ELSEVIER

Perfect simulation for marked point processes

M.N.M. van Lieshout[a,*], R.S. Stoica[b,1]

[a]*CWI, P.O. Box 94079, 1090 GB Amsterdam, The Netherlands*
[b]*INRA - Biometrie, Domaine St. Paul, site Agroparc, 84914 Avignon Cedex 9, France*

Abstract

Some recently proposed exact simulation methods are extended to the case of marked point processes. Four families of algorithms are considered: coupling from the past, the clan of ancestors technique, the Gibbs sampler, and a Metropolis–Hastings algorithm based on birth and death proposals. From a theoretical point of view, conditions are given under which the algorithms yield unbiased samples in finite time. For practical application, a C++ library for marked processes is described. The various algorithms are tested on several models, including the Widom–Rowlinson mixture model, multi-type pairwise interaction processes, and the Candy line segment model. A simulation study is 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 is the fastest method among the ones analysed, while coupling from the past is applicable to the widest range of parameter values, and usually faster than the Metropolis–Hastings sampler. The latter two methods tend to be cumbersome if the underlying model is neither attractive nor repulsive. 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 in practice.
© 2006 Elsevier B.V. All rights reserved.

Keywords: Exact simulation; Markov chain Monte Carlo simulation; Markov marked point process

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 (1996). 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 Barndorff-Nielsen et al. (2001). The difficulty with such methods is the need for careful burn-in and for convergence diagnostics to assess whether the underlying Markov chain has

* Corresponding author. Centrum voor Wiskunde en Informatica, Kruislaan 413, 1098 SJ Amsterdam, The Netherlands. Tel.: +31 20 592 4008; fax: +31 20 592 4199.

E-mail addresses: colette@cwi.nl (M.N.M. van Lieshout), Radu.Stoica@avignon.inra.fr (R.S. Stoica).

reached its stationary distribution. Exact simulation methods in contrast automatically terminate 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 Häggström and Nelander (1998, 1999) and further extended in Møller (1999). 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 (Fill and Huber, 2000).

Murdoch and Green (1998) 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 (Fill, 1998; Fill et al., 2000), 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 (Møller, 1989; Preston, 1977; Ripley, 1977) or the Metropolis–Hastings paradigm (Geyer and Møller, 1994; Green, 1995; Ogata and Tanemura, 1984). 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 (Berthelsen and Møller, 2001; Ferrari et al., 2002; Kendall, 1998; Kendall and Møller, 2000). The second approach is a propose-accept strategy based on the local characteristics.

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 Häggström et al. (1999), Mira et al. (2001), Thőnnes (1999). 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.

The purpose of this paper is to generalise those exact simulation algorithms proposed for point processes that are based on the local characteristics to marked patterns, and to investigate their relative efficiency in terms of computational load by means of an extensive simulation study involving a range of well-known models. The coupling from the past and clan of ancestors algorithms for marked point processes are straightforward adaptations of existing algorithms (Kendall and Møller, 2000; Ferrari et al., 2002), although the speed-up that can be gained in practice by use of the incompatibility index to form the ancestor set has received little attention. The Gibbs sampler and Metropolis–Hastings algorithm require more care, as the discretisation involved depends on the mark space and various strategies can be imagined. Our choices for the Gibbs sampler are being inspired by ideas in estimation theory (Goulard et al., 1996); for the Metropolis–Hastings algorithm, we try to stay as close as possible to the classic Metropolis updates.

The plan of this paper is as follows. We briefly review basic facts on marked point processes in Section 2. Then, in Section 3.1, we discuss coupling from the past, turn to the clan of ancestors algorithm in Section 3.2, develop a Metropolis–Hastings algorithm in Section 4, and consider Gibbs sampling of a lattice approximation in Section 5. For the novel Metropolis–Hastings and Gibbs samplers, full proofs of convergence are included. Results on the speed of convergence are reported in Section 6. The paper closes with a critical evaluation.

2. Marked point processes

2.1. Definitions

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 (Daley and Vere-Jones, 1988). In other words, realisations of Y 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 $i = 1, \ldots, n$, with $k_i \neq k_j$ for $i \neq j$.

Let ν_M be a probability measure on the Borel σ-algebra $\mathscr{B}(M)$. 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.

The *Papangelou conditional intensity* of a marked point process with density f is defined for $(k, m) \in (K \times M) \backslash \mathbf{y}$ 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 $\{\mathbf{y} : f(\mathbf{y}) = 0\}$. It may be interpreted as the conditional probability of finding a point at k with mark m conditional on the configuration elsewhere being $\mathbf{y} \backslash \{(k, m)\}$.

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 the next subsections satisfy these assumptions. A density is said to be repulsive if $\lambda((k, m); \mathbf{y})$ is decreasing in its second argument with respect to set inclusion, attractive if it is increasing. Such properties correspond to our intuitive notion of e.g. repulsion since in that case the more marked points there are, the harder it is to introduce yet another one, and the smaller the conditional intensity.

2.1.1. Widom–Rowlinson mixture model

The Widom–Rowlinson mixture model for penetrable spheres (Widom and Rowlinson, 1970) has mark space $M = \{1, 2\}$ and density

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

with respect to a unit rate Poisson process labelled according to the symmetric Bernoulli distribution. Thus, particles with a different label keep at least a distance $R > 0$ away from each other, causing particles of the same type to cluster in the holes left by the configuration of particles of the other type. The parameters $\beta_m > 0$ govern the intensity of type m points, and $\alpha = \alpha(\beta_1, \beta_2, R) \in (0, \infty)$ is the normalising constant that ensures that f integrates to 1. The conditional intensity for adding a point of type 1 at $k \in K$ to the multi-type point pattern $\mathbf{y}$ that does not contain $(k, 1)$ 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 the addition of a point of type 2. The Widom–Rowlinson model is hereditary and locally stable with $\Lambda = \max\{\beta_1, \beta_2\}$. Furthermore, $\lambda((k, m); \mathbf{y}') \geqslant \lambda((k, m); \mathbf{y})$ for all $\mathbf{y}' \subseteq \mathbf{y}$ and all $(k, m) \in K \times M$.

2.1.2. Multi-type pairwise interaction process

Consider the mark space $M = \{1, \ldots, I\}$ for some $I \in \mathbb{N}$ equipped with the uniform probability distribution ν_M. Multi-type pairwise interaction processes (Baddeley and Møller, 1989; Ripley and Kelly, 1977) are defined by a density f with respect to the dominating Poisson process that is of the form

$$f(\mathbf{y}) = \alpha \prod_{(k,m) \in \mathbf{y}} \beta_m \prod_{(u,i) \neq (v,j) \in \mathbf{y}} \gamma_{ij}(\|u - v\|) \tag{2}$$

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 $i, j \in M$. Model (2) is hereditary and locally stable with upper bound $\Lambda = \max_{m \in M} \beta_m$. The Papangelou conditional intensity

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

$(k, m) \notin \mathbf{y}$, is decreasing in $\mathbf{y}$ with respect to the inclusion order.

2.1.3. Candy model

The Candy model (Stoica et al., 2004; Lieshout and Stoica, 2003) is a line segment process. The segments are characterised by the position of their centre, their length $l \in [l_{\min}, l_{\max}]$ for some $0 < l_{\min} < l_{\max} < \infty$, and their orientation $\theta \in [0, \pi)$. The orientation space is equipped with the complete metric $\rho(\theta, \theta') = \min\left\{ |\theta - \theta'|, \pi - |\theta - \theta'| \right\}$ that identifies 0 and π. Thus, 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

$$f(\mathbf{y}) = \alpha \beta^{n(\mathbf{y})} \prod_{i=1}^{n(\mathbf{y})} \exp\left[\frac{l_i - l_{\max}}{l_{\max}} \right] \times \gamma_{\mathrm{r}}^{n_{\mathrm{r}}(\mathbf{y})} \gamma_{\mathrm{o}}^{n_{\mathrm{o}}(\mathbf{y})} \tag{3}$$

with respect to a unit rate Poisson process marked uniformly and independently. The model parameters are $\gamma_{\mathrm{r}}, \gamma_{\mathrm{o}} \in (0, 1)$ and $\beta > 0$. The sufficient statistics $n(\mathbf{y})$, $n_{\mathrm{r}}(\mathbf{y})$ and $n_{\mathrm{o}}(\mathbf{y})$ 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 given $\tau, \delta \in (0, \pi/2)$, define the relation $\sim_{\mathrm{r}}$ on $K \times M$ by

$$y = (k, l, \theta) \sim_{\mathrm{r}} y' = \left(k', l', \theta'\right)$$

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

Then $n_{\mathrm{r}}(\mathbf{y})$ is the number of pairs of different points in $\mathbf{y}$ that are $\sim_{\mathrm{r}}$-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 $\sim_{\mathrm{o}}$ on $K \times M$ as follows: $y \sim_{\mathrm{o}} y' \Leftrightarrow \left\| k - k' \right\| > \max\left\{l, l'\right\}/2$ and either exactly one endpoint of y is a member of $Z\left(y'\right)$ or exactly one endpoint of y' is a member of $Z(y)$. Then $n_{\mathrm{o}}(\mathbf{y})$ is the number of $\sim_{\mathrm{o}}$ neighbour pairs in $\mathbf{y}$ with the property that $\rho\left(\theta, \theta'\right) > \tau$. Generalisations of (3) may be obtained by distinguishing several types of connection between the segments. The Candy model density (3) is strictly positive, and locally stable with $\Lambda = \beta$. The conditional intensity at $\xi = (k, l, \theta)$ is given by

$$\lambda(\xi; \mathbf{y}) = \beta \quad \exp\left[\frac{l - l_{\max}}{l_{\max}} \right] \gamma_{\mathrm{r}}^{n_{\mathrm{r}}(\xi; \mathbf{y})} \gamma_{\mathrm{o}}^{n_{\mathrm{o}}(\xi; \mathbf{y})},$$

and is decreasing with respect to the inclusion order. Here we use the notation $n_{\mathrm{r}}(\xi; \mathbf{y})$ for the number of $\sim_{\mathrm{r}}$ neighbours of ξ in the marked point pattern $\mathbf{y}$. Similarly $n_{\mathrm{o}}(\xi; \mathbf{y})$ denotes the number of $\sim_{\mathrm{o}}$ neighbours of ξ in $\mathbf{y}$ for which the angle property holds.

3. Perfect simulation using spatial birth-and-death processes

In this section, we consider the generalisations of two exact sampling methods based on the classic idea to simulate from spatial point process models by means of running a spatial birth-and-death process (Møller, 1989; Preston, 1977; Ripley, 1977). The first technique, coupling from the past (Kendall, 1998; Kendall and Møller, 2000) carries over immediately to the marked case; regarding the clans of ancestors method (Ferrari et al., 2002), as the incompatibility index involved is heavily dependent on the interaction structure, and the latter of course on the type of mark, although the method carries over immediately from a theoretical point of view, in practice more care is needed.

The study of exact simulation algorithms was pioneered in the seminal paper by Propp and Wilson (1996). Their coupling from the past (CFTP) algorithm was developed with discrete probability distributions on graphs in mind, for example the critical Ising model. Imagine m parallel chains are run from some time $-T < 0$ until 0, one for each initial state. The chains are coupled so that each of them uses the same pseudo-random numbers. Hence, if at any stage the initial state influence has worn off, then all sample paths are identical at some time, and remain so until time 0. We say that the m chains have coalesced. Consequently, had we sampled for an infinite amount of time, the same result would have been obtained, an exact sample from the equilibrium distribution. Of course, it is not practically feasible to run many parallel chains, nor is our state space finite. However, a tool kit of tricks such as exploiting a partial order or stochastic bounds can be used, as exemplified by the algorithms studied in this paper.

3.1. Coupling from the past for marked point processes

Coupling from the past techniques for point processes were first described in Kendall (1998) for a special model, the general method for locally stable point processes can be found in Kendall and Møller (2000), while Lieshout and Baddeley (2002) present an adaptive variation on the theme. The dynamics are those of a spatial birth-and-death process (Preston, 1977) 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, $t \leqslant 0$, 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. $D(t-) = D(t) \backslash \{(k, m)\}$, the marked point (k, m) is added to $L^{-T}(t-)$ if $V_t \leqslant \alpha_{\min}\left(U^{-T}(t-), L^{-T}(t-), (k, m)\right)$, and to $U^{-T}(t-)$ if $V_t \leqslant \alpha_{\max}\left(U^{-T}(t-), L^{-T}(t-), (k, m)\right)$.
(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

$$\alpha_{\min}(U, L, (k, m)) := \min\{\lambda((k, m); \mathbf{y})/\Lambda : L \subseteq \mathbf{y} \subseteq U\}$$

and

$$\alpha_{\max}(U, L, (k, m)) := \max\{\lambda((k, m); \mathbf{y})/\Lambda : L \subseteq \mathbf{y} \subseteq U\}. \tag{4}$$

It should be stressed that values of the random processes generated in earlier runs are re-used when time is doubled.

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 $\lambda\left((k, m); \mathbf{y}'\right) \leqslant \lambda((k, m); \mathbf{y})$, respectively, $\lambda\left((k, m); \mathbf{y}'\right) \geqslant \lambda((k, m); \mathbf{y})$ for all $\mathbf{y}' \subseteq \mathbf{y}$ and all $(k, m) \in (K \times M) \backslash \mathbf{y}$.

Doing so, the birth acceptance probabilities (4) simplify considerably. Indeed, if f defines an attractive marked point process, $\Lambda \alpha_{\min}(U, L, (k, m)) = \lambda((k, m); L)$, respectively, $\Lambda \alpha_{\max}(U, L, (k, m)) = \lambda((k, m); U)$, for $(k, m) \notin U$. Similarly, for repulsive densities, $\Lambda \alpha_{\min}(U, L, (k, m)) = \lambda((k, m); U)$ and $\Lambda \alpha_{\max}(U, L, (k, m)) = \lambda((k, m); L)$.

The Widom–Rowlinson, multi-type pairwise interaction and Candy models are all repulsive, hence Algorithm 1 may be implemented conveniently.

3.2. 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 Ferrari et al. (2002) 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 $B(k, r) = \{(u, i) \in K \times M, \|u - k\| \leqslant r\}$. The equation is clearly satisfied for r equal to 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 $v_M(\cdot)$. 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 $\Lambda v \times v_M$ 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 thus obtained 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. $D(t-) = D(t)\backslash\{(k, m)\}$ 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.
   - Return $Y(0)$.

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, as will be illustrated in Section 6. In the formulation of the algorithm above, we added all marked points in the ball $B(k, r)$ to the ancestor set for ease of notation. However, not all such points necessarily influence the conditional intensity. Instead, the more refined *incompatibility index*

$$I((k, m), (l, n)) := \mathbf{1}\left\{\sup_{\mathbf{y}} |\lambda((k, m); \mathbf{y}) - \lambda((k, m); \mathbf{y} \cup \{(l, n)\})| > 0\right\} \tag{5}$$

could be used (see Section 6), that is, only marked points $(l, n) \neq (k, m)$ with $I((k, m), (l, n)) = 1$ are added to A in the backward sweep. Also, one might generate the ancestors themselves directly, see Ferrari et al. (2002). The formulation above in terms of space–time cylinders is an adaptation to our marked point process context of the presentation in Garcia (2000), see also Ferrari et al. (2002, p. 80), Berthelsen and Møller (2002).

The correctness proofs of Algorithms 1–2 are adaptations of their counterparts for classic (un-marked) point processes, and are left to the reader.

4. Exact Metropolis–Hastings algorithm

4.1. Description

For (unmarked) locally stable point processes, Kendall and Møller (2000) proposed a dominated coupling from the past algorithm using discretisation and Metropolis–Hastings dynamics depending on the current configuration. Here we propose an easier alternative with fixed proposal probabilities for each cell and generalise our algorithm to marked point processes.

Recall that a Metropolis–Hastings algorithm is a proposal-acceptance technique (Geyer and Møller, 1994). A typical update if the current state is the marked point pattern $\mathbf{y}$ (assumed to have positive density $f(\mathbf{y})$) is to opt for a birth with probability $p_b \in (0, 1)$, for a death with the complementary probability $p_d = 1 - p_b$. In case of a birth, a new marked point ξ is sampled from the distribution $v \times v_M/v(K)$, and the proposal is accepted with probability

$$\min\left\{1, \frac{v(K)\lambda(\xi; \mathbf{y})}{p(1 + n(\mathbf{y}))}\right\},$$

where $p = p_b/(1 - p_b)$. If a death is proposed and $\mathbf{y}$ is empty, the state $\mathbf{y}$ remains unchanged. Otherwise, a point ξ is selected uniformly from $\mathbf{y}$; its deletion is accepted with probability

$$\min\left\{1, \frac{pn(\mathbf{y})}{v(K)\lambda(\xi; \mathbf{y}\backslash\{\xi\})}\right\}.$$

For a proof that the dynamics described above yield an unbiased sample from f in the long run, see for example Geyer (1999) and the references therein. In order to transform these update dynamics into an exact simulation algorithm, care has to be taken, especially with respect to death transitions. Since dominated coupling into and from the past is based on the idea of designing a pair of chains, coupled to some dominating one, that maintains the inclusion order and merges eventually (Kendall and Møller, 2000), it would be convenient if the Hastings ratio for accepting death proposals would be 1, regardless of the current configuration. The ratio is clearly bounded from below by $p/(v(K)\Lambda)$, but the bound may be smaller than 1. To overcome this problem, following (Kendall and Møller, 2000), we discretise K and update in a stripwise fashion.

More specifically, let $K = \bigcup_{i=1}^{n_K} K_i$ be a finite partition such that $v(K_i) > 0$ for all cells K_i, $i = 1, \ldots, n_K$. If the current configuration is $\mathbf{y}$ and the strip to be visited is $K_i \times M$, with probability p_b^i a new marked point ξ is generated with a v-uniformly distributed location restricted to K_i and mark distribution v_M; with the complementary probability p_d^i, a randomly chosen marked point located in K_i is proposed for deletion—if there is such a point. We assume that $p_b^i = 1 - p_d^i \in (0, 1)$ for all $i = 1, \ldots, n_K$. Hastings ratios similar to those above are computed to decide whether to accept or not. The procedure is well defined on the set of configurations with strictly positive density, and converges to f (for a proof, see the next subsection).

In order to describe the coupling based on the dynamics just described, let us start with the dominating chain D in which all proposals are accepted. The D-chain visits strips at random with equal probabilities. If strip $K_i \times M$ is being visited, with probability p_b^i, a new marked point is generated with location and mark distributed according to the v-uniform distribution on K_i and to v_M, respectively; with the complementary probability $p_d^i = 1 - p_b^i$, a marked point is chosen uniformly among those with location in K_i and deleted from D with the provision that if $K_i \times M$ contains no marked points, nothing is done.

Provided the birth proposal probability for each strip is strictly positive and less than the proposal probability for a death, $0 < p_b^i < \frac{1}{2}$, the detailed balance equations for the number of marked points in strip i

$$\pi_i(n) p_b^i = \pi_i(n+1) p_d^i, \quad n = 0, 1, 2, \ldots,$$

have a unique solution

$$\pi_i(n) = \left(\frac{p_b^i}{p_d^i}\right)^n \left(1 - \frac{p_b^i}{p_d^i}\right), \tag{6}$$

that is, the number of marked points in $K_i \times M$ has a shifted geometric distribution with success probability $1 - p_b^i/p_d^i$. Given there are n marked points with locations in cell K_i, under the equilibrium distribution they are i.i.d. with locations following the probability distribution $v(\cdot)/v(K_i)$ and marks chosen according to v_M. Clearly, the chain is reversible, and patterns in different strips $K_i \times M$ are independent.

Next, turn to the target dynamics. The Hastings ratio for a death transition from $\mathbf{y} \neq \emptyset$ to $\mathbf{y} \setminus \{y_j\}$ with $y_j \in (K_i \times M) \cap \mathbf{y}$ is given by

$$\frac{p_i n(\mathbf{y} \cap (K_i \times M))}{v(K_i) \lambda(y_j; \mathbf{y} \setminus \{y_j\})},$$

where $p_i = p_b^i/p_d^i$ is the ratio of birth and death proposal probabilities for the strip $K_i \times M$, $i = 1, \ldots, n_K$. Note that if we restrict the Markov chain to the set $\{\mathbf{y} : f(\mathbf{y}) > 0\}$, as f is hereditary by assumption, the Hastings ratio is well-defined. A sufficient condition for it to be bounded from below by 1 is $p_i \geqslant \Lambda v(K_i)$ where Λ denotes the upper bound on the conditional intensity. Under this condition, proposals to delete a marked point are always accepted.

Let $L \subseteq U$ be two finite marked point patterns, and set, for $(k, m) \notin U$,

$$\alpha_{\min}(U, L, (k, m), i) := \min\left\{\frac{v(K_i) \lambda((k, m); \mathbf{y})}{p_i(1 + n(\mathbf{y} \cap (K_i \times M)))} : L \subseteq \mathbf{y} \subseteq U\right\}$$

and

$$\alpha_{\max}(U, L, (k, m), i) := \max\left\{\frac{v(K_i) \lambda((k, m); \mathbf{y})}{p_i(1 + n(\mathbf{y} \cap (K_i \times M)))} : L \subseteq \mathbf{y} \subseteq U\right\} \tag{7}$$

for the bounds on the Hastings ratio for the birth of a point (k, m) in strip i based on marked point patterns sandwiched in between L and U.

Based on the ingredients described above, we propose the following algorithm. As before, random variables set in previous runs are re-used when doubling time.

Algorithm 3. Let K be partitioned in cells in such a way that $0 < \Lambda v(K_i) < 1$ and fix p_b^i in such a way that $p_i \in [\Lambda v(K_i), 1)$ for all $i = 1, \ldots, n_K$. Let $V_t, t = -1, -2, \ldots$, be a family of independent, uniformly distributed random variables on $\{1, \ldots, n_K\}$ and let U_t 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 independent (6) distributed strip counts, and for which, given a strip contains n points, the locations are scattered v-uniformly over the cell and marked i.i.d. according to v_M.

(1) Extend $D(\cdot)$ backwards to time $-T$ as follows. With probability $p_d^{V_t}$ delete a randomly picked marked point from $K_{V_t} \times M$; otherwise add a marked point ξ_t with v-uniform location in K_{V_t} and mark distributed according to v_M independently of other random variables.
(2) Generate $L^{-T}(\cdot)$ (lower process) and $U^{-T}(\cdot)$ (upper process) forward in time as follows:
   - set $L^{-T}(-T) = \emptyset$ and $U^{-T}(-T) = D(-T)$;
   - in case $D(\cdot)$ experiences a backward birth, i.e. $D(t) = D(t+1) \cup \{(k, m)\}$, a point is deleted from $L^{-T}(t)$ and $U^{-T}(t)$ according to the random permutation mechanism of Kendall and Møller (2000);
   - if $D(\cdot)$ experiences a backward death, i.e. $D(t) = D(t+1) \setminus \{(k, m)\}$, the marked point (k, m) is added to $L^{-T}(t)$ if $U_t \leqslant \alpha_{\min} (U^{-T}(t), L^{-T}(t), (k, m), V_t)$ and to $U^{-T}(t)$ if $U_t \leqslant \alpha_{\max} (U^{-T}(t), L^{-T}(t), (k, m), V_t)$.
(3) If $U^{-T}(0) = L^{-T}(0)$ stop. Else set $T = 2T$ and repeat.
(4) Return $U^{-T}(0)$.

If f is repulsive, (7) reduce to $\alpha_{\min}(U, L, (k, m), i) = (v(K_i) \lambda((k, m); U)) / (p_i + p_i n (U \cap (K_i \times M)))$, respectively, to $\alpha_{\max}(U, L, (k, m), i) = (v(K_i) \lambda((k, m); L)) / (p_i + p_i n (L \cap (K_i \times M)))$. In the attractive case, no such reduction is possible in general, but we may use the slightly looser bounds $\alpha_{\min}(U, L, (k, m), i) = (v(K_i) \lambda((k, m); L)) / (p_i + p_i n (U \cap (K_i \times M)))$ and $\alpha_{\max}(U, L, (k, m), i) = (v(K_i) \lambda((k, m); U)) / (p_i + p_i n (L \cap (K_i \times M)))$.

4.2. Proof of correctness

We need to check the conditions in Kendall and Møller (2000, Theorem 2.1).

In order to do so, first we have to specify the dynamics of the dominating process, and its associated random variables, in greater detail. Apart from the V_t and U_t used in handling the selection of strips and forward births in subprocesses, associate with each $D(t)$ a random permutation $\Sigma(t)$ of its marked points in the strip given by V_t. This permutation serves to specify the order in which points may die (in the forward sense), both in D and its subprocesses. Thus, in case of a death proposal at time t, if $D(t) \cap (K_{V_t} \times M) = \mathbf{y} = \{y_1, \ldots, y_n\}$ and $\Sigma(t) = (y_{j_1}, \ldots, y_{j_n})$ for some permutation $(j_1, \ldots, j_n)$ of the set $\{1, \ldots, n\}$, the point to die in the dominating process is y_{j_1}. The distribution of the Σ-process is specified conditionally on D and V as follows. If $D(t+1)$ is obtained from $D(t)$ by a death in the strip indexed by V_t, assign rank 1 to the marked point that is being removed, and let the ranks of the marked points in $D(t+1)$ with location in the cell indexed by V_t follow a uniformly distributed permutation, independently of all other random variables. Otherwise, let $\Sigma(t)$ be a uniformly distributed permutation of the marked points in $D(t) \cap (K_{V_t} \times M)$. Thus, $\Sigma(t)$ depends on V_t, $D(t)$ and $D(t+1)$. The time-reversibility of $D(\cdot)$, the independence of the U_t and V_t, and the construction of $\Sigma(\cdot)$ now imply the joint stationarity of $\{(D(t), V_t, \Sigma(t), U_t) : -\infty < t < \infty\}$. Note that conditionally given $D(t) = \mathbf{y}$ and $V_t = i$, permutation σ of $\{1, \ldots, n(\mathbf{y} \cap (K_i \times M))\}$ has marginal conditional probability $P(\sigma | \mathbf{y}, i)$ equal to

$$p_b^i \frac{1}{n(\mathbf{y} \cap (K_i \times M))!} + p_d^i \frac{1}{n(\mathbf{y} \cap (K_i \times M))} \frac{1}{(n(\mathbf{y} \cap (K_i \times M)) - 1)!} = \frac{1}{n(\mathbf{y} \cap (K_i \times M))!},$$

i.e. a uniform distribution if $n(\mathbf{y} \cap (K_i \times M)) > 0$, and a degenerate one otherwise. Sample paths can be generated as follows. In $D(0)$, cell i contains a shifted geometrically (6) distributed number of points (with success probability $1 - p_i$) that are scattered v-uniformly and marked according to v_M independently of each other and of other cells.

The random number U_0 is uniformly distributed on $(0, 1)$, and V_0 is uniformly distributed on the set $\{1, \dots, n_K\}$, independently of other sources of randomness. Because of its time reversibility, D may easily be extended backwards into the past, say up to $-T$, and forward, say up to T. The same is true for the i.i.d. U_t and V_t. Conditionally on the paths of D and V, realisations of $\Sigma(-T), \dots, \Sigma(-1), \Sigma(0), \dots, \Sigma(T-1)$ are then generated in a forward direction as explained above.

Deaths in subprocesses are handled by the removal of that marked point that is assigned the lowest rank by Σ. We need to verify that this choice amounts to the deletion of a uniformly distributed point from those in the strip being updated. Thus, let Σ be a uniformly distributed permutation of $A = \{y_1, \dots, y_n\}$, and consider its restriction to the subset $B = \{y_{i_1}, \dots, y_{i_m}\}$ of A. Note that each permutation σ of the n marked points has probability $1/n!$, that the order of the m points in B is fixed by σ, and that there are $(n-m)!$ permutations to fill the remaining ranks that can be chosen in $\binom{n}{n-m}$ different ways consistent with σ. Hence any permutation of B has probability $1/m!$, as it should. Moreover, a death transition implemented in this way respects the inclusion order: after the deletion of the point with lowest Σ-rank, the modified B is a subset of the modified A. In practice, only those parts of the permutation that are strictly necessary are generated, as in Kendall and Møller (2000, procedure `MHDeath`).

In summary, the upper and lower processes are adapted functionals of the stationary marked dominating process $\{(D(t), V_t, \Sigma(t), U_t) : -\infty < t < \infty\}$. By construction the sandwiching and funnelling properties Kendall and Møller (2000, (2.4)–(2.6)) hold.

Secondly, since the shifted geometric probability distribution assigns positive mass to 0, and the dominating process is in equilibrium, D extended backwards will almost surely reach state $\emptyset$, so that the algorithm will almost surely terminate.

It remains to consider the target process $Y^{-T}(\cdot)$ defined as follows. Set $Y^{-T}(-T) = \emptyset$ and apply the same dynamics as for $U^{-T}(\cdot)$ and $L^{-T}(\cdot)$ except that if $Y^{-T}(t) = \mathbf{y}$ the birth at time t of a marked point (k, m) in the strip indexed by V_t is accepted if

$$U_t \leqslant \frac{v\left(K_{V_t}\right) \lambda((k, m); \mathbf{y})}{p_{V_t}\left(1 + n\left(\mathbf{y} \cap \left(K_{V_t} \times M\right)\right)\right)}.$$

Clearly, $L^{-T}(t) \subseteq Y^{-T}(t) \subseteq U^{-T}(t)$ for all integers $0 \geqslant t \geqslant -T$.

The process $Y(t)$ for $t = 0, 1, 2, \dots$ with $Y(0) = \emptyset$ is defined analogously based on a forward run of the dominating process and its associated random variables. Note that $Y(t)$ exhibits the dynamics of a Metropolis–Hastings Markov process. If the current configuration is $\mathbf{y}$, a strip $K_i \times M$ is chosen uniformly. With probability p_b^i a new marked point ξ is generated in $K_i \times M$ from the probability distribution $v \times v_M / v(K_i)$, and accepted with probability $(v(K_i) \lambda(\xi; \mathbf{y})) / (p_i + p_i n(\mathbf{y} \cap K_i \times M)) \leqslant 1$, where as before $p_i = p_b^i / (1 - p_b^i)$. With the complementary probability p_d^i, a randomly chosen marked point located in K_i is deleted, unless the strip is empty in which case no update occurs. The procedure is well defined on the set of configurations with strictly positive density.

By construction, f is an invariant density for the transition kernel. Since deaths are always accepted, and $p_d^i > \frac{1}{2}$, for any marked point pattern $\mathbf{y}$, the probability of reaching the empty set is at least $(2n_K)^{-n(\mathbf{y})} > 0$. Hence the Metropolis–Hastings sampler is irreducible with respect to the Dirac measure δ_0 on $\emptyset$, and consequently with respect to f (cf. Geyer, 1999, Section 3.8). Moreover, if the current state is $\emptyset$, the probability of staying put is at least $\sum_i p_d^i / n_K > \frac{1}{2} \geqslant (2n_K)^{-1}$. Consequently, sets of the form $\{\mathbf{y} : n(\mathbf{y}) \leqslant n\}$ (and their measurable subsets) are small with respect to $(2n_K)^{-n}\delta_0$. Since self-transitions occur with positive probability, the sampler is aperiodic. Therefore, $Y(t)$ converges to the distribution specified by f weakly and in total variation (note that the convergence holds for f-almost all initial states, but recall $Y(0) = \emptyset$, the atom of the limit distribution). To get rid of the null set, we may use the drift condition Meyn and Tweedie (1993, Theorem 9.1.8). Define a function V by

$$V(\mathbf{y}) = c(\mathbf{y}) A^{m(\mathbf{y})},$$

where $m(\mathbf{y}) = \max_i n(\mathbf{y} \cap (K_i \times M))$, $c(\mathbf{y})$ is the multiplicity of the maximum, and $A > n_K$. Note that V is strictly positive, and its level sets $\{\mathbf{y} : V(\mathbf{y}) \leqslant c\}$ for $c > 0$ are small. Indeed, $V(\mathbf{y}) \leqslant c$ implies that the cardinality of $\mathbf{y}$ is bounded by $n_K \log(c) / \log(A)$ or 0 depending on whether c exceeds A or not. Now, for any $\varepsilon > 0$, one may choose an integer $N_\varepsilon > 0$ such that the acceptance probability of a birth in a near-optimal slice $K_i \times M$ with $n(\mathbf{y} \cap (K_i \times M)) \geqslant m(\mathbf{y}) - 1$

does not exceed ε for any $\mathbf{y}$ with $n(\mathbf{y}) \geqslant N_\varepsilon$. Updates in strips with less than $m(\mathbf{y}) - 1$ marked points do not affect the value of $V(\cdot)$. For $\mathbf{y}$ with $n(\mathbf{y}) \geqslant N_\varepsilon$,

$$\mathbb{E}\left[V\left(Y\left(t+1\right)\right)|Y(t) = \mathbf{y}\right] - V(\mathbf{y})$$

$$\leqslant \frac{\max_i p_{\mathrm{b}}^i}{n_K}\left(A + n_K - 2\right)\varepsilon V(\mathbf{y}) - \frac{\min_i p_{\mathrm{d}}^i}{n_K}\left(1 - \frac{n_K}{A}\right)V(\mathbf{y}).$$

One may choose ε such that $0 < \max_i p_{\mathrm{b}}^i (A + n_K - 2)\varepsilon < \min_i p_{\mathrm{d}}^i (1 - n_K/A)$, in which case the expected increase in V is negative. We conclude that the Y-chain is Harris recurrent, and the proof is complete.

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 (Häggström et al., 1999), 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 (Lieshout, 2000). Lattice approximations of marked point processes on the other hand are quite common, especially in the context of pseudo-likelihood estimation (Besag et al., 1982; Goulard et al., 1996), and for such processes, Gibbs sampling is the natural way to proceed. Some careful tuning is necessary, though, due to the fact that not every point process on a product space can be seen as a marked point process.

5.1. Approximation by a lattice process

As in Section 4, let $K = \bigcup_{i=1}^{n_K} K_i$ be a finite partition of the location space such that $0 < v(K_i) < \infty$ for all $i = 1, \ldots, n_K$, and suppose the mark space is partitioned similarly as $M = \bigcup_{j=1}^{n_M} M_j$ with $0 < v_M(M_j) \leqslant 1$ for $j = 1, \ldots, n_M$. Hence, the induced partition of the product space $K \times M = \bigcup_{i=1}^{n_K} \bigcup_{j=1}^{n_M} (K_i \times M_j)$. We also choose a representative ξ_{ij} for each marked cell $K_i \times M_j$, for example the centroid, and denote the set of all representatives by Ξ.

Following (Besag et al., 1982; Goulard et al., 1996), we approximate the density f of Y by an auto-logistic distribution. More specifically, 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 $\sum_{j=1}^{n_M} n_{ij} \leqslant 1$ for all $i = 1, \ldots, n_K$. If we interpret n_{ij} as a presence/absence indicator for 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 $N_i = (N_{i1}, \ldots, N_{i n_M})$ almost surely takes values in $S := \{\mathbf{0}, \mathbf{e}_1, \ldots, \mathbf{e}_{n_M}\}$, the set consisting of the unit and zero vectors.

In order to specify the joint probability mass function $\mathbb{P}$ of the random field N, recall that, since f is hereditary by assumption, by the Hammersley–Clifford theorem (Baddeley and Møller, 1989; Lieshout, 2000; Ripley and Kelly, 1977) it can by factorised as a product

$$f(\mathbf{y}) = \prod_{\mathbf{z} \subseteq \mathbf{y}} \varphi(\mathbf{z}) \tag{8}$$

of so-called interaction functions φ. Indeed, the multi-type pairwise interaction process density (2) is already given in this form, with $\varphi(\mathbf{z}) = 1$ for patterns $\mathbf{z}$ that consist of more than two marked points. For the Widom–Rowlinson density (1), $\varphi(\emptyset) = f(\emptyset) = \alpha$, $\varphi(\{(k, m)\}) = \beta_m$, and $\varphi(\{(u, 1), (v, 2)\}) = \mathbf{1}\{\|u - v\| > R\}$. The Candy density (3) satisfies (8) with $\varphi(\emptyset) = f(\emptyset) = \alpha$, $\varphi(\{y\})$ equal to $\beta \exp\left[(l - l_{\max})/l_{\max}\right]$ for a singleton $y = (k, l, \theta)$, and $\varphi(\{y, y'\})$ equal to γ_{r} if $y \sim_{\mathrm{r}} y'$, to γ_{o} if $y \sim_{\mathrm{o}} y'$ and the pair $\{y, y'\}$ satisfies the small orientation distance constraint. For other configurations $\mathbf{z}$, $\varphi(\mathbf{z}) = 1$.

The probability mass function of N is obtained from (8) by restricting the product to subsets of Ξ, and correcting for the mass of the partition cells (Goulard et al., 1996), that is,

$$\mathbb{P}(\mathbf{n}) \propto \prod_{\{i,j\}} \left[v(K_i)v_M(M_j)\right]^{n_{ij}} \prod_{\emptyset \neq \mathbf{z} \subseteq \Xi} \varphi(\mathbf{z})^{\eta(\mathbf{n}; \mathbf{z})} \tag{9}$$

for $\mathbf{n}$ in the state space. Here, $\eta(\mathbf{n}; \mathbf{z}) = \prod_{\xi_{ij} \in \mathbf{z}} n_{ij}$, which takes values 0 or 1. By convention, $0^0 = 1$.

5.2. Local characteristics

The local behaviour of a random field is captured by its one point conditional probabilities, which we refer to in the sequel as the local characteristics. 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

$$\mathbb{P}\left(N_i = \mathbf{e}_j | \mathbf{n}_{-i}\right) = \frac{v\left(K_i\right) v_M\left(M_j\right) \prod_{\mathbf{z}} \varphi\left(\mathbf{z} \cup \left\{\xi_{ij}\right\}\right)^{\eta(\mathbf{n}_{-i}; \mathbf{z})}}{1 + v\left(K_i\right) \sum_{j'=1}^{n_M} v_M\left(M_{j'}\right) \prod_{\mathbf{z}} \varphi\left(\mathbf{z} \cup \left\{\xi_{ij'}\right\}\right)^{\eta(\mathbf{n}_{-i}; \mathbf{z})}} \tag{10}$$

for $j \neq 0$, provided the conditioning event has strictly positive mass under $\mathbb{P}$. The products range over $\mathbf{z} \subseteq \Xi \setminus \left\{\xi_{ij'} : j' = 1, \ldots, n_M\right\}$, and $\eta\left(\mathbf{n}_{-i}; \mathbf{z}\right)$ is defined as before. Often, as in the examples presented in Section 2, (10) does not involve all $\xi_{i'j}$, $i' \neq i$, but only a small subset near ξ_{ij} (which may be called the neighbourhood of ξ_{ij} or the marked cell it represents) as the interaction function φ reduces to 1 for most configurations containing ξ_{ij}. In this case, (10) is truly local.

The random field N is said to be monotone with respect to a given order $\leqslant$ on S if for all $s \in S$, $\mathbb{P}\left(N_i \leqslant s | \mathbf{n}_{-i}\right) \geqslant \mathbb{P}\left(N_i \leqslant s | \tilde{\mathbf{n}}_{-i}\right)$ whenever $\mathbf{n}_{-i} \leqslant \tilde{\mathbf{n}}_{-i}$ stripwise, $\mathbb{P}\left(\mathbf{n}_{-i}\right) > 0$ and $\mathbb{P}\left(\tilde{\mathbf{n}}_{-i}\right) > 0$. The Widom–Rowlinson mixture model is monotone with respect to the total order $\mathbf{e}_1 < \mathbf{0} < \mathbf{e}_2$ on S (Häggström and Nelander, 1999). The same is true for bivariate pairwise interaction processes for which $\gamma_{ii} \equiv 1$ and $\gamma_{ij}(t) > 0$ for all $t > 0$.

To see this, consider a bivariate pairwise interaction process with density (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

$$\mathbb{P}\left(N_i = \mathbf{e}_1 | \mathbf{n}_{-i}\right) \geqslant \mathbb{P}\left(N_i = \mathbf{e}_1 | \tilde{\mathbf{n}}_{-i}\right)$$

and

$$\mathbb{P}\left(N_i = \mathbf{e}_2 | \mathbf{n}_{-i}\right) \leqslant \mathbb{P}\left(N_i = \mathbf{e}_2 | \tilde{\mathbf{n}}_{-i}\right)$$

for all i and all pairs of configurations for which $\mathbf{n}_{-i} \leqslant \tilde{\mathbf{n}}_{-i}$ stripwise. With the notation $\xi_{i1} = (k_i, 1)$, $\xi_{i2} = (k_i, 2)$ for the marked representatives of cell K_i, and $\alpha_i = v\left(K_i\right)/2$, note that $\mathbb{P}\left(N_i = \mathbf{e}_1 | \mathbf{n}_{-i}\right)$ is equal to

$$\frac{\alpha_i \beta_1 \prod_{k \in n_{-i}(2)} \gamma_{12}\left(\|k_i - k\|\right)}{1 + \alpha_i \beta_1 \prod_{k \in n_{-i}(2)} \gamma_{12}\left(\|k_i - k\|\right) + \alpha_i \beta_2 \prod_{k \in n_{-i}(1)} \gamma_{12}\left(\|k_i - k\|\right)}$$

$$= \frac{1}{\left[1/\left(\alpha_i \beta_1\right) + \left(\beta_2/\beta_1\right) \prod_{k \in n_{-i}(1)} \gamma_{12}\left(\|k_i - k\|\right)\right] \prod_{k \in n_{-i}(2)} \gamma_{12}(\|k_i - k\|)^{-1} + 1},$$

where we use the notation $n_{-i}(m)$ for the locations of representatives $\xi_{i'j}$, $i' \neq i$, that are assigned mark $m \in \{1, 2\}$ in $\mathbf{n}$. 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.

For the Candy model, we have been unable to find any order in the local characteristics given by

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

where the pair interaction function φ is as computed in the beginning of this section. Here representatives ξ_{ij} are of the form $\left(k_i, m_j\right)$, with $m_j = \left(l_j, \theta_j\right)$.

5.3. Gibbs sampler for lattice process

The exact Gibbs sampler of Häggström and Nelander (1999) for random fields on a finite lattice is based on the coupling from the past idea (Propp and Wilson, 1996). However, they suggested, in order to avoid having to run

many coupled chains simultaneously, to keep track of the set of possible values at each time and each lattice cell (or, equivalently, each representative). The state space of the Markov chain thus obtained consists of subsets of the value set of the random field distribution to sample from. In our case, the lattice is represented by Ξ, and each n_{ij} takes values in $\{0, 1\}$.

A naive implementation of the exact Gibbs sampler described above, however, would be rather inefficient, since it does not exploit the constraints that $\sum_{j=1}^{n_M} n_{ij} \leqslant 1$ for all $i = 1, \ldots, n_K$. The solution we propose below implements random strip updating. In more detail, suppose that the current state is $\mathbf{n} = (\mathbf{n}_1, \ldots, \mathbf{n}_{n_K})$ where $\mathbf{n}_i \in S$ is the vector of values n_{ij}, $j = 1, \ldots, n_M$, in strip i. Thus, each $\mathbf{n}_i$ lies in the n_K fold Cartesian product $S^{\otimes n_K}$ of S with itself. At each time t, a strip $K_i \times M$ is chosen at random, its current value $\mathbf{n}_i$ replaced by a sample from the local characteristic (10), and the values at other strips are left unchanged. Since S has finitely many values, a single pseudo-random number u suffices to perform the update once the strip is selected. In terms of a transition kernel, say ψ, the new state is $\psi(\mathbf{n}, u, i)$.

If we apply coupling from the past to chains from every initial state (of positive probability), and keep track of the union of possible states at each point in time rather than of individual paths, we obtain the following algorithm. As before, random variables set in previous runs are re-used when doubling time.

Algorithm 4. Let $V_t, t = -1, -2, \ldots$, be a family of independent, uniformly distributed random variables on $\{1, \ldots, n_K\}$ and let U_t be a family of independent, uniformly distributed random variables on $(0, 1)$. Suppose that $\psi : S^{\otimes n_K} \times (0, 1) \times \{1, \ldots, n_K\} \to S$ is a transition kernel for the Gibbs sampler, that is, $\psi(\mathbf{n}, U, v)$ alters the state $\mathbf{n}$ in the n_K-fold Cartesian product of S at strip v in such a way that the new state is distributed according to the $\mathbb{P}$-conditional distribution given $\mathbf{n}_{-v}$. Initialise $T = 1$.

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

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

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

By Häggström and Nelander (1999, Theorem 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 throughout this paper we assume that (8) is locally stable, for all $\mathbf{n}_{-i}$ with $\mathbb{P}(\mathbf{n}_{-i}) > 0$,

$$\mathbb{P}(N_i = \mathbf{0}|\mathbf{n}_{-i}) = \frac{1}{1 + v(K_i) \sum_{j=1}^{n_M} v_M(M_j) \prod_{\mathbf{z}} \varphi(\mathbf{z} \cup \{\xi_{ij}\})^{\eta(\mathbf{n}_{-i}; \mathbf{z})}}$$

$$\geqslant \frac{1}{1 + v(K_i) \sum_{j=1}^{n_M} v_M(M_j) \Lambda} \geqslant \frac{1}{1 + K^* \Lambda} > 0,$$

where $K^* = \max \{v(K_i) : i = 1, \ldots, n_K\}$, the lower bound condition holds.

We conclude this section with some remarks on implementation. Like Algorithm 1, Algorithm 4 is particularly appealing in case of monotonicity with respect to some conveniently chosen order (Propp and Wilson, 1996; Häggström and Nelander, 1998). In this case, if the transition kernel is chosen in such a way that it respects the order, for example

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

(which we shall use in the simulations in Section 6), Algorithm 4 effectively reduces to the original Propp–Wilson algorithm (1996) 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. The procedure yields an unbiased sample in finite time provided at least one of the two initial states has positive probability.

Furthermore, one often saves on computational effort if the new state $\mathcal{X}_{V_t}^{-T}(t+1)$ is computed explicitly only if one knows in advance that it is a singleton, and to set $\mathcal{X}_{V_t}^{-T}(t+1) = S$ otherwise (Häggström and Nelander, 1999; Kendall and Montana, 2002; Murdoch and Green, 1998). Indeed, since

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

uniformly in $i \in \{1, \ldots, n_K\}$ and $\mathbf{n}_{-i}$, we may choose the transition kernel governing Algorithm 4 in such a way that $\psi(\chi, u, v) = \mathbf{0}$ whenever $u \leqslant 1/\left(1 + \Lambda \max_i v\left(K_i\right)\right)$. Thus, for small realised values of U_t, regardless of $\mathcal{X}^{-T}(t)$, the strip V_t is assigned value $\mathbf{0}$. For $U_t > 1/\left(1 + \Lambda \max_i v\left(K_i\right)\right)$, if $\mathcal{X}_v^{-T}(t)$ is not a singleton for some neighbour v of the current strip V_t, in the sense that the local characteristic at V_t depends on v, no update is made, i.e. $\mathcal{X}_{V_t}^{-T}(t+1) = S$. The papers cited above use various names for such an approach, we shall refer to it as a 'multigamma' coupler.

6. Simulation study

6.1. Design

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 (1), the Candy model (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

$$\gamma_{12}(t) = \begin{cases} 1 - (1-\gamma)\mathbf{1}\{t \leqslant r\} & \textit{Strauss}; \\ 1 - 1/\left(1 + (t/\sigma)^2\right)^2 \mathbf{1}\{t \leqslant 3\sigma\} & \textit{Cauchy}. \end{cases}$$

All models are sampled on the unit square.

We plot for each of the four algorithms presented in the previous sections the average number of 'jumps' over 25 independent samples as a function of intensity and interaction range parameters. Note that different vertical scales were used in the figures.

For the Widom–Rowlinson model, the intensity parameter is $\beta_1 = \beta_2$, and the hard core distance R is taken as the range parameter r. For the bivariate Strauss and Cauchy models, the intensity parameter is $\beta_1 = \beta_2$ as well. Additionally, in the Strauss case, the strength of interaction is set to $\gamma = 0.5$. The range parameter of the Strauss interaction function is r, for the Cauchy model, the range is $r = 3\sigma$. The intensity parameter values start at 1 and increase with steps of 1 to a maximum that depends on the algorithm. The range starts at 0.005 and increases with steps of 0.005, except for the Gibbs sampler which uses a different discretisation (see Section 6.6).

The parameters of the Candy model are chosen as follows. The intensity parameter β has minimum value 1, and is incremented with step size 1. The orientation and rejection parameters are 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 $r = 1.25\,l_{\max}$. To see this, 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 $5l_{\max}/4$. In this simulation study, we assume the length distribution to be concentrated on $l_{\max}$, which we assign minimum value 0.005 and increment by steps of size 0.005, except for the Gibbs sampler.

6.2. Implementation

The marked point process models described in Section 2—as well as a range of others—have been implemented in C++ in the software library MPPLIB (Steenbeek et al., 2002–2003) developed by A.G. Steenbeek and others. The object oriented nature of this programming language is well suited to our purposes, as it allows us to define a few general classes for (marked) points, marked or classic point processes, and samplers, and to introduce derived classes for new models and algorithms when they are needed. The general classes contain (virtual) member functions for the basic operations, the derived classes contain model or algorithm specific variables and functions, and implement the virtual members of their parent class as appropriate.
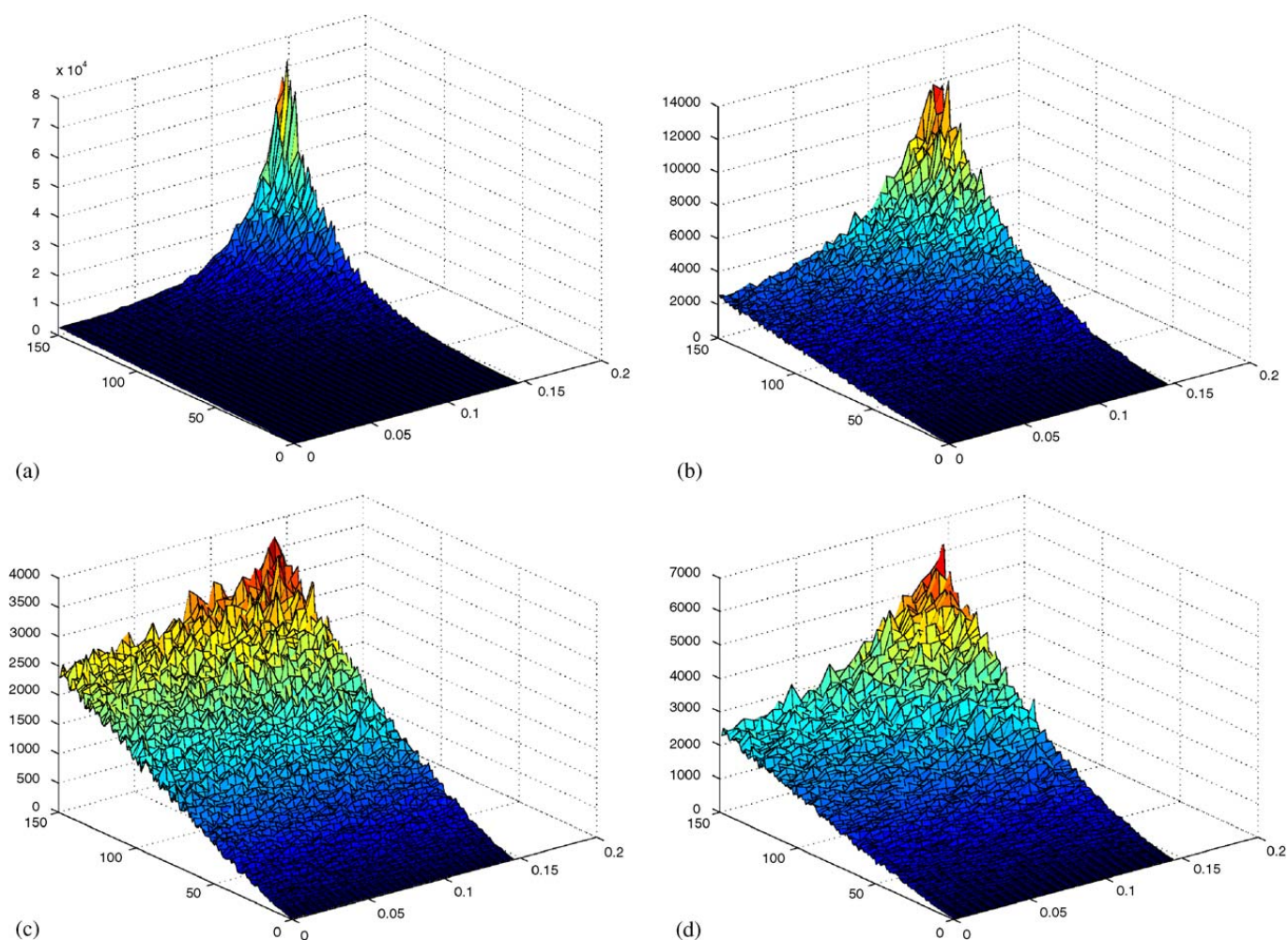
Fig. 1. Average number of jumps over 25 independent runs of the CFTP algorithm against intensity and range for (a) Widom–Rowlinson, (b) Strauss, (c) Cauchy, and (d) Candy.

For example the class `Prior` for a point process observed in a rectangle contains member functions for setting and retrieving the rectangle, and virtual ones for the computation of the conditional intensity, the setting or retrieving of model parameters, the generation of appropriate marked points, and for reporting samples or statistics.

Most Monte Carlo samplers encountered in the literature are based on a few simple updates: births and deaths of a single marked point. Some also allow for changes in the position or mark. However, as the details vary considerably between algorithms, these are hidden for the user in the non-public parts of the sampler classes. Public virtual member functions deal with initialisation, definition of the model to sample from, and generation of realisations. Since samplers often need to keep track of coupled sample paths, a clean up facility is essential. Discretisation of the mark space is model dependent and handled by the corresponding classes.

6.3. Coupling from the past

For Algorithm 1, a jump is defined as a birth or death transition in the dominating chain. Note that the actual number of calculations involved in a run of the algorithm is larger, due the successive doubling of time T. For example, if the coalescence time is $T = -4$, the jumps between $T = -2$ and $T = -4$ are thinned only once in the forward loop, those between $T = -1$ and $T = -2$ are thinned twice, and finally the jumps before $T = -1$ are thinned in three forward runs.

In Fig. 1 we plot the average number of jumps against intensity and range parameters. The upper bounds are 150 and 0.15, respectively, i.e. in the Candy model, the maximum length is 0.12. 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
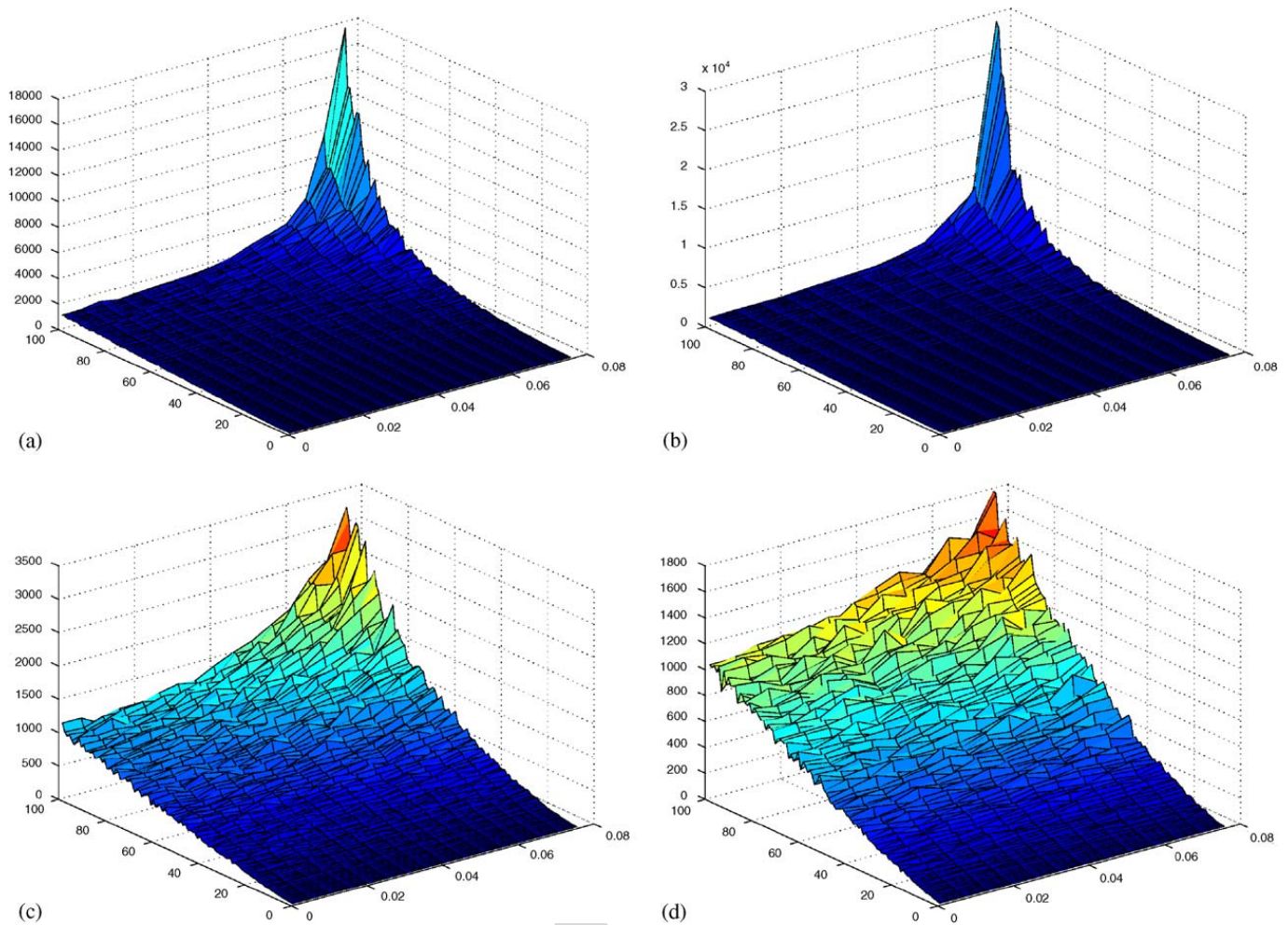
Fig. 2. Average number of jumps over 25 independent runs of the clan of ancestors algorithm against intensity and range for (a) multi-type pairwise interaction processes using the fixed range neighbourhood criterion, (b) Candy model using the fixed range neighbourhood criterion, (c) multi-type pairwise interaction processes using the incompatibility index, and (d) Candy model using the incompatibility index.

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.

6.4. Clan of ancestors

For Algorithm 2, a jump is defined as a birth or death transition in the backward sweep. The total number of jumps is a good measure of the computational load of the clan of ancestors algorithm, as no doubling scheme is involved.

The average number of jumps as a function of the parameter values is plotted in Figs. 2(a,b). Now, the maximal intensity is 100, the largest range is 0.08 for the pairwise interaction models and 1.25 times the maximal length 0.065 for the Candy model. 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.

Next, we use the incompatibility index (5) that describes the neighbourhood structure more precisely than a ball of fixed radius. Indeed, for the Widom–Rowlinson model and both multi-type pairwise interaction processes, the

Table 1
Average coalescence time and mean number of jumps for simulating 500 samples of a Poisson process with $\beta = 50$, depending on the proposal probabilities

p_b	Time	Jumps
0.01	34471.94	690.80
0.10	4816.90	965.64
0.25	4048.90	2023.12
0.45	24707.07	22233.11

Table 2
Average coalescence time and mean number of jumps for simulating 500 samples of a Poisson process with $\beta = 50$, with proposal probability $p_b = 0.1$

Mesh	Time	Jumps
25	6520.83	1304.47
30	9437.18	1890.16
35	13828.10	2768.14
40	17580.03	3518.75

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 $(l, n) \sim_o (k, m)$ when updating the ancestor set.

If we again take 25 independent samples and average the number of jumps for each pair of parameters, we obtain the plots in Figs. 2(c,d). The gain in efficiency due to smaller ancestor sets compared to those resulting from a fixed range neighbourhood is obvious, and more pronounced as 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 due to its different incompatibility index.

6.5. Metropolis–Hastings

Algorithm 3 leaves some freedom in choosing the partition in strips and the birth proposal probabilities. The effect of different choices is examined in the following experiments. We assume equal birth proposal probabilities $p_b^i \equiv p_b$, and partition the unit square $K = [0, 1] \times [0, 1]$ as an $m \times m$ grid of small squares of equal volume $1/m^2$.

In the first experiment, the mesh size m is adapted to p_b as $\lceil \sqrt{(\Lambda/p)} \rceil$ where $p = p_b/(1 - p_b)$. We simulate a Poisson point process of intensity $\Lambda = 50$ using Algorithm 3 for different values of p_b. Note that any independent marking may be applied without affecting the running time of the algorithm. For each p_b-value, 500 independent realisations are generated, and the average coalescence time computed. We also record the mean number of actual jumps in the dominating chain D by discarding the proposals to delete a point from an empty strip. The results are listed in Table 1.

Next, the influence of the mesh size m for fixed $p_b = 0.10$ is investigated (cf. Table 2). As in the previous case, 500 independent samples from a Poisson point process of rate 50 are used to compute the average coalescence time and number of jumps.

Note that for $p_b = 0.10$, $\lceil \sqrt{(\Lambda/p)} \rceil = 22$. Table 2 indicates that increasing m leads to longer coalescence times. Table 1 suggests that a long coalescence time is required for both small and large values of p_b, while intermediate values ($p_b = 0.10$ or 0.25) give good results.

In Fig. 3 we plot the mean coalescence time over 25 runs of the exact Metropolis–Hastings algorithm of Section 4 against intensity and range parameters. It should be noted that the actual number of calculations is larger, as the method is based on successive doubling. The upper bound on intensity is 150, the one on range 0.15. The birth proposal probability is $p_b = 0.1$, and the unit plane is divided into an $m \times m$ grid of equal volume cells with $m = \lceil \sqrt{(\Lambda/p)} \rceil$.
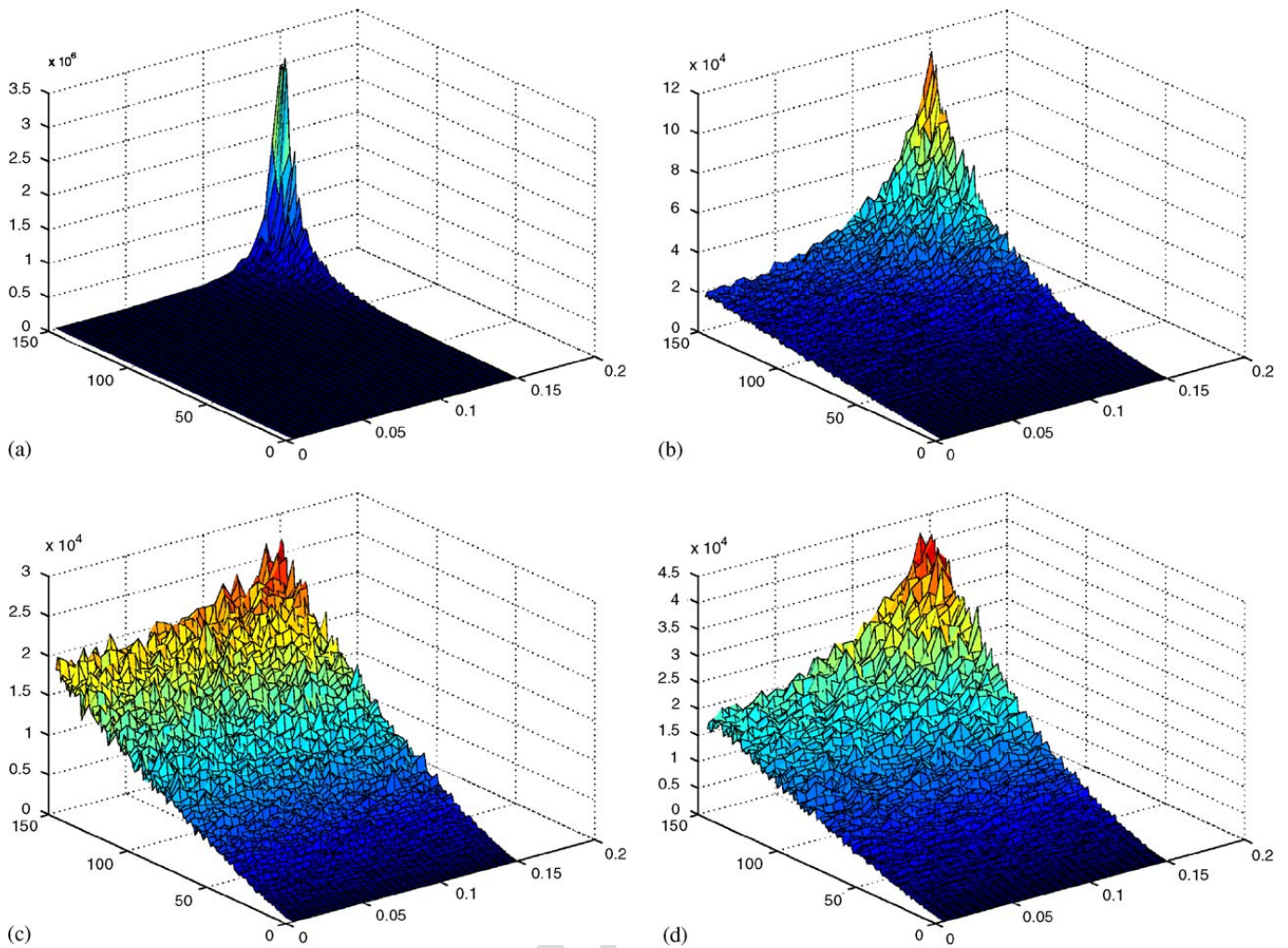
Fig. 3. Average coalescence time over 25 independent runs of the Metropolis–Hastings algorithm against intensity and range for (a) Widom–Rowlinson, (b) Strauss, (c) Cauchy, (d) Candy. In all cases, the mesh was adapted to the birth proposal probability $p_b = 0.10$.

Similar considerations as for Algorithm 1 apply: the stronger the interaction, the longer it takes to obtain a sample. The Metropolis–Hastings algorithm is slower than coupling from the past using spatial birth-and-death processes.

6.6. Gibbs sampler

For the Gibbs sampler, at each time step the state is updated. Thus, in Fig. 4, we plot the average coalescence time 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 use the monotone strip based Gibbs sampler. For the Candy model, we have chosen the stripwise multigamma coupling.

For all models, 20×20 lattices were used. Due to the discretisation effect, only a few range parameter values have to be considered. We take $r = 0.005, 0.055, 0.075, 0.105, 0.115$ and 0.155. The upper bound on the intensity parameter was 100. For the Widom–Rowlinson, as well as the bivariate Strauss and Cauchy models, the mark space is divided in two bins. Note that for the Cauchy model γ_{12} is not a step function, resulting in a somewhat different discretisation effect than for the other two models! For the Candy model, the lengths $l = 0.8 * r$ are given by the discretised values $r = 0.005, 0.055, 0.075$ and 0.105. The orientation interval is partitioned in five sub-intervals with equal lengths.
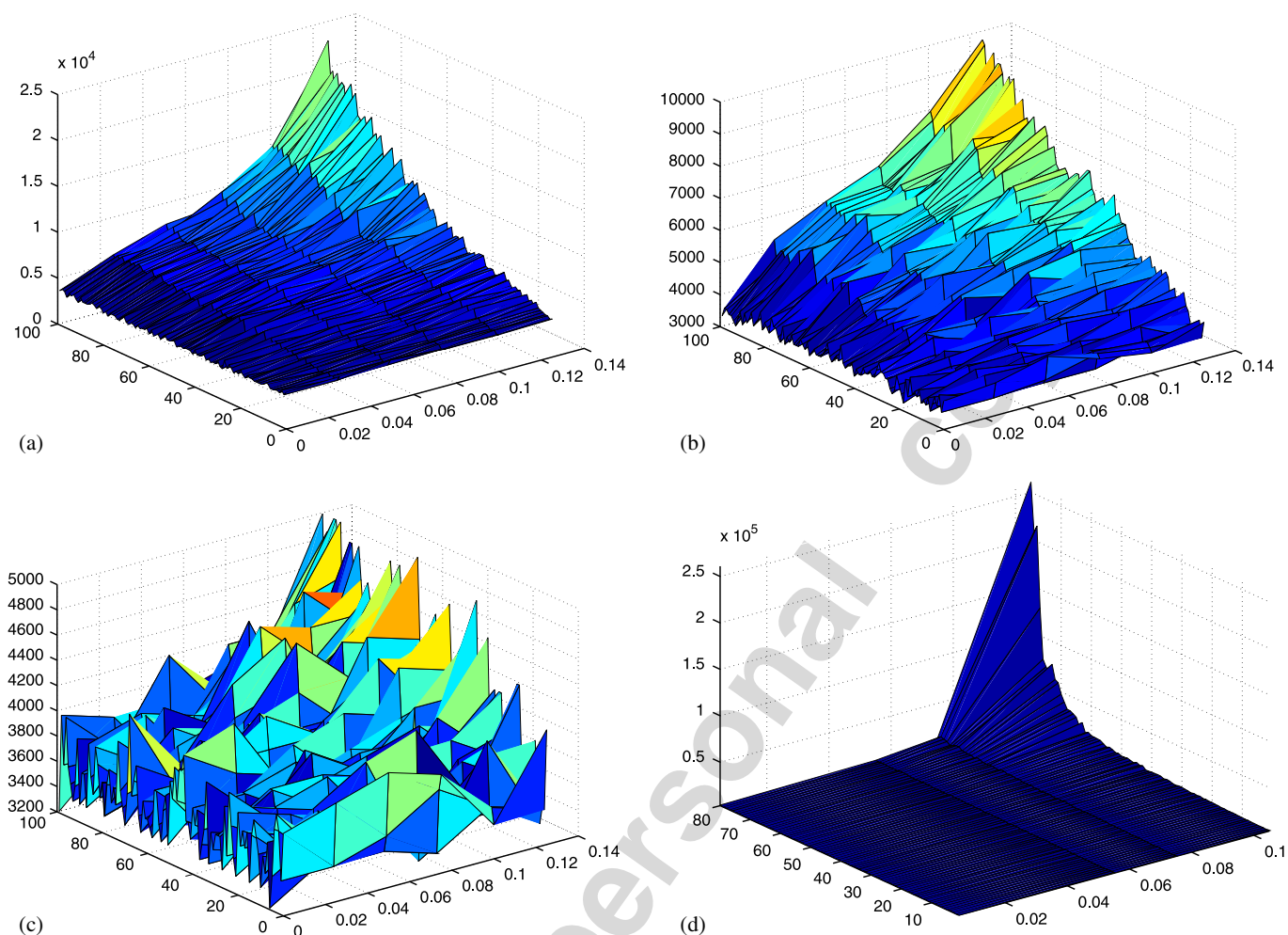
Fig. 4. Average coalescence time over 25 independent runs of the Gibbs sampler against intensity and range for (a) Widom–Rowlinson, (b) Strauss, (c) Cauchy, and (d) Candy. For (a), (b), and (c), the monotone strip based Gibbs sampler was used, for (d) the stripwise multigamma sampler.

Fig. 4 indicates 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 Fig. 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. Four families of algorithms were proposed: coupling from the past, the clan of ancestors technique, the Gibbs sampler and a Metropolis–Hastings algorithm based on birth and death proposals. The various algorithms were 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.

In summary, 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, and usually faster than the Metropolis–Hastings sampler. The latter two methods seem practically infeasible in general if the underlying model is neither attractive nor repulsive.

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, it would be interesting to investigate extensions of the Metropolis–Hastings dynamics (Kendall and Møller, 2000) that allow for the construction of change moves in order to improve the mixing properties of the simulated Markov chain. Speed performances of the algorithms can be improved if the Markovian structure of the models can be efficiently exploited. On a longer perspective, we would like to use those ideas presented in Steinsland (2006) that may lead to parallel algorithms for exact simulation applied to marked point processes.

Acknowledgements

References

Baddeley, A.J., Møller, J., 1989. Nearest-neighbour Markov point processes and random sets. Internat. Statist. Rev. 57, 89–121.

Barndorff-Nielsen, O.E., Cox, D.R., Klüppelberg, C., 2001. Complex Stochastic Systems. Chapman and Hall, CRC, Boca Raton.

Berthelsen, K.K., Møller, J., 2001. Spatial Jump Processes and Perfect Simulation. Department of Mathematical Sciences, Aalborg University.

Berthelsen, K.K., Møller, J., 2002. A primer on perfect simulation for spatial point processes. Fifth Brazilian School in Probability (Ubatuba, 2001). Bull. Braz. Math. Soc. (N.S.) 33 (3), 351–367.

Besag, J.E., Milne, R., Zachary, S., 1982. Point process limits of lattice processes. J. Appl. Probab. 19 (1), 210–216.

Daley, D.J., Vere-Jones, D., 1988. An Introduction to the Theory of Point Processes. Springer, New York.

Ferrari, P.A., Fernández, R., Garcia, N.L., 2002. Perfect simulation for interacting point processes, loss networks and Ising models. Stochastic Process. Appl. 102 (1), 63–88.

Fill, J.A., 1998. An interruptible algorithm for perfect sampling via Markov chains. Ann. Appl. Probab. 8 (1), 131–162.

Fill, J.A., Huber, M., 2000. The randomness recycler: a new technique for perfect sampling. In: Proceedings of the 41st Annual Symposium on Foundations of Computer Science.IEEE Computer Society Press, Los Alamitos, pp. 503–511.

Fill, J.A., Machida, M., Murdoch, D.J., Rosenthal, J.S., 2000. Extension of Fill's perfect rejection sampling algorithm to general chains. Random Structures Algorithms 17 (3–4), 290–316.

Garcia, N.L., 2000. Perfect simulation of spatial processes. Resenhas 4 (3), 283–325.

Geyer, C., 1999. Likelihood inference for spatial point processes. In: Barndorff-Nielsen, O., Kendall, W.S., Lieshout, M.N.M. van (Eds.), Stochastic Geometry, Likelihood and Computation. Chapman and Hall, CRC, Boca Raton, pp. 79–140.

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

Goulard, M., Sǎrkkǎ, A., Grabarnik, P., 1996. Parameter estimation for marked Gibbs point processes through the maximum pseudo-likelihood method. Scand. J. Statist. 23 (3), 365–379.

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

Häggström, O., Nelander, K., 1998. Exact sampling from anti-monotone systems. Statist. Neerlandica 52 (3), 360–380.

Häggström, O., Nelander, K., 1999. On exact simulation of Markov random fields using coupling from the past. Scand. J. Statist. 26 (3), 395–411.

Häggström, O., Lieshout, M.N.M. van, Møller, J., 1999. Characterization results and Markov chain Monte Carlo algorithms including exact simulation for some spatial point processes. Bernoulli 5 (4), 641–658.

Kendall, W.S., 1998. Perfect simulation for the area-interaction point process. In: Accardi, L., Heyde, C. (Eds.), Probability Towards 2000. Springer, New York, pp. 218–234.

Kendall, W.S., Møller, J., 2000. Perfect simulation using dominating processes on ordered spaces, with application to locally stable point processes. Adv. Appl. Probab. 32 (3), 844–865.

Kendall, W.S., Montana, G., 2002. Small sets and Markov transition densities. Stochastic Process. Appl. 99 (2), 177–194.

Lieshout, M.N.M. van, 2000. Markov Point Processes and their Applications. Imperial College Press, World Scientific Publishing, London, Singapore.

Lieshout, M.N.M. van, Baddeley, A.J., 2002. Extrapolating and interpolating spatial patterns. In: Lawson, A.B., Denison, D.G.T. (Eds.), Spatial Cluster Modelling. Chapman and Hall, CRC, Boca Raton, pp. 61–86.

Lieshout, M.N.M. van, Stoica, R.S., 2003. The Candy model: properties and inference. Statist. Neerlandica 57 (2), 177–206.

Meyn, S.P., Tweedie, R.L., 1993. Markov Chains and Stochastic Stability. Springer, London.

Mira, A., Møller, J., Roberts, G.O., 2001. Perfect slice samplers. J. Roy. Statist. Soc. Ser. B Statist. Methodol. 63 (3), 593–606.

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

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

Murdoch, D.J., Green, P.J., 1998. Exact sampling from a continuous state space. Scand. J. Statist. 25 (3), 483–502.

Ogata, Y., Tanemura, M., 1984. Maximum likelihood analysis of spatial point patterns. J. Roy. Statist. Soc. Ser. B 46 (3), 496–518.

Preston, C.J., 1977. Spatial birth-and-death processes. Bull. Internat. Statist. Inst. 46, 371–391.

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

Ripley, B.D., 1977. Modelling spatial patterns. With discussion. J. Roy. Statist. Soc. Ser. B 39 (2), 172–212.

Ripley, B.D., Kelly, F.P., 1977. Markov point processes. J. London Math. Soc. 15 (2), 188–192.

Steenbeek, A.G., Lieshout, M.N.M. van, Stoica, R.S., 2002–2003. With contributions from P. Gregori and K.K., Berthelsen. MPPBLIB, a C++ library for marked point processes. CWI.

Steinsland, I., 2006. Parallel exact sampling and evaluation of Gaussian Markov random fields. Comput. Statist. Data Anal., Available online 13 February 2006.

Stoica, R., Descombes, X., Zerubia, J., 2004. A Gibbs point process for road extraction from remotely sensed images. Internat. J. Comput. Vision 57, 121–136.

Thőnnes, E., 1999. Perfect simulation of some point processes for the impatient user. Adv. Appl. Probab. 31 (1), 69–87.

Widom, B., Rowlinson, J.S., 1970. A new model for the study of liquid–vapor phase transition. J. Chem. Phys. 52, 1670–1684.