additionally suppose that f is Markov [21] with respect to some symmetric relation $\sim$ on $K \times M$. This is not really a restriction, since all marked points may be defined to be related. By the Hammersley–Clifford theorem [1, 32], Markov marked point processes are characterised by the fact that their density

$$f(\mathbf{y}) = \prod_{\text{cliques } \mathbf{z} \subseteq \mathbf{y}} \phi(\mathbf{z}) \tag{5.1}$$

factorises over sets of mutually related marked points (the *cliques*). By convention, the empty set is a clique, as is a set consisting of a single marked point. We set $\phi(\mathbf{z}) = 1$ if $\mathbf{z}$ is not a clique. Doing so, one derives

$$\lambda((k,m);\mathbf{y}) = \prod_{\mathbf{z} \subseteq \mathbf{y}} \phi(\mathbf{z} \cup \{(k,m)\}).$$

Example 1. Widom–Rowlinson mixture model (continued)
Define $(u,i) \sim (v,j)$ if and only if $i \neq j$ and $||u-v|| \leq R$. Then the Widom–Rowlinson density (2.1) can be factorised as in (5.1) with $\phi(\emptyset) = f(\emptyset)$, $\phi(\{(k,m)\}) = \beta_k$ and $\phi(\{(u,1),(v,2)\}) = \mathbf{1}\{||u-v|| > R\}$. For all other marked point patterns $\mathbf{z}$, we have $\phi(\mathbf{z}) = 1$.

Example 2. Multi-type pairwise interaction process (continued)
The density (2.2) is already factorised with respect to the relation $(u,i) \sim (v,j) \Leftrightarrow \gamma_{ij}(||u-v||) < 1$.

Example 3. Candy model (continued)
Define a relation $\sim$ on $K \times M$ by $y \sim y'$ if and only if $y \sim_r y'$ or $y \sim_o y'$. Note that a pair of marked points cannot be related by both $\sim_r$ and $\sim_o$ simultaneously. Now, the Candy density (2.3) satisfies (5.1) with $\phi(\emptyset) = f(\emptyset)$, $\phi(\{(k,l,\theta)\}) = \beta \exp[(l - l_{\max})/l_{\max}]$ and $\phi(\{y,y'\})$ equal to γ_r if $y \sim_r y'$, to γ_o if $y \sim_o y'$ and the pair $\{y,y'\}$ satisfies the small orientation distance constraint. For other sets, ϕ reduces to 1.

Let $K = \bigcup_{i=1}^{n_K} K_i$ be a finite partition of the location space such that $0 < \nu(K_i) < \infty$ for all i. The mark space is partitioned similarly as $M = \bigcup_{j=1}^{n_M} M_j$ with $0 < \nu_M(M_j) \leq 1$ for all j. Hence, $K \times M = \bigcup_{i=1}^{n_K} \bigcup_{j=1}^{n_M} (K_i \times M_j)$. A representative ξ_{ij} is chosen for each cell $K_i \times M_j$, for example the centroid. The set of all representatives is denoted by Ξ.

Following [5, 14], we approximate (5.1) by an auto-logistic distribution. Thus, let N be a random field on the lattice induced by Ξ taking binary values as follows. The state space

is finite and consists of vectors (n_{ij}), $i = 1, \ldots, n_K$, $j = 1 \ldots n_M$ such that $n_{ij} \in \{0, 1\}$ and $\sum_{j=1}^{n_M} n_{ij} \leq 1$ for all $i = 1, \ldots, n_K$. If we interpret n_{ij} as the indicator of a marked point in $K_i \times M_j$, the condition on $\sum_j n_{ij}$ ensures that each strip $K_i \times M$ contains at most a single marked point. Thus, the random vector $N_i = (N_{i1}, \ldots, N_{in_M})$ almost surely takes values in $S := \{\mathbf{0}, \mathbf{e}_1, \cdots, \mathbf{e}_{n_M}\}$, the set consisting of the unit and zero vectors. The probability mass function of N is defined as

$$\mathbb{P}(\mathbf{n}) \propto \prod_{\{i,j\}} [\nu(K_i)\nu_M(M_j)]^{n_{ij}} \prod_{\emptyset \neq \mathbf{z} \subseteq \Xi} \phi(\mathbf{z})^{\eta(\mathbf{n};\mathbf{z})} \tag{5.2}$$

for $\mathbf{n}$ in the state space. In words, (5.2) is the lattice version of (5.1). Here, $\eta(\mathbf{n}; \mathbf{z}) = \prod_{\xi_{ij} \in \mathbf{z}} n_{ij}$, which takes values 0 or 1. By convention, $0^0 = 1$. The discrete counterpart of the conditional intensity is straightforwardly derived. Denote by $\mathbf{n}_{-i}$ the vector consisting of the $n_{i'j}$ for $i' \neq i$. Then the local characteristic at strip i is given by

$$\mathbb{P}(N_i = \mathbf{e}_j | \mathbf{n}_{-i}) = \frac{\nu(K_i)\nu_M(M_j) \prod_{\mathbf{z}} \phi(\mathbf{z} \cup \{\xi_{ij}\})^{\eta(\mathbf{n}_{-i};\mathbf{z})}}{1 + \nu(K_i) \sum_{j'=1}^{n_M} \nu_M(M_{j'}) \prod_{\mathbf{z}} \phi(\mathbf{z} \cup \{\xi_{ij'}\})^{\eta(\mathbf{n}_{-i};\mathbf{z})}} \tag{5.3}$$

for $j \neq 0$, provided the conditioning event has strictly positive mass under $\mathbb{P}$. The products range over $\mathbf{z} \subseteq \Xi \setminus \{\xi_{ij'} : j' = 1, \ldots, n_M\}$, and $\eta(\mathbf{n}_{-i}; \mathbf{z})$ is defined as before.

Example 1. Widom–Rowlinson mixture model (continued)
For the two-type mixture model defined by (2.1), the natural mark discretisation is in two bins corresponding to the two labels. Then $S = \{\mathbf{0}, \mathbf{e}_1, \mathbf{e}_2\}$ *and, for* $\mathbf{n}_{-i}$ *such that* $\mathbb{P}(\mathbf{n}_{-i}) > 0$,

$$\frac{\mathbb{P}(N_i = \mathbf{e}_1 | \mathbf{n}_{-i})}{\mathbb{P}(N_i = \mathbf{0} | \mathbf{n}_{-i})} = \frac{1}{2} \nu(K_i)\, \beta_1\, \mathbf{1}\{c_2(\mathbf{n}_{-i}) = 0\}$$

if we write $\xi_{i1} = (k_i, 1)$ *and denote by* $c_2(\mathbf{n}_{-i})$ *the cardinality of the set* $\{\xi_{i'2} = (k_{i'}, 2) : ||k_i - k_{i'}|| \leq R; \mathbf{n}_{i'} = \mathbf{e}_2\}$. *An expression for the conditional probability of* $\mathbf{e}_2$ *relative to that of* $\mathbf{0}$ *at strip* i *is obtained by interchanging* 1 *and* 2 *in the formula above.*

Example 2. Multi-type pairwise interaction process (continued)
For pairwise interaction densities of the form (2.2), the number of mark bins n_M *in the lattice approximation is bounded from above by the cardinality* I *of the mark space* M. *Write* $\xi_{ij} = (k_i, m_j)$ *for* $i = 1, \ldots, n_K$, $j = 1, \ldots, n_M$. *Then,*

$$\frac{\mathbb{P}(N_i = \mathbf{e}_j | \mathbf{n}_{-i})}{\mathbb{P}(N_i = \mathbf{0} | \mathbf{n}_{-i})} = \nu(K_i)\, \nu_M(M_j)\, \beta_{m_j} \prod_{\xi_{i'j'} = (k_{i'}, m_{j'}), i' \neq i} \gamma_{m_j m_{j'}}(||k_i - k_{i'}||)^{n_{i'j'}}$$

for all $\mathbf{e}_j \in S \setminus \{\mathbf{0}\}$ *and* $\mathbf{n}_{-i}$ *not in the set of states with zero probability.*

Example 3. Candy model (continued)
The Candy model exhibits pairwise interactions only, hence the local characteristics (5.3) can be written as

$$\frac{\mathbb{P}(N_i = \mathbf{e}_j | \mathbf{n}_{-i})}{\mathbb{P}(N_i = \mathbf{0} | \mathbf{n}_{-i})} = \nu(K_i)\, \nu_M(M_j)\, \beta\, \exp\left[\frac{l_j - l_{\max}}{l_{\max}}\right] \prod_{\xi_{i'j'}, i' \neq i} \phi(\xi_{ij}, \xi_{i'j'})^{n_{i'j'}}$$

where the pair interaction function ϕ *is as computed in the beginning of this section. As in Example 2, we take representatives* ξ_{ij} *of the form* (k_i, m_j), *where for the Candy model* $m_j = (l_j, \theta_j)$.

5.2 Gibbs sampler for lattice process

The exact Gibbs sampler of [18] follows the CFTP idea proposed by Propp and Wilson in [30]. However, if one allows for arbitrary probability mass functions, in order to avoid having to run many coupled chains simultaneously, for each cell $K_i \times M_j$ or strip $K_i \times M$ the set of possible values is kept. The algorithm below implements strip updating, so that the set $\mathcal{X}_i(t)$ is a subset of S; cell updating may be implemented with obvious modifications.

Algorithm 3. *Suppose that $\psi : S^{\otimes n_K} \times (0,1) \times \{1,\dots,n_K\} \to S$ is a transition kernel for the Gibbs sampler, that is, $\psi(\chi,u,v)$ alters the state χ_v at strip v based on a realisation u of a random number in such a way that the new state is distributed according to the $\mathbb{P}$-conditional distribution given χ_{-v}. Let V_t, $t = -1,-2,\dots$, be a family of independent, uniformly distributed random variables on $\{1,\dots,n_K\}$ and let U_t be a family of independent, uniformly distributed random variables on $(0,1)$. Initialise $T = 1$.*

1. *Let $\mathcal{X}_i^{-T}(-T) = S$ for all $i = 1,\dots,n_K$.*
2. *Extend forwards from time $-T$ by setting*

$$\mathcal{X}_{V_t}^{-T}(t+1) = \{s \in S : \exists \chi \in \mathcal{X}^{-T}(t) : \psi(\chi, U_t, V_t) = s\}$$

   and leaving the sets kept at other strips unchanged.
3. *If for all strips $K_i \times M$ the cardinality of $\mathcal{X}_i^{-T}(0)$ is one, stop. Else, set $T = 2T$ and repeat.*

It is often convenient from a computational point of view to compute the new state $\mathcal{X}_{V_t}^{-T}(t+1)$ explicitly only if one knows in advance that it is a singleton, and to set $\mathcal{X}_{V_t}^{-T}(t+1) = S$ otherwise. An example of such an implementation is the multigamma coupler to be described below. Further examples can be found in e.g. [18, 27].

Although generally applicable (see the Appendix) to all (hereditary) locally stable Markov marked point process models, Algorithm 3 is particularly appealing in case of monotonicity [30] or anti-monotonicity [17]. Recall that the random field N with probability mass function $\mathbb{P}$ is monotone with respect to a given order $\leq$ on S if for all $s \in S$, $\mathbb{P}(N_i \leq s|\mathbf{n}_{-i}) \geq \mathbb{P}(N_i \leq s|\tilde{\mathbf{n}}_{-i})$ whenever $\mathbf{n}_{-i} \leq \tilde{\mathbf{n}}_{-i}$ stripwise, $\mathbb{P}(\mathbf{n}_{-i}) > 0$ and $\mathbb{P}(\tilde{\mathbf{n}}_{-i}) > 0$. Suppose that the transition kernel ψ were chosen in such a way that it respects the order. Then, Algorithm 3 effectively reduces to the original Propp–Wilson algorithm [30] based on two coupled chains. More specifically, an upper chain U^{-T} is initialised at time $-T$ with the maximum, a lower chain L^{-T} with the minimum. Both chains are run forwards up to time 0 using the same kernel ψ and random numbers provided by U_t and V_t; if $U^{-T}(0) = L^{-T}(0)$, the algorithm terminates and outputs the common state, otherwise the running time is doubled and the procedure repeated.

A convenient example of a transition kernel that respects a given order is

$$\psi(\mathbf{n}, u, i) = \max\{j = 0,\dots,n_M : \mathbb{P}(N_i \geq \mathbf{e}_j|\mathbf{n}_{-i}) \geq u\}$$

and we shall use this kernel in the simulations in Section 6 below. Note that the Markov properties of Y induce Markovianity of its random field approximation N, so computations may be carried out locally.

For the special case of the Widom–Rowlinson mixture model, monotonicity for the strip sampler with respect to the total order $\mathbf{e}_1 < \mathbf{0} < \mathbf{e}_2$ on S can be established [18]. The same is true for bivariate pairwise interaction processes for which $\gamma_{ii} \equiv 1$ and $\gamma_{ij}(t) > 0$ for all $t > 0$, see the Appendix for a derivation. In general, however, such a property does not seem to hold, and we resort to multigamma implementations. Indeed, since

$$\mathbb{P}(N_i = \mathbf{0} | \mathbf{n}_{-i}) \geq \frac{1}{1 + \Lambda \max_i \nu(K_i)} > 0$$

uniformly in $i \in \{1, \ldots, n_K\}$ and $\mathbf{n}_{-i}$, we may choose the transition kernel governing Algorithm 3 in such a way that $\psi(\chi, u, v) = \mathbf{0}$ whenever $u \leq 1/(1 + \Lambda \max_i \nu(K_i))$. Thus, for small realised values of U_t, regardless of $\mathcal{X}^{-T}(t)$, the strip V_t is assigned value $\mathbf{0}$. For $U_t > 1/(1 + \Lambda \max_i \nu(K_i))$, if $\mathcal{X}_v^{-T}(t)$ is not a singleton for some neighbour v of the current strip V_t, no update is made, i.e. $\mathcal{X}_{V_t}^{-T}(t+1) = S$.

6. Simulation study

In this section, we present a simulation study to assess the range of applicability and relative efficiency of the various exact simulation algorithms. The models we consider are the Widom–Rowlinson model (2.1) the Candy model (2.3), and two multi-type point processes. For the latter, we restrict ourselves to $I = 2$, and between type interaction only (i.e. $\gamma_{ii} \equiv 1$). For the interaction function, we choose

$$\gamma_{12}(t) = \begin{cases} 1 - (1-\gamma)\mathbf{1}\{t \leq r\} & \textbf{Strauss} \\ 1 - 1/(1 + (t/\sigma)^2)^2 \mathbf{1}\{t \leq 3\sigma\} & \textbf{Cauchy} \end{cases}$$

All models are sampled on the unit square.

6.1 Coupling from the past

In Figure 1 we plot the total number of jumps against intensity and radius parameters. It should be noted that the actual number of calculations is larger, as the method is based on successive doubling.

For the Widom–Rowlinson model, the intensity parameter is $\beta_1 = \beta_2$ which ranged between 1 and 150 with steps of 1. The hard core distance between points of different type was taken to be in between 0.005 and 0.15 with steps of 0.005.

For the bivariate Strauss and Cauchy models, the intensity parameter $\beta_1 = \beta_2$ also ranged between 1 and 150 with steps of 1. The range parameter of the Strauss interaction function is r, which we allowed to vary between 0.005 and 0.15 with steps of 0.005. We set the strength of interaction equal to $\gamma = 0.5$. For the Cauchy interaction function, the range $r = 3\sigma$. Again, we allowed r to vary between 0.005 and 0.15 with steps of 0.005.

The parameters of the Candy model were chosen as follows. The orientation and rejection parameters were set to $\gamma_o = \gamma_r = 0.5$, with $\tau = 0.1$ and $\delta = 0.05$. The range of interaction is determined by the length of the segments as $r = 1.25 l_{\max}$. Thus, we assumed the length distribution was concentrated on $l_{\max}$, which we let vary between 0.005 and 0.12 with steps of 0.005.

For each combination of parameters, 25 independent samples were generated by means of Algorithm 1, and the average number of jumps recorded. The results are plotted in Figure 1. It can be seen that the stronger the interaction, the longer it takes to obtain a sample. Indeed,

for the same intensity and range parameters, the interaction in the Widom–Rowlinson model is of hard core type, whereas the Strauss interaction function is a positive constant. The Cauchy interaction is more severely repulsive than that of the Strauss model at very short interpoint distances, but less so for most of the range. Finally in the Candy model the maximum strength of interaction is the same as that of the Strauss bivariate interaction process, but only a subset of segments within the interaction range actually contributes to the conditional intensity.

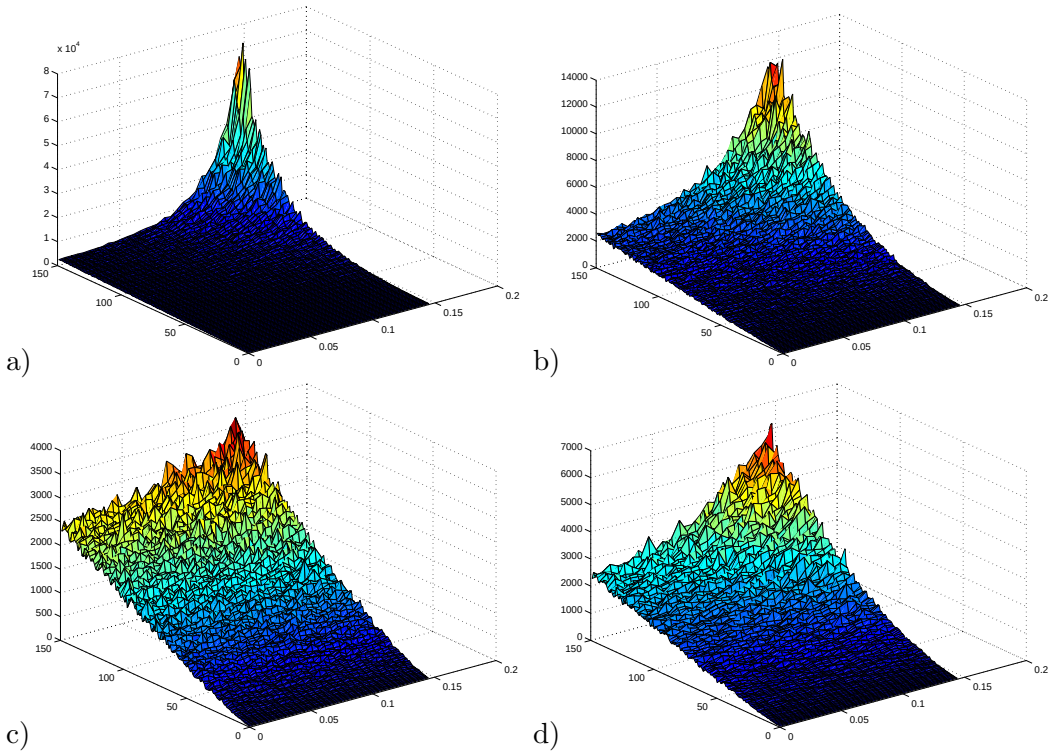

Figure 1: CFTP algorithm a) Widom-Rowlinson b) Strauss c) Cauchy d) Candy.

6.2 Clan of ancestors

For Algorithm 2, the convergence time may be measured by the number of jumps before finishing the backward sweep. In contrast to coupling from the past, this number reflects the computational load well, as no doubling scheme is needed.

For the Widom–Rowlinson model, the intensity parameter is $\beta_1 = \beta_2$ which ranged between 1 and 100 with steps of 1. The hard core distance between points of different type was taken to be in between 0.005 and 0.08 with steps of 0.005.

For the bivariate Strauss and Cauchy models, the intensity parameter $\beta_1 = \beta_2$ also ranged between 1 and 100 with steps of 1. The range parameter of the Strauss interaction function is r, which we allowed to vary between 0.005 and 0.08 with steps of 0.005. We set the strength of interaction equal to $\gamma = 0.5$. For the Cauchy interaction function, the range $r = 3\sigma$. Again, we allowed r to vary between 0.005 and 0.08 with steps of 0.005.

The parameters of the Candy model were chosen as follows. The orientation and rejection parameters were set to $\gamma_o = \gamma_r = 0.5$, with $\tau = 0.1$ and $\delta = 0.05$. The range of interaction is determined by the length of the segments as $r = 1.25 l_{\max}$. Thus, we assumed the length distribution was concentrated on $l_{\max}$, which we let vary between 0.005 and 0.065 with steps of 0.005.

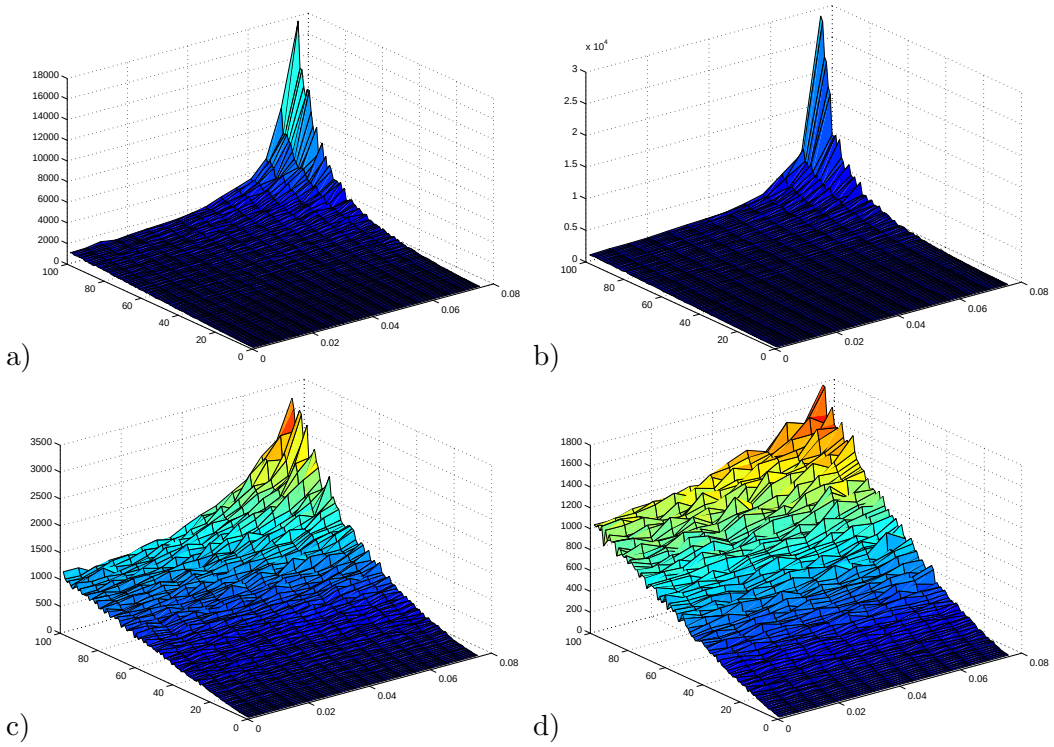

Figure 2: Clan of ancestors algorithm a) multi-type fixed range b) Candy fixed range c) multi-type incompatibility index d) Candy incompatibility index.

For each combination of parameters, 25 independent samples were generated by means of Algorithm 2, and the average number of jumps recorded. The results are plotted in Figures 2(ab). The criterion for the backward sweep to end is determined only by Λ and r. Thus, the interaction structure is irrelevant, and all models take equally long to sample from. Compared to coupling from the past, within the considered range of parameters, the clan of ancestors algorithm is generally faster. On the other hand, coupling from the past is applicable over a wider range of parameter values.

If we use the incompatibility index, the neighbourhood structure becomes important. For the Widom–Rowlinson model and both multi-type pairwise interaction processes, the conditional intensity of a point (k, m) depends only on points of type other than m that are within distance R of k. For the Candy model, we may restrict to points (l, n) that satisfy either $(l, n) \sim_r (k, m)$ or $(l, n) \sim_o (k, m)$.

If we take 25 independent samples and average the number of jumps for each pair of parameters, we obtain the plots in Figures 2(cd). The gain in efficiency is obvious, and

stronger the more is saved in terms of neighbour counts. Note that the strength of interaction does not matter; indeed identical plots are obtained for all except the Candy model.

6.3 Gibbs sampler

In Figure 3, we plot the total number of jumps against intensity and range parameters. As for coupling from the past, the actual computational load is larger due to the successive doubling involved.

For the Widom–Rowlinson and the bivariate cross-interaction models, we used the monotone strip based Gibbs sampler. For the Candy model, we have chosen the stripwise multigamma coupling.

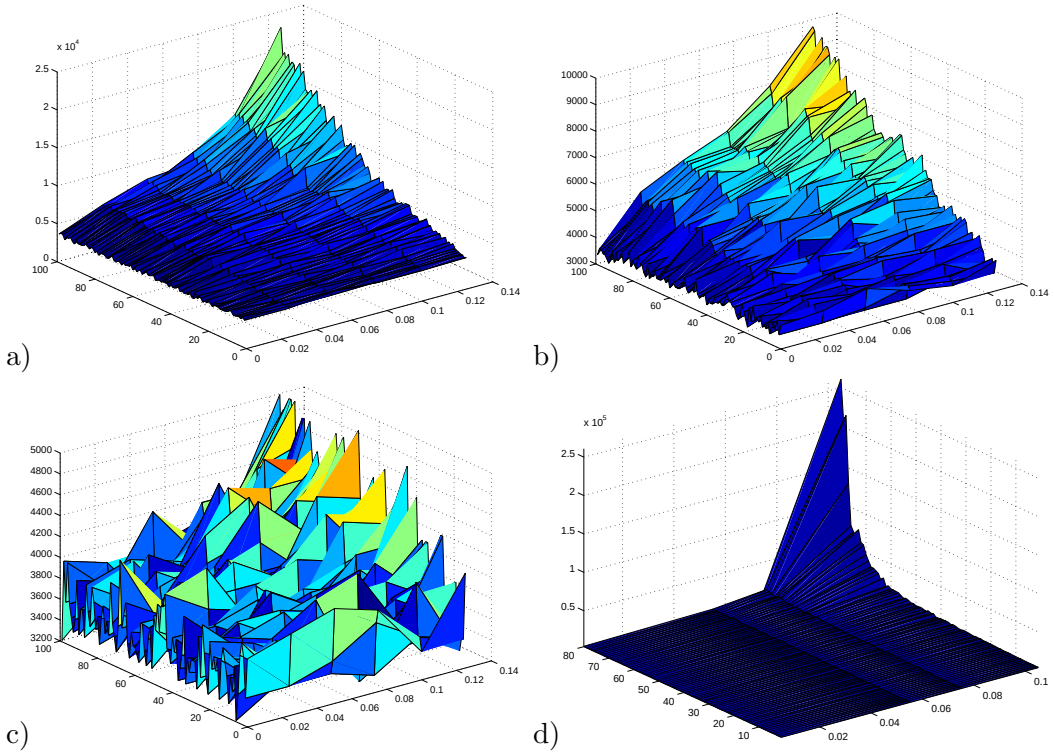

Figure 3: Gibbs algorithm a) Widom-Rowlinson b) Strauss c) Cauchy d) Candy.

For all models, 20×20 lattices were used. As before, for the Widom–Rowlinson model, the intensity parameter $\beta_1 = \beta_2$ ranged between 1 and 100 with steps of 1. Due to the discretisation effect, only a few hard core distance values R have to be considered. We took $r = R = 0.005$, 0.055, 0.075, 0.105, 0.115 and 0.155. The mark space was divided in two bins.

For the bivariate Strauss and Cauchy models, the intensity parameter $\beta_1 = \beta_2$ also ranged between 1 and 100 with steps of 1. The range parameter of the Strauss interaction function is r, while for the Cauchy interaction function, the range is $r = 3\sigma$. The values of r were chosen in an identical fashion as for the Widom–Rowlinson model (note though that for the Cauchy model γ_{12} is not a step function, resulting in a somewhat different discretisation effect!). We set the strength of interaction in the Strauss model equal to $\gamma = 0.5$. The natural mark

discretisation is to allocate each type its own bin.

The orientation and rejection parameters of the Candy model were set to $\gamma_o = \gamma_r = 0.5$, with $\tau = 0.1$ and $\delta = 0.05$. The lengths $l = 0.8 * r$ were given by the discretised values $r = 0.005$, 0.055, 0.075 and 0.105. The intensity parameter β was ranging from 1 to 80 with step size 1. The orientation interval was partitioned in 5 sub-intervals with equal lengths.

For each combination of parameters, 25 independent samples were generated by means of Algorithm 3, and the average number of jumps recorded. The results are plotted in Figure 3. It can be seen that monotonicity not only results in easier algorithms, but also improves the speed of convergence. Furthermore, if a model exhibits stronger interaction, it takes longer to obtain a sample. Indeed, the same efficiency ranking as in Figure 1 may be observed for the Widom–Rowlinson, Strauss and Cauchy models.

7. Conclusion

In this work, we extended some recently proposed exact simulation methods to the case of marked point processes. Three families of algorithms were proposed: coupling from the past, the clan of ancestors technique and the Gibbs sampler. The various algorithms have been tested on several models, including the Widom-Rowlinson mixture model, multi-type pairwise interaction processes and the Candy line segment model. A simulation study was carried out in order to analyse the proposed methods in terms of speed of convergence in relation to the parameters of the model.

For the range of models investigated, the clan of ancestors algorithm using the incompatibility index was the fastest method among the ones analysed in this work, while coupling from the past was applicable to the widest range of parameter values.

If one is prepared to approximate by discretisation, a proper choice of Gibbs sampler makes it possible to obtain samples from models that lack monotonicity or have such a high local stability bound as to rule out coupling from the past or clan of ancestor approaches from a practical point of view.

For the future we intend to investigate extensions of the Metropolis–Hastings dynamics [20] that allow for the construction of change moves in order to improve the mixing properties of the simulated Markov chain.

Acknowledgments

The research was carried out while the second author was at CWI. It is a pleasure to express our thanks to A.G. Steenbeek for expert programming assistance and to K.K. Berthelsen for helpful comments.

Appendix

Proof: (Correctness Algorithm 1) The correctness of Algorithm 1 can be shown by arguments similar to those used by [20] in the unmarked case. Indeed

- the dominating process $D(\cdot)$ is in equilibrium;
- $L^{-T}(t) \subseteq L^{-S}(t) \subseteq U^{-S}(t) \subseteq U^{-T}(t) \subseteq D(t)$ for all $-S \leq -T \leq t \leq 0$;
- due to the coupling by using the same $V_{t,(k,m)}$ in both the U and the L process, once $L^{-T}(t) = U^{-T}(t)$ for some $t \in [-T, 0]$, the two processes proceed as one;
- the D-process will almost surely reach state $\emptyset$, hence termination is guaranteed.

Next, set $Y^{-T}(-T) = \emptyset$ and define a process $Y^{-T}(\cdot)$ on $[-T, 0]$ that evolves just like the upper and lower processes, except that if $Y^{-T}(t-) = \mathbf{y}$ the birth at time t of a marked point (k, m) is accepted if $V_{t,(k,m)} \leq \lambda((k, m); \mathbf{y})/\Lambda$. Therefore, $Y^{-T}(\cdot)$ exhibits the dynamics of a spatial birth-and-death process with equilibrium distribution defined by $f(\cdot)$. Clearly, $L^{-T}(t) \subseteq Y^{-T}(t) \subseteq U^{-T}(t)$ for all $t \geq -T$ and in particular for $t = 0$, so that, as the algorithm terminates almost surely, with probability one $\lim_{T\to\infty} Y^{-T}(0)$ is well-defined. As $D(\cdot)$ is in equilibrium and time-reversible, the distribution of $Y^{-T}(0)$ is the same as it would be if run forward from time 0 (coupled to the dominating process as before) over a time period of length T. In conclusion, Algorithm 1 returns a sample from $f(\cdot)$. □

Proof: (Correctness Algorithm 2) To see that Algorithm 2 terminates, note that the D-process is time reversible and in equilibrium. The assumptions imply that $D(\cdot)$ almost surely reaches $\emptyset$, at which time T also $A_T = \emptyset$. If in the forward sweep all births were subjected to the conditional intensity test, the resulting process would also be in detailed balance with equilibrium distribution $f(\cdot)$.

Looking in more detail at the forwards birth transition, one notices that the decision whether a point about to be added to $D(t)$ is also kept in $Y(t)$ is made on the basis of the conditional intensity, which by the fixed range Markov assumption depends only on the R-neighbours of the point under consideration. Hence, points not in A_t may safely be disregarded.

As an aside, the clan of ancestors algorithm can also be used to sample infinite volume measures within a bounded window, see thm 3.18 in [7]. □

Proof: (Correctness Algorithm 3) By [18, thm 3.1], if the conditional probability that $N_i = \mathbf{0}$ given the values at other strips is uniformly bounded from below by some $\delta > 0$, the algorithm terminates almost surely and results in an unbiased sample from $\mathbb{P}$. Now, since we assume that (5.1) is locally stable,

$$\begin{aligned}\mathbb{P}(N_i = \mathbf{0}|\mathbf{n}_{-i}) &= \frac{1}{1 + \nu(K_i)\sum_{j=1}^{n_M} \nu_M(M_j) \prod_{\mathbf{z}} \phi(\mathbf{z} \cup \{\xi_{ij}\})^{\eta(\mathbf{n}_{-i};\mathbf{z})}} \\ &\geq \frac{1}{1 + \nu(K_i)\sum_{j=1}^{n_M} \nu_M(M_j)\Lambda} \geq \frac{1}{1 + K^*\Lambda} > 0\end{aligned}$$

where $K^* = \max\{\nu(K_i) : i = 1, \ldots, n_K\}$, and the proof is complete. □

Proof: (Monotonicity bivariate cross-interaction processes) Consider a bivariate pairwise interaction process with density (2.2). Suppose $\gamma_{ii} \equiv 1$ for $i \in \{1, 2\}$ and $\gamma_{12}(t) > 0$ for all $t > 0$. Upon discretising K, in order to prove monotonicity with respect to the order $\mathbf{e}_1 < \mathbf{0} < \mathbf{e}_2$ on S, we have to establish

$$\begin{aligned} \mathbb{P}(N_i = \mathbf{e}_1 | \mathbf{n}_{-i}) &\geq \mathbb{P}(N_i = \mathbf{e}_1 | \tilde{\mathbf{n}}_{-i}) \\ \mathbb{P}(N_i = \mathbf{e}_2 | \mathbf{n}_{-i}) &\leq \mathbb{P}(N_i = \mathbf{e}_2 | \tilde{\mathbf{n}}_{-i}) \end{aligned}$$

for all i and all pairs of configurations for which $\mathbf{n}_{-i} \leq \tilde{\mathbf{n}}_{-i}$ stripwise. With the notation $\xi_{i1} = (k_i, 1)$, $\xi_{i2} = (k_i, 2)$ for the marked representatives of cell K_i, and $\alpha_i = \nu(K_i)/2$, note that

$$\begin{aligned} \mathbb{P}(N_i = \mathbf{e}_1 | \mathbf{n}_{-i}) &= \frac{\alpha_i \beta_1 \prod_{k \in n_{-i}(2)} \gamma_{12}(||k_i - k||)}{1 + \alpha_i \beta_1 \prod_{k \in n_{-i}(2)} \gamma_{12}(||k_i - k||) + \alpha_i \beta_2 \prod_{k \in n_{-i}(1)} \gamma_{12}(||k_i - k||)} \\ &= \frac{1}{\left[\frac{1}{\alpha_i \beta_1} + \frac{\beta_2}{\beta_1} \prod_{k \in n_{-i}(1)} \gamma_{12}(||k_i - k||)\right] \prod_{k \in n_{-i}(2)} \gamma_{12}(||k_i - k||)^{-1} + 1} \end{aligned}$$

where we use the notation $n_{-i}(m)$ for the points in the neighbourhood of k_i that have mark $m \in \{1, 2\}$. Since by assumption γ_{12} takes values in $(0, 1]$, and every k in $\tilde{n}_{-i}(1)$ belongs also to $n_{-i}(1)$, while a k in $n_{-i}(2)$ is contained in $\tilde{n}_{-i}(2)$, the required ordering between the local characteristics follows. A similar argument is valid for the conditional probabilities of label 2, and we are done. □

References

1. A.J. Baddeley and J. Møller. Nearest-neighbour Markov point processes and random sets. *International Statistical Review*, 57:89–121, 1989.
2. O.E. Barndorff–Nielsen, D.R. Cox and C. Klüppelberg. *Complex stochastic systems*. Boca Raton: CRC Press/Chapman and Hall, 2000.
3. K.K. Berthelsen and J. Møller. A primer on perfect simulation for spatial point processes. *Research Report R-01-2026*, Department of Mathematical Sciences, Aalborg University, 2001.
4. K.K. Berthelsen and J. Møller. Spatial jump processes and perfect simulation. *Research Report R-01-2008*, Department of Mathematical Sciences, Aalborg University, 2001.
5. J.E. Besag, R. Milne and S. Zachary. Point process limits of lattice process. *Journal of Applied Probability*, 19:210–216, 1982.
6. D.J. Daley and D. Vere-Jones. *An introduction to the theory of point processes*. New York: Springer Verlag, 1988.
7. P.A. Ferrari, R. Fernández and N.L. Garcia. Perfect simulation for interacting point processes, loss networks and Ising models. *Stochastic Processes and their Applications*, 102:63–88, 2002.
8. J.A. Fill. An interruptible algorithm for perfect sampling via Markov chains. *Annals of Applied Probability*, 8:131–162, 1998.
9. J.A. Fill, M. Machida, D.J. Murdoch and J.S. Rosenthal. Extension of Fill's perfect rejection sampling algorithm to general chains. *Random Structures and Algorithms*, 17:290–316, 2000.
10. J.A. Fill and M. Huber. The randomness recycler: A new technique for perfect sampling. *Proceedings of the 41st Annual Symposium on Foundations of Computer Science*, 503–511, 2000.
11. J.A. Fill and M. Huber. The randomness recycler approach to perfect sampling. *Invited*

Papers Meeting on Perfect Simulation, 53rd Session of the International Statistical Institute, Seoul, Korea, August 2001.

12. N.L. Garcia. Perfect simulation of spatial processes. *Resenhas IME-USP*, 4:281–324, 2000.
13. C.J. Geyer and J. Møller. Simulation procedures and likelihood inference for spatial point processes. *Scandinavian Journal of Statistics*, 21:359–373, 1994.
14. M. Goulard, P. Grabarnik and A. Särkkä. Parameter estimation for marked Gibbs point processes through the maximum pseudo-likelihood method. *Scandinavian Journal of Statistics*, 23:365–379, 1996.
15. P.J. Green. Reversible jump MCMC computation and Bayesian model determination. *Biometrika*, 82:711-732, 1995.
16. O. Häggström, M.N.M. van Lieshout and J. Møller. Characterization results and Markov chain Monte Carlo algorithms including exact simulation for some point processes. *Bernoulli*, 5:641–659, 1999.
17. O. Häggström and K. Nelander. Exact sampling for anti-monotone systems. *Statistica Neerlandica*, 52:360–380, 1998.
18. O. Häggström and K. Nelander. On exact simulation of Markov random fields using coupling from the past. *Scandinavian Journal of Statistics*, 26:395–411, 1999.
19. W.S. Kendall. Perfect simulation for the area-interaction point process. In *Proceedings of the Symposium on Probability towards the year 2000*, L. Accardi and C. Heyde (Eds.) Berlin: Springer-Verlag, 1998.
20. W.S. Kendall and J. Møller. Perfect simulation using dominating processes on ordered spaces, with application to locally stable point processes. *Advances in Applied Probability (SGSA)*, 32:844–865, 2000.
21. M.N.M. van Lieshout. *Markov point processes and their applications*. London/Singapore: Imperial College Press/World Scientific Publishing, 2000.
22. M.N.M. van Lieshout and A.J. Baddeley. Extrapolating and interpolating spatial pattern. In *Spatial cluster modelling*, A.B. Lawson and D.G.T. Denison (Eds.) Boca Raton: Chapman and Hall/CRC Press, 2002.
23. M.N.M. van Lieshout and R.S. Stoica. The Candy model revisited: properties and inference. *Statistica Neerlandica*, 57:1–30, 2003.
24. A. Mira, J. Møller and G.O. Roberts. Perfect slice sampler. *Journal of the Royal Statistical Society, Series B*, 63:593–606, 2001.
25. J. Møller. On the rate of convergence of spatial birth-and-death processes. *Annals of the Institute of Statistical Mathematics*, 41:565–581, 1989.
26. J. Møller. Perfect simulation of conditionally specified models. *Journal of the Royal Statistical Society, Series B*, 61:251–264, 1999.
27. D.J. Murdoch and P.J. Green. Exact sampling from a continuous state space. *Scandinavian Journal of Statistics*, 25:483–502, 1998.
28. Y. Ogata and M. Tanemura. Estimation for interaction potentials of spatial point patterns through the maximum likelihood procedure. *Annals of the Institute of Statistical*

Mathematics, 33:315–338, 1981.

29. C.J. Preston. Spatial birth-and-death processes. *Bulletin of the International Statistical Institute*, 46:371–391, 1977.
30. J.G. Propp and D.B. Wilson. Exact sampling with coupled Markov chains and applications to statistical mechanics. *Random Structures and Algorithms*, 9:223–252, 1996.
31. B.D. Ripley. Modelling spatial patterns. *Journal of the Royal Statistical Society, Series B*, 39:172–192, 1977.
32. B.D. Ripley and F.P. Kelly. Markov point processes. *Journal of the London Mathematical Society*, 15:188–192, 1977.
33. R. Stoica, X. Descombes and J. Zerubia. A Gibbs point process for road extraction in remotely sensed images. Research Report 3923, INRIA Sophia Antipolis, 2000.
34. E. Thönnes. Perfect simulation of some point processes for the impatient user. *Advances in Applied Probability*, 31:69–87, 1999.
35. B. Widom and J.S. Rowlinson. A new model for the study of liquid-vapor phase transition. *Journal of Chemical Physics*, 52:1670–1684, 1970.