

Characterization results and Markov chain Monte Carlo algorithms including exact simulation for some spatial point processes

OLLE HÄGGSTRÖM,¹ MARIE-COLETTE N.M. VAN LIESHOUT² and JESPER MØLLER³

¹*Department of Mathematics, Chalmers University of Technology, S-412 96 Göteborg, Sweden. E-mail: olleh@chalmers.math.se*

²*Centre for Mathematics and Computer Science, POBox 94079, 1090 GB Amsterdam, The Netherlands. E-mail: Marie-Colette.van.Lieshout@cw.nl*

³*Department of Mathematical Sciences, Aalborg University, Fredrik Bajers Vej 7E, DK-9220 Aalborg Ø, Denmark. E-mail: jm@math.auc.dk*

The area-interaction process and the continuum random-cluster model are characterized in terms of certain functional forms of their respective conditional intensities. In certain cases, these two point process models can be derived from a bivariate point process model which in many respects is simpler to analyse and simulate. Using this correspondence we devise a two-component Gibbs sampler, which can be used for fast and exact simulation by extending the recent ideas of Propp and Wilson. We further introduce a Swendsen–Wang type algorithm. The relevance of the results within spatial statistics as well as statistical physics is discussed.

Keywords: area-interaction process; continuum random-cluster model; exact simulation; Gibbs sampling; Markov chain Monte Carlo; nearest-neighbour Markov point process; Papangelou conditional intensity; penetrable sphere model; phase transition; spatial point processes; Swendsen–Wang algorithm; Widom–Rowlinson mixture model

1. Introduction

One of the most popular classes of models for spatial point patterns exhibiting interactions between the points is that of Markov point processes (Ripley and Kelly 1977); the similar concept of a Gibbs point process is heavily used in statistical physics (Ruelle 1969). Initially, attention was focused on the special case of pairwise interaction models, defined in terms of a simple interaction function on pairs of points that are less than a certain distance apart. These models are appealing since they are easy to interpret – they usually form an exponential family whose sufficient statistics are related to natural interaction potentials – and they are by their nature particularly amenable to Markov chain Monte Carlo (MCMC) simulation and other iterative statistical techniques (see, for example Besag and Green 1993; Geyer 1998; Geyer and Møller 1994; or Møller 1999; and the references therein). However,

although they are quite flexible models for regular patterns, pairwise interaction processes do not seem to be able to produce clustered patterns in sufficient variety.

The area-interaction model of Baddeley and van Lieshout (1995), in contrast, is a Markov point process that can be used both for clustered and repulsive models by varying its interaction parameter. More recent generalizations are suggested by Kendall *et al.* (1999) and van Lieshout and Molchanov (1998), while lattice models of a somewhat similar nature are studied in Møller and Waagepetersen (1998) and in Häggström (1998). Repulsive area interaction is particularly useful in object recognition to avoid multiple response caused by the occlusion of many overlapping objects (Baddeley and van Lieshout 1992). Indeed, in an iterative recognition algorithm, if the introduction of a new object to a scene is considered, a natural penalty term is based on the amount of overlap between the new object and the existing ones. Our characterization theorem in Section 2 implies that under some mild conditions, the repulsive area-interaction model is the only candidate. The area-interaction model for clustered patterns has been applied to liquid–vapour equilibrium in chemistry by Widom and Rowlinson (1970), and may also be useful as a ‘selfish herd’ model in biology (Hamilton 1971).

In this paper we study the area-interaction process (Baddeley and van Lieshout 1995) and two related processes, namely the continuum random-cluster model (Chayes *et al.* 1995; Møller 1994; 1999) and the penetrable spheres mixture model (Widom and Rowlinson 1970; Hammersley *et al.* 1975; Rowlinson 1980; 1990). The latter model is highly relevant in statistical physics, in providing the best-known example of a phase transition in a continuous setting.

The plan of this paper is as follows. After providing some background material on spatial point processes, Section 2 gives characterization results for the models mentioned above, providing additional motivation for their use. In the rest of the paper, we devise algorithms for MCMC simulation of the models, including an exact simulation algorithm based on Gibbs sampling. Readers interested only in these more practical aspects may start with Section 3 and then turn to Section 4. If one is interested only in Swendsen–Wang type algorithms, it suffices to consider Sections 3 and 5.

More specifically, in Section 3 we introduce a two-component Gibbs sampler for the mixture model. A problem of this sampler, shared by most MCMC simulation techniques, is that the available rigorous bounds for rates of convergence are not good enough to be useful in practice, so that one is forced to use the (not particularly satisfactory) method of running the chain for a reasonably long time and just hoping that it is close to the stationary distribution. Therefore, it is highly remarkable that Propp and Wilson (1996) recently found a simulation technique which gives a sample from the Ising or Potts model that has *exactly* the right distribution, and which works in practice even for fairly large systems. In Section 4 we demonstrate how the two-component Gibbs sampler can be combined with Propp and Wilson’s ideas in order to obtain exact samples from the mixture model (and hence also from the attractive area-interaction model and some particular cases of the continuum random-cluster model). Recently, and independently of our work, Kendall (1998) has also demonstrated a way to apply the Propp–Wilson techniques in a more general point process setting. While Kendall uses a coupling construction of spatial birth-and-death processes, our approach seems much simpler to present and implement. The possibility of doing exact simulation for general point process models by replacing spatial

birth-and-death processes with Markov chains generated by the Metropolis–Hastings algorithm is being studied by Kendall and Møller (1999).

Section 4 also contains some empirical findings. We show that exact samples from the mixture model can be obtained in only a few steps as long as the rate of the underlying Poisson processes is small or moderate. An important motivation for doing simulation is to obtain a better understanding of the phase transition phenomenon. It is well known that the mixture model exhibits a phase transition for sufficiently large values of the model parameter, but not for small values (see, for example Chayes *et al.* 1995). The open problem is to decide whether the onset of the phase transition is marked by a unique critical point and, if so, to find its (approximate) value. Here we will initiate a study in this direction based on exact simulation.

Finally, in Section 5, we introduce an alternative MCMC algorithm for the above point process models. This algorithm is similar in spirit to the Swendsen and Wang (1987) algorithm for simulation of the Ising model and was simultaneously and independently discovered by Chayes and Machta (1998).

2. Characterization results

In this section, we establish some characterization results for the area-interaction process and the continuum random-cluster process; the results are of interest in their own right, but they also motivate the interest in the closely connected penetrable spheres mixture model in Sections 3–5. Before stating the characterization results, we need to introduce some background material and notation.

We consider spatial point processes X on a bounded Borel set $A \subset \mathbb{R}^d$, defined by their density $f(\cdot)$ with respect to a unit-rate Poisson process. Let Ω_A denote the space of all finite point configurations in A without multiple points, that is,

$$\Omega_A = \{\mathbf{x} \subseteq A \mid n(\mathbf{x}) < \infty\}.$$

Here $n(\mathbf{x})$ denotes the number of points n in the configuration $\mathbf{x} = \{x_1, \dots, x_n\}$.

The Papangelou conditional intensity (Daley and Vere-Jones 1988; Kallenberg 1984) of X is given by

$$\lambda^*(\mathbf{x}, u) = \begin{cases} \frac{f(\mathbf{x} \cup \{u\})}{f(\mathbf{x})}, & \text{if } f(\mathbf{x}) > 0, \\ 0, & \text{otherwise,} \end{cases} \tag{2.1}$$

for $\mathbf{x} \in \Omega_A$ and $u \in A \setminus \mathbf{x}$. In fact, there is a one-to-one correspondence between $f(\cdot)$ and $\lambda^*(\cdot, \cdot)$ if the probability density is hereditary, that is, whenever $f(\mathbf{y}) > 0$ implies $f(\mathbf{x}) > 0$ for all $\mathbf{x} \subseteq \mathbf{y}$.

The conditional intensity can be used to define a Markov property for point processes. Given a symmetric neighbourhood relation \sim on A , X is a Markov process with respect to \sim in the Ripley and Kelly (1977) sense if its density is hereditary and $\lambda^*(\mathbf{x}, u)$ depends only on u and its neighbours $\{x_i \in \mathbf{x} : u \sim x_i\}$ in \mathbf{x} . Heuristically, $\lambda^*(\mathbf{x}, u) du$ can readily

be interpreted as the conditional probability of having a point in the infinitesimal region du centred at u given the rest of the pattern is \mathbf{x} .

For modelling purposes it may thus be useful to establish characterization results in terms of $\lambda^*(\cdot, \cdot)$. Strauss (1975) and Kelly and Ripley (1976) considered the Strauss process

$$f(\mathbf{x}) = \alpha \beta^{n(\mathbf{x})} \gamma^{s(\mathbf{x})} \tag{2.2}$$

where α is the normalizing constant, $\beta > 0$ and $\gamma \in [0, 1]$ are model parameters, and $s(\mathbf{x})$ denotes the number of pairs of points $\xi, \eta \in \mathbf{x}$ such that $\xi \sim \eta$. Any symmetric relation \sim (defined on an arbitrary space) may be considered here, but usually $\xi \sim \eta$ if and only if $\|\xi - \eta\| \leq R$ (for some $R > 0$ fixed in advance). Then, if $\gamma = 0$, (2.2) defines a hard core process where no points are allowed to be within distance R of each other. Assuming that A contains pairwise distinct points $\xi, \eta, \zeta_1, \zeta_2, \dots$ satisfying $\xi \sim \eta$ and $\xi \sim \zeta_1, \eta \not\sim \zeta_1, \xi \sim \zeta_2, \eta \not\sim \zeta_2, \dots$, the proof of Theorem 1 in Kelly and Ripley (1976) can be modified to the present situation (where we have excluded the case of multiple points), whereby the Strauss process is seen to be uniquely characterized by two properties: its density is hereditary, and the conditional intensity is of the form

$$\lambda^*(\mathbf{x}, u) = g(n(\mathbf{x} \cap B_u)) \tag{2.3}$$

for all $\mathbf{x} \in \Omega_A$ and $u \in A \setminus \mathbf{x}$, where $g : \mathbb{N}_0 \rightarrow [0, \infty)$ and B_u denotes the closed ball centred at u with radius R .

Strauss (1975) suggested (2.2) with $\gamma > 1$ as a model for the clustering of Californian redwood seedlings around older stumps but, as pointed out by Kelly and Ripley (1976), (2.2) is only well defined for $0 \leq \gamma \leq 1$. Note that the model exhibits interactions between pairs of points only. Pairwise interaction models appear to be a useful and flexible class of models for regular patterns, but probably not so for clustered patterns (Diggle *et al.* 1994; Gates and Westcott 1986; and Møller 1999).

A more promising way of modelling attraction between points in a spatial pattern is to allow interaction terms of higher order as in Baddeley and van Lieshout (1995), Geyer (1999) and Møller (1999), or to generalize the Markov property to depend on the configuration as in the definition of *nearest-neighbour Markov point processes* (cf. Baddeley and Møller 1989). More precisely, we shall consider nearest-neighbour Markov point processes defined with respect to the connected component relation $\sim_{\mathbf{x}}$ on $\mathbf{x} \in \Omega_A$ given by

$$\xi \sim_{\mathbf{x}} \eta \Leftrightarrow \xi = x_1 \sim x_2 \sim \dots \sim x_m = \eta \tag{2.4}$$

for some subconfiguration $\{x_1, \dots, x_m\} \subseteq \mathbf{x}$ and where \sim is a given symmetric relation on A . Then, as shown in Baddeley *et al.* (1996), the general definition of a nearest-neighbour point process is equivalent to having a density of the form

$$f(\mathbf{x}) = \alpha \prod_{\mathbf{y} \in C(\mathbf{x})} \Phi(\mathbf{y}), \tag{2.5}$$

where $\alpha > 0$ is a normalizing constant, $C(\mathbf{x})$ is the set of (maximal) connected components defined by \mathbf{x} and $\Phi(\cdot) \geq 0$ satisfies certain regularity conditions (if $f(\cdot) > 0$ then strict positivity of $\Phi(\cdot)$ is the only condition).

The Ripley and Kelly (1977) Markov point processes as well as certain Poisson cluster processes (Daley and Vere-Jones 1988) are special cases of (2.5); cf. Baddeley *et al.* (1996). Lattice processes with a density similar in form to (2.5) have recently been studied in Møller and Waagepetersen (1998), where among other things characterization results similar to those below are discussed.

For specificity, we henceforth let $u \sim v$ in (2.4) be the usual relation: $u \sim v$ if and only if $\|u - v\| \leq R$, where $R > 0$ is given.

In the rest of this section we consider two models of the form (2.5) with respectively $\Phi(\mathbf{y}) = \beta^{n(\mathbf{y})}/\gamma$ and $\Phi(\mathbf{y}) = \beta^{n(\mathbf{y})}\gamma^{-|U_{\mathbf{y}}|}$, where $U_{\mathbf{y}} = \cup_i B_{y_i}$ and $|\cdot|$ denotes Lebesgue measure. In both cases, the densities are well defined for all $\beta, \gamma > 0$ and the models exhibit regularity for $0 < \gamma < 1$ and clustering for $\gamma > 1$. The hard core process may be considered as a limiting case of both models.

First, consider the continuum random-cluster model which has $\Phi(\mathbf{y}) = \beta^{n(\mathbf{y})}/\gamma$ and density

$$f(\mathbf{x}) = \alpha \beta^{n(\mathbf{x})} \gamma^{-c(\mathbf{x})}, \tag{2.6}$$

where $c(\mathbf{x})$ denotes the number of connected components in $U_{\mathbf{x}}$. This model seems to have been rediscovered many times, for example in Chayes *et al.* (1995) and Møller (1994); the earliest appearance in the literature we are aware of is in Klein (1982). We use the name ‘continuum random-cluster model’ because of the strong analogy with the random-cluster representation of the Ising–Potts models introduced by Fortuin and Kasteleyn (1972); see Borgs and Chayes (1996) and Grimmet (1995) for recent reviews. We have the following characterization result.

Theorem 1. *A density $f(\cdot)$ is a continuum random-cluster process if and only if*

$$\lambda^*(\mathbf{x}, u) = g(c(\mathbf{x} \cup \{u\}) - c(\mathbf{x})) \tag{2.7}$$

for all $\mathbf{x} \in \Omega_A$, $u \in A \setminus \mathbf{x}$ and a function $g : \mathbb{Z} \rightarrow (0, \infty)$.

Note that

$$c(\mathbf{x}, u) = 1 + c(\mathbf{x}) - c(\mathbf{x} \cup \{u\})$$

is the number of ‘clusters’ $U_{\mathbf{y}}$, $\mathbf{y} \in C(\mathbf{x})$, generated by \mathbf{x} which are intersected by the disc associated with u , so (2.7) is equivalent to $f(\mathbf{x} \cup \{u\})/f(\mathbf{x}) = g(c(\mathbf{x}, u))$.

Note also that the result still holds if the fixed range relation \sim is replaced by any other symmetric relation defined on a finite measure space; see Møller (1999). For instance, one may consider configurations of path-connected sets such as discs, where two discs are related if and only if they overlap each other.

Moreover, the positivity condition on $\lambda^*(\cdot, \cdot)$ is necessary: if $g(n) = \beta \mathbf{1}\{n = 1\}$ then (2.7) gives a hard core process (all $\|x_i - x_j\| \geq 2R$).

Proof. Clearly, if $f(\cdot)$ belongs to the family of continuum random-cluster processes, its conditional intensity is of the form specified in (2.7). On the other hand, assuming (2.7) holds, we proceed to prove that $f(\cdot)$ is a continuum random-cluster process by induction with respect to $n(\mathbf{x})$, the number of points. Setting

$$\beta = g(0) \quad \text{and} \quad \gamma = \frac{g(0)}{g(1)}, \quad (2.8)$$

then (2.6) holds for $n(\mathbf{x}) \leq 1$. Assuming the statement holds for configurations with up to $n(\mathbf{x}) = n$ points and writing $k = c(\mathbf{x} \cup \{u\}) - c(\mathbf{x})$,

$$f(\mathbf{x} \cup \{u\}) = f(\mathbf{x})\lambda^*(\mathbf{x}, u) = \alpha\beta^n\gamma^{-c(\mathbf{x})}g(k).$$

We can always arrange the points in the configuration $\mathbf{y} = \mathbf{x} \cup \{u\}$ so that either $k = 0$ or $k = 1$. If $k = 0$, then

$$f(\mathbf{y}) = \alpha\beta^n\gamma^{-c(\mathbf{x})}\beta = \alpha\beta^{n+1}\gamma^{-c(\mathbf{y})}.$$

If $k = 1$, then

$$f(\mathbf{y}) = \alpha\beta^n\gamma^{-c(\mathbf{x})}\beta/\gamma = \alpha\beta^{n+1}\gamma^{-c(\mathbf{y})}.$$

□

Next turn to the model specified by $\Phi(\mathbf{y}) = \beta^{n(\mathbf{y})}\gamma^{-|U_{\mathbf{y}}|}$ or, equivalently, by density

$$f(\mathbf{x}) = \alpha\beta^{n(\mathbf{x})}\gamma^{-|U_{\mathbf{x}}|} \quad (2.9)$$

This is the area-interaction model studied by Baddeley and van Lieshout (1995), a generalization of the penetrable sphere model – introduced by Widom and Rowlinson (1970) in statistical physics – where $\gamma > 1$. Contrary to the continuum random-cluster model, (2.9) is Markov in the Ripley and Kelly (1977) sense but it has interactions of arbitrarily high order; see Baddeley and van Lieshout (1995). In the one-dimensional case, it can also be presented as a *pairwise* interaction sequential neighbours Markov process (Baddeley and Møller 1989), since

$$f(\mathbf{x}) = \alpha\gamma^{-2R}\beta^n \exp[-(\log \gamma) \sum_{i=1}^{n-1} \min(x_{i+1} - x_i, 2R)]$$

for $\mathbf{x} = \{x_1, \dots, x_n\}$, with $x_1 < x_2 < \dots < x_n$.

Theorem 2. *Given a bounded Borel set $A \subset \mathbb{R}^2$ containing an open ball of radius $3R$, a density $f : \Omega_A \rightarrow (0, \infty)$ is an area-interaction density if and only if*

$$\lambda^*(\mathbf{x}, u) = g(|B_u \cap U_{\mathbf{x}}|)$$

for all $\mathbf{x} \in \Omega_A$, $u \in A \setminus \mathbf{x}$ and a left-continuous function $g : [0, \pi R^2] \rightarrow (0, \infty)$.

Note that although the characterization Theorem 2 is stated for the planar case only, generalizations to higher dimensions are straightforward. Clearly, we need the positivity condition $g(\cdot) > 0$, since the conditional intensity of a hard core process

$$f(\mathbf{x}) = \alpha\beta^{n(\mathbf{x})}\mathbf{1}\{\|x_i - x_j\| \geq 2R \text{ for all } i \neq j\}$$

can be written as

$$\lambda^*(\mathbf{x}, u) = \beta\mathbf{1}\{\rho(u, \mathbf{x}) \geq 2R\} = \beta\mathbf{1}\{|B_u \cap U_{\mathbf{x}}| = 0\}.$$

The hard core process does not fall within the class of area-interaction processes, although it can be seen as a limiting case.

As noted in the introduction, Theorem 2 has implications for applications in, for example, computer vision and biology where a spatial model is required whose conditional intensity is a function g of the volume of overlap. If the function g is positive and left-continuous, then an area-interaction model is the only possible choice.

In order to prove Theorem 2 we need the following lemma.

Lemma 1. *If $g : [0, \pi R^2] \rightarrow (0, \infty)$ is left-continuous and $g(s+t)g(0)/(g(s)g(t)) = 1$ for all $s, t \in [0, \pi R^2]$ such that $s+t \in [0, \pi R^2]$, then $g(s) = g(0)\gamma^s$ for some $\gamma > 0$ and all $s \in [0, \pi R^2]$.*

Proof. Without loss of generality, take $R = 1$ and extend the function $g(\cdot)$ onto the whole of $(0, \infty)$ as follows:

$$g(k\pi + s) = \left(\frac{g(\pi)}{g(0)}\right)^k g(s), \quad s \in (0, \pi], k \in \mathbb{N}.$$

Then $g(\cdot)$ is left-continuous on $(0, \infty)$. Moreover, for $s, t \in [0, \pi]$ with $s+t > \pi$,

$$\frac{g(0)g(s+t)}{g(s)g(t)} = \frac{g(\pi)g(s+t-\pi)}{g(s)g(t)}.$$

Now choose $c_1, c_2 > 0$ such that $s - c_1\pi \geq 0, t - c_2\pi \geq 0$, and $c_1 + c_2 = 1$. Then

$$\begin{aligned} \frac{g(0)g(s+t)}{g(s)g(t)} &= \frac{g(\pi)g(s-c_1\pi)g(t-c_2\pi)}{g(s)g(t)g(0)} \\ &= \frac{g(\pi)g(s-c_1\pi)g(t-c_2\pi)g(0)g(0)}{g(0)g(c_1\pi)g(s-c_1\pi)g(c_2\pi)g(t-c_2\pi)} \\ &= \frac{g(\pi)g(0)}{g(c_1\pi)g(c_2\pi)}. \end{aligned}$$

The latter expression equals 1 by assumption, since $c_1 + c_2 = 1$. Hence, $g(t) = g(0)e^{-\lambda t}$ where $\lambda = (-1/\pi)\log(g(\pi)/g(0))$. In particular, for all $s \in [0, \pi]$,

$$g(s) = g(0)\gamma^s, \quad \gamma = \left(\frac{g(\pi)}{g(0)}\right)^{1/\pi} > 0. \quad \square$$

Proof of Theorem 2. If $f(\cdot)$ is an area-interaction process,

$$\frac{f(\mathbf{x} \cup \{u\})}{f(\mathbf{x})} = \beta\gamma^{-|B_u|+|B_u \cap U_{\mathbf{x}}|} = g(|B_u \cap U_{\mathbf{x}}|)$$

for the continuous function $(g(t) = \beta\gamma^{-\pi R^2}\gamma^t$.

To prove the reverse statement, take $s, t \in [0, \pi R^2]$ such that $s+t \in [0, \pi R^2]$. We will

show that $g(s + t)g(0) = g(s)g(t)$, so assume without loss of generality that $s \neq 0$ and $t \neq 0$.

Choose $\varepsilon > 0$. Since A contains a ball B of radius $3R$, there exist u (the centre of B) and $v \in A$, $\mathbf{y} \in \Omega_A$, such that $|B_u \cap B_v| = s$ and $|D| = t' \in [t - \varepsilon, t]$, where $D = B_v \cap U_{\mathbf{y}}$ and $\|y_i - u\| = 2R$ for all $y_i \in \mathbf{y}$. Here we use the Heine–Borel theorem to ensure that the configuration \mathbf{y} is finite and $U_{\mathbf{y}}$ covers an area t of B_v up to the given precision ε without covering any area of B_u . Now

$$\begin{aligned} g(s)g(t') &= \frac{f(\{u, v\} \cup \mathbf{y})}{f(\{v\} \cup \mathbf{y})} \frac{f(\{v\} \cup \mathbf{y})}{f(\mathbf{y})} \\ &= \frac{f(\{u, v\} \cup \mathbf{y})}{f(\{u\} \cup \mathbf{y})} \frac{f(\{u\} \cup \mathbf{y})}{f(\mathbf{y})} = g(s + t')g(0). \end{aligned}$$

As $\varepsilon \downarrow 0$, $t' \rightarrow t$. Since $t' \leq t$, by left continuity $g(s + t)g(0) = g(s)g(t)$.

Consequently, by Lemma 1, $g(s) = g(0)\gamma^s$, for some $\gamma > 0$. Now, letting $\beta = g(0)\gamma^{\pi R^2} = g(\pi R^2)$, we obtain $g(s) = \beta\gamma^{s - \pi R^2}$ and hence

$$f(\{u\}) = f(\emptyset)g(0) = \alpha\beta\gamma^{-\pi R^2} = \alpha\beta\gamma^{-|U_{\{u\}}|}.$$

By induction with respect to the number of points $n(\mathbf{x}) = n$,

$$\begin{aligned} f(\mathbf{x} \cup \{u\}) &= g(|B_u \cap U_{\mathbf{x}}|)f(\mathbf{x}) \\ &= g(0)\gamma^{|B_u \cap U_{\mathbf{x}}|} \alpha\beta^n \gamma^{-|U_{\mathbf{x}}|} \\ &= \alpha\beta^{n+1} \gamma^{-\pi R^2 + |B_u \cap U_{\mathbf{x}}| - |U_{\mathbf{x}}|} \\ &= \alpha\beta^{n+1} \gamma^{-|U_{\mathbf{x} \cup \{u\}}|}. \end{aligned} \quad \square$$

3. Relationship to a mixture model and Gibbs sampling

The area-interaction model and a particular case of the continuum random-cluster model are both related to Widom and Rowlinson’s (1970) penetrable spheres mixture model as described below. So far this correspondence has mainly been used for studying phase transition behaviour; we return to this in Section 4. In the present section we shall investigate the relationships for the purpose of simulation: we propose a two-component Gibbs sampler which is also used in the exact simulation procedure discussed in Section 4.

The mixture model is a bivariate point process (X, Y) , where X and Y refer to two different types of point configurations. The joint density of (X, Y) at (\mathbf{x}, \mathbf{y}) with respect to the product measure of two independent unit-rate Poisson processes is

$$f_{\text{mix}}(\mathbf{x}, \mathbf{y}) = \alpha_{\text{mix}} \beta_1^{n(\mathbf{x})} \beta_2^{n(\mathbf{y})} \mathbf{1}\{d(\mathbf{x}, \mathbf{y}) > R\}, \tag{3.1}$$

where α_{mix} is the normalizing constant, $\beta_1, \beta_2 > 0$ are model parameters, and $d(\mathbf{x}, \mathbf{y})$ is the shortest distance between a point in \mathbf{x} and a point in \mathbf{y} . In other words, only points of different types interact and they are not allowed to be within a distance R of each other.

Hence the conditional distribution of X given Y is a homogeneous Poisson process on $A \setminus U_Y$ with intensity β_1 (see Widom and Rowlinson 1970).

Writing π for the distribution of a unit-rate Poisson process on A and integrating out the second component, the marginal density of X with respect to π is

$$\begin{aligned} f(\mathbf{x}) &= \int \alpha_{\text{mix}} \beta_1^{n(\mathbf{x})} \beta_2^{n(\mathbf{y})} \mathbf{1}\{d(\mathbf{x}, \mathbf{y}) > R\} d\pi(\mathbf{y}) \\ &= \int \alpha_{\text{mix}} \beta_1^{n(\mathbf{x})} \beta_2^{n(\mathbf{y})} \mathbf{1}\{\mathbf{y} \cap U_{\mathbf{x}} = \emptyset\} d\pi(\mathbf{y}) \\ &= \alpha_{\text{mix}} \beta_1^{n(\mathbf{x})} \exp[\beta_2 |A \setminus U_{\mathbf{x}}| - |A|]. \end{aligned}$$

Provided $|U_{\mathbf{x}}|$ is redefined as the Lebesgue measure on A , or alternatively, A is rectangular with periodic boundary conditions (making A into a torus) and balls B_u are redefined with respect to geodesic distance $\|\cdot\|$, we find that

$$f(\mathbf{x}) = \alpha_1 \beta_1^{n(\mathbf{x})} (e^{\beta_2})^{-|U_{\mathbf{x}}|},$$

where $\alpha_1 = \alpha_{\text{mix}} e^{-(1-\beta_2)|A|}$. Hence X is a penetrable sphere or attractive area-interaction model with parameters $(\beta_1, e^{\beta_2}) \in (0, \infty) \times (1, \infty)$. Similarly, Y is an area-interaction model with parameters $(\beta_2, e^{\beta_1}) \in (0, \infty) \times (1, \infty)$ and normalizing constant $\alpha_2 = \alpha_{\text{mix}} e^{-(1-\beta_1)|A|}$.

Turning attention to the superposition $Z = X \cup Y$, its density (with respect to π) can be written as

$$\begin{aligned} f_s(\mathbf{z}) &= e^{-|A|} \sum f_{\text{mix}}(\mathbf{z}_1, \mathbf{z}_2) \\ &= e^{-|A|} \alpha_{\text{mix}} \sum \beta_1^{n(\mathbf{z}_1)} \beta_2^{n(\mathbf{z}_2)} \mathbf{1}\{d(\mathbf{z}_1, \mathbf{z}_2) > R\}, \end{aligned}$$

where the sum is over all ordered partitions of \mathbf{z} into two groups \mathbf{z}_1 and \mathbf{z}_2 . In the symmetric case $\beta_1 = \beta_2 = \beta$ this reduces to

$$f_s(\mathbf{z}) = \alpha_{\text{mix}} e^{-|A|} \beta^{n(\mathbf{z})} N(\mathbf{z})$$

where $N(\mathbf{z})$ denotes the number of partitions such that

$$(\cup_{\eta \in \mathbf{z}_1} B(\eta, R/2)) \cap (\cup_{\zeta \in \mathbf{z}_2} B(\zeta, R/2)) = \emptyset.$$

Since $N(\mathbf{z}) = 2^{c(\mathbf{z})}$, the superposition density is a continuum random-cluster model with radii $R/2$ and parameters β and $\gamma = \frac{1}{2}$. This relation is exploited by Georgii and Häggström (1996) in order to understand the phase transition of the mixture model in percolation terms.

This construction can be straightforwardly extended to the case of a multitype point process $(X^{(1)}, \dots, X^{(k)})$ with density proportional to

$$\beta^{n(\mathbf{z})} \mathbf{1}\{d(\mathbf{x}^{(i)}, \mathbf{x}^{(j)}) > R, \forall i \neq j\} \tag{3.2}$$

for $\mathbf{z} = \mathbf{x}^{(1)} \cup \dots \cup \mathbf{x}^{(k)}$. The superposition is then a continuum random-cluster model but with $\gamma = 1/k$. However, for $k \geq 3$, $X^{(i)}$ is not an area-interaction process.

For simulating the area-interaction model X or the continuum random-cluster model Z , it is tempting to apply MCMC methods to the mixture model (X, Y) of (3.1), since this avoids calculation of the areas of overlapping balls in U_X . Since the conditional distribution of one type of points given the others is particularly easy to sample from, we opted for a Gibbs sampler approach. In the bivariate case this is given by alternating between the conditional distribution of $X|Y$ and $Y|X$: initialize with any \mathbf{y}_0 , for example from a Poisson process (rate β_2). Then the sweeps of the *two-component Gibbs sampler* are given by the following steps for each $k = 0, 1, \dots$:

1. $\mathbf{x}_{k+1} \sim \text{Poisson}(\beta_1)$ on $A \setminus U_{\mathbf{y}_k}$;
2. $\mathbf{y}_{k+1} \sim \text{Poisson}(\beta_2)$ on $A \setminus U_{\mathbf{x}_{k+1}}$.

The Poisson processes in steps 1 and 2 above are easily implemented by thinning a Poisson process in the whole window A (that is, just delete those points lying in $U_{\mathbf{y}_k}$ or $U_{\mathbf{x}_{k+1}}$).

In fact it is possible to specify a similar two-component Gibbs sampler for the inhibitory case of the area-interaction process. In the first step $X|Y = \mathbf{y}$ is $\text{Poisson}(\beta_1)$ restricted to the event $H_{\mathbf{x}, \mathbf{y}} = \{U_{\mathbf{x}} \supseteq \mathbf{y}\}$. Note that this may be too slow in practice, so it would be preferable to replace this step with a single point updating procedure. The other step, where $Y|X = \mathbf{x}$ is a $\text{Poisson}(\beta_2)$ process on $U_{\mathbf{x}}$, is easily performed by thinning. The equilibrium density becomes

$$\tilde{f}_{\text{mix}}(\mathbf{x}, \mathbf{y}) = \tilde{\alpha}_{\text{mix}} \beta_1^{n(\mathbf{x})} \beta_2^{n(\mathbf{y})} \mathbf{1}\{\mathbf{y} \subseteq U_{\mathbf{x}}\},$$

and arguments similar to those above give the marginal density for X as

$$\tilde{f}(\mathbf{x}) = \tilde{\alpha}_{\text{mix}} e^{-|A|} \beta_1^{n(\mathbf{x})} (e^{-\beta_2})^{-|U_{\mathbf{x}}|}, \quad (3.3)$$

again writing $|U_{\mathbf{x}}|$ for the Lebesgue measure of $U_{\mathbf{x}}$ either restricted to A or in the geodesic sense. Hence X (but not Y !) is an inhibitory area-interaction model.

4. Exact simulation

We now combine Propp and Wilson's (1996) ideas of exact simulation based on coupling Markov chains from the past with the two-component Gibbs sampler introduced in Section 3.

Our set-up differs from that in Propp and Wilson (1996) mainly in two respects. Firstly, the state space in Propp and Wilson (1996) is finite, while ours is infinite. Secondly, our state space does not have any maximal (or minimal) element with respect to the partial order introduced below. The first difference turns out to be inconsequential, but the second requires special treatment.

Introduce a partial order \preceq on the space $\Omega_A^2 = \Omega_A \times \Omega_A$ of mixed configurations, given by $(\mathbf{x}, \mathbf{y}) \preceq (\mathbf{x}', \mathbf{y}')$ (or $(\mathbf{x}', \mathbf{y}') \succeq (\mathbf{x}, \mathbf{y})$) if $\mathbf{x} \subseteq \mathbf{x}'$ and $\mathbf{y} \supseteq \mathbf{y}'$. In order to be able to adapt the Propp–Wilson ideas we verify first that the two-component Gibbs sampler respects the partial order \preceq . Let \preceq_d denote stochastic domination with respect to \preceq ; that is, two Ω_A^2 -valued random elements (X, Y) and (X', Y') satisfy $(X, Y) \preceq_d (X', Y')$ if there exists a

coupling of (X, Y) and (X', Y') such that $(X, Y) \preceq (X', Y')$ almost surely (for a general discussion of coupling and stochastic domination, see Lindvall 1992).

Lemma 2. Fix $(\mathbf{x}, \mathbf{y}), (\mathbf{x}', \mathbf{y}') \in \Omega_A^2$ such that $(\mathbf{x}, \mathbf{y}) \preceq (\mathbf{x}', \mathbf{y}')$. Let, for $i = 0, 1, \dots$, (X_i, Y_i) be the Ω_A^2 -valued random element obtained by starting with $(X_0, Y_0) = (\mathbf{x}, \mathbf{y})$ and running i iterations of the two-component Gibbs sampler in Section 3, and define (X'_i, Y'_i) similarly. Then $(X_i, Y_i) \preceq_d (X'_i, Y'_i)$ for all i .

Proof. The case $i = 0$ is trivial. To prove the case $i = 1$ we consider the following coupling. Let $Z_{1,x}$ and $Z_{1,y}$ be two independent Poisson processes on A with rates β_1 and β_2 , respectively. Let $X_1 = Z_{1,x} \setminus U_{Y_0}$ and $X'_1 = Z_{1,x} \setminus U_{Y'_0}$, and then let $Y_1 = Z_{1,y} \setminus U_{X_1}$ and $Y'_1 = Z_{1,y} \setminus U_{X'_1}$. Clearly, this gives the right marginal distributions of (X_1, Y_1) and (X'_1, Y'_1) . Since $Y_0 \supseteq Y'_0$ we obtain $U_{Y_0} \supseteq U_{Y'_0}$ whence $X_1 \subseteq X'_1$. This in turn implies $U_{X_1} \subseteq U_{X'_1}$ so that $Y_1 \supseteq Y'_1$, and we have $(X_1, Y_1) \preceq (X'_1, Y'_1)$. The cases $i = 2, 3, \dots$ follow similarly by induction. \square

We call an element $(\mathbf{x}, \mathbf{y}) \in \Omega_A^2$ *quasimaximal* if $\mathbf{y} = \emptyset$ and $U_{\mathbf{x}} \supseteq A$. Similarly, (\mathbf{x}, \mathbf{y}) is called *quasiminimal* if $\mathbf{x} = \emptyset$ and $U_{\mathbf{y}} \supseteq A$.

Lemma 3. Fix $(\mathbf{x}, \mathbf{y}) \in \Omega_A^2$, and for $i = 0, 1, \dots$, let (X_i, Y_i) be the Ω_A^2 -valued random element obtained by taking $(X_0, Y_0) = (\mathbf{x}, \mathbf{y})$ and running i iterations of the two-component Gibbs sampler. Define (X'_i, Y'_i) similarly, with an arbitrary distribution of (X'_0, Y'_0) . If (\mathbf{x}, \mathbf{y}) is quasimaximal, then

$$(X_i, Y_i) \succeq_d (X'_i, Y'_i)$$

for all $i \geq 1$, while if (\mathbf{x}, \mathbf{y}) is quasiminimal, then

$$(X_i, Y_i) \preceq_d (X'_i, Y'_i)$$

for all $i \geq 1$.

Proof. We only need to prove the lemma for $i = 1$, as the general case follows using Lemma 2. However, the case $i = 1$ follows directly if we use the same coupling as in the proof of Lemma 2. \square

We are now ready to describe the algorithm for exact simulation. For $i = 0, -1, -2, \dots$, let $Z_{i,x}$ and $Z_{i,y}$ be independent Poisson processes on A with respective intensities β_1 and β_2 . Let k_1, k_2, \dots be a strictly increasing sequence of positive integers, and let (\mathbf{x}, \mathbf{y}) and $(\mathbf{x}', \mathbf{y}')$ be fixed elements of Ω_A^2 such that (\mathbf{x}, \mathbf{y}) is quasiminimal and $(\mathbf{x}', \mathbf{y}')$ is quasimaximal. Then, for $i = 1, 2, \dots$, we generate two coupled Markov chains in accordance to the two-component Gibbs sampler by setting

$$({}^i X_{-k_i}, {}^i Y_{-k_i}) = (\mathbf{x}, \mathbf{y}), \quad ({}^i X'_{-k_i}, {}^i Y'_{-k_i}) = (\mathbf{x}', \mathbf{y}'),$$

and for $j = 1, \dots, k_i$,

$$\begin{aligned} {}^iX_{j-k_i} &= Z_{j-k_i,x} \setminus U^i Y_{j-1-k_i}, & {}^iY_{j-k_i} &= Z_{j-k_i,y} \setminus U^i X_{j-k_i} \\ {}^iX'_{j-k_i} &= Z_{j-k_i,x} \setminus U^i Y'_{j-1-k_i}, & {}^iY'_{j-k_i} &= Z_{j-k_i,y} \setminus U^i X'_{j-k_i}. \end{aligned}$$

Clearly $({}^iX_{j-k_i}, {}^iY_{j-k_i}) \preceq ({}^iX'_{j-k_i}, {}^iY'_{j-k_i})$ for all $j = 0, \dots, k_i$. Letting $I = \inf\{i \in \mathbb{N} : ({}^iX_0, {}^iY_0) = ({}^iX'_0, {}^iY'_0)\}$, (taking $\inf \emptyset = \infty$), we show below that $I < \infty$ almost surely. We stop when $i = I$ and set $(X_0, Y_0) = ({}^I X_0, {}^I Y_0)$.

Theorem 3. *The above algorithm terminates almost surely, and the distribution of the obtained sample (X_0, Y_0) is given by (3.1).*

Proof. Note that, for any $m \geq 0$, if

$$Z_{-m-1,y} = \emptyset \tag{4.1}$$

then for any i such that $k_i \geq m$ we obtain $({}^iX_{-m}, {}^iY_{-m}) = ({}^iX'_{-m}, {}^iY'_{-m})$ and hence also $({}^iX_0, {}^iY_0) = ({}^iX'_0, {}^iY'_0)$. That $I < \infty$ almost surely now follows from the observation that with probability 1, (4.1) occurs for some m . Let M denote the smallest such m . Moreover, define Markov chains $({}^iX''_{j-k_i}, {}^iY''_{j-k_i})$, $j = 0, \dots, k_i$, $i = 1, 2, \dots$, in exactly the same way as the chains $({}^iX, {}^iY)$ and $({}^iX', {}^iY')$ except that $({}^iX''_{-k_i}, {}^iY''_{-k_i}) = (X, Y)$, where (X, Y) follows the stationary distribution (3.1). Then, for $k_i \geq M$, we have that ${}^iX_{-M} = {}^iX'_{-M} = {}^iX''_{-M} = Z_{-M,x}$. It follows then from Lemmas 2 and 3 and the coupling construction that $(X_0, Y_0) = ({}^iX_0, {}^iY_0) = ({}^iX'_0, {}^iY'_0) = ({}^iX''_0, {}^iY''_0)$ when $k_i \geq M$, so $(X_0, Y_0) = \lim_{i \rightarrow \infty} ({}^iX''_0, {}^iY''_0)$ almost surely. Hence the distribution of (X_0, Y_0) is given by the stationary distribution (3.1). □

Propp and Wilson give an argument for preferring the sequence $k_i = 2^i$, and we have used the same in our simulation studies.

Fixing $R > 0$ and the dimension $d \geq 2$, let us say $d = 2$, it is known that phase transition behaviour occurs in the penetrable sphere model in the symmetric case $\beta = \log \gamma$, whenever β is very large, and does not occur when β is very small. This is a consequence of the phase transition behaviour in the mixture model (3.1), which was demonstrated by Ruelle (1971) and later by Chayes *et al.* (1995) and Georgii and Häggström (1996) using percolation arguments analogous to the random-cluster derivation (see, for example, Borgs and Chayes 1996) of the phase transition occurring in Ising and Potts models. Similarly to the Ising and Potts models, phase transition means that infinite-volume limits fail to be unique. In particular, realizations tend, even for large systems, to be dominated by a single type of point (despite the symmetry of the model).

It is widely believed that the occurrence of phase transition is monotone in β , in which case there exists a β_c (depending on d and R) such that

$$\begin{cases} \beta > \beta_c & \Rightarrow \text{phase transition,} \\ \beta < \beta_c & \Rightarrow \text{no phase transition,} \end{cases} \tag{4.2}$$

but this is not known. In statistical physics, this is sometimes referred to as the ‘sharpness of the phase transition problem’. Of course, (4.2) is a statement about infinite-volume limits, but

the effect should already be visible for moderately large A . A measure of the amount of consensus is the fraction ϕ of pairs of points that are of the same type,

$$\phi = \frac{\binom{n_1}{2} + \binom{n_2}{2}}{\binom{n}{2}},$$

where n_i is the number of points of type i , and $n = n_1 + n_2$. Note that ϕ is undefined when $n = 0, 1$, in which cases we will arbitrarily set $\phi = 0$. For large systems, one expects the following behaviour; ϕ should be close to $\frac{1}{2}$ with high probability if $\beta < \beta_c$, it should start increasing rapidly as β approaches β_c , and tend to 1 as $\beta \rightarrow \infty$.

We have simulated independent samples from the symmetric mixture model using our exact simulation procedure for various values of β and R . We took $A = [0, 1]^2$ with the restricted Lebesgue measure. Here it is important to note that reducing the interaction radius is equivalent to enlarging the size of the window A . Hence any phase transition behaviour for given R that is visible in a large enough window A is also apparent in $[0, 1]^2$ for small enough R .

Figure 1 depicts the simulated mean of ϕ over 100 simulations against the canonical parameter $\theta = \log \beta$ of the mixture distribution (considered as an exponential family). It can be seen that the Monte Carlo approximation of $E_\theta \phi$ is indeed close to $1/2$ for small values of θ , while it increases to 1 as θ increases. The rapid increase becomes more apparent when the interaction range is smaller – or equivalently, the window size is bigger.

The phase transition behaviour can also be observed from plots of the fraction $p = \max(n_1, n_2)/(n_1 + n_2)$ of points of the most frequent type. Plots of the Monte Carlo mean of p (Figure 1) look very similar to the plots for ϕ , and with increasing θ , the histograms of p (which we omit) become bimodal, due to the fact that realizations tend to consist predominantly of one type. We have included the 0.05, 0.25, 0.75 and 0.95 Monte Carlo quantiles for p and ϕ to indicate the variability.

Examples of simulated realizations of point patterns for $\theta = \log(30), \log(80)$ at radius $R = 0.2$ and for $\theta = \log(100), \log(200)$ at radius $R = 0.1$ can be seen in Figure 2. This figure also shows that for larger θ , one of the components dominates.

Finally, in Figure 3 we plot Monte Carlo estimates based on 100 simulations of $E_\theta I$, the expected number of steps until coalescence. The plot demonstrates the feasibility of exact simulation: for small to moderate θ , convergence is reached in only a few steps. However, as θ approaches phase transition, the number of steps needed to obtain coalescence increases rapidly and it would not be computationally feasible to use many more than the 100 samples used in Figures 1 and 3 or to increase the window size (or reduce R) much further. One may hope that this will be practicable in the future, with faster computers and perhaps more ingenious implementations of the algorithm.

5. A Swendsen–Wang type algorithm

In this section we present an algorithm which is similar to one of Swendsen and Wang (1987). This algorithm has independently been discovered by Chayes and Machta (1998).

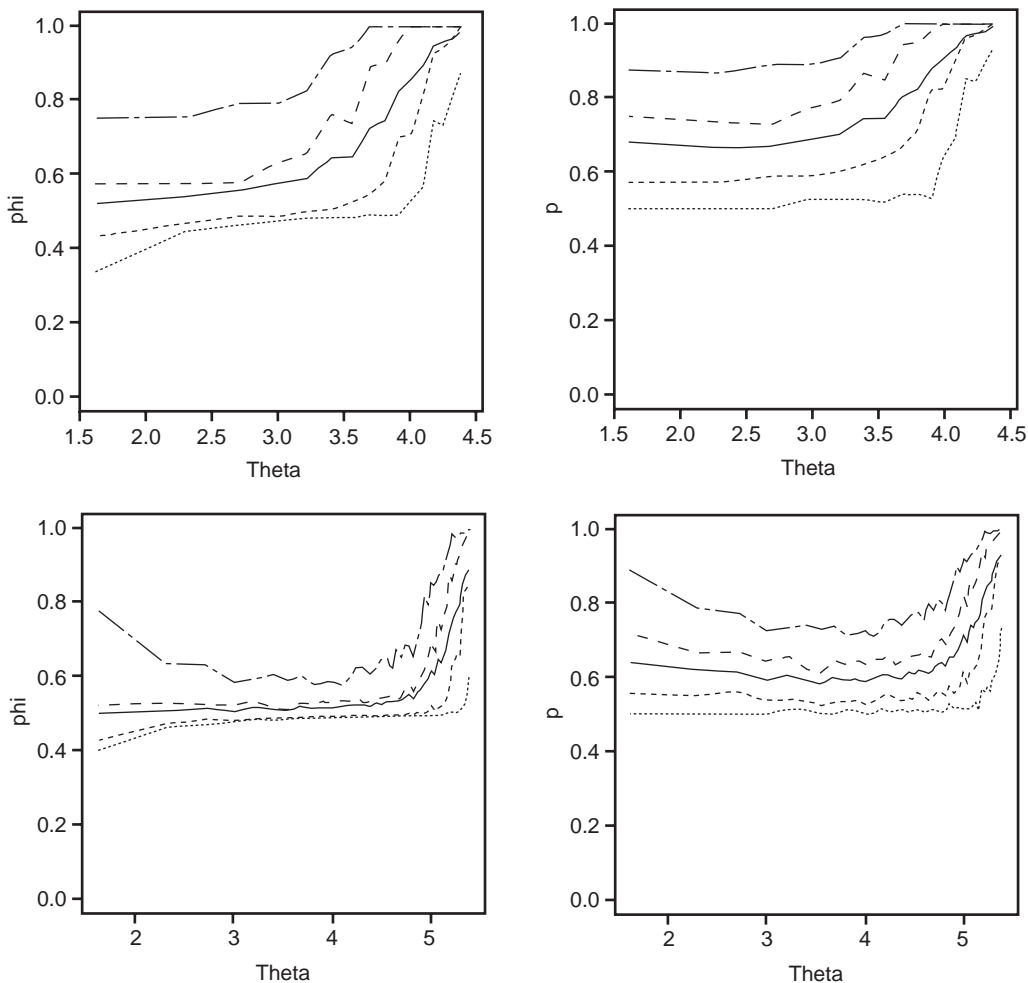


Figure 1. Monte Carlo estimates of the means of ϕ (left) and p (right) as a function of θ (solid line) at interaction radius $R = 0.2$ (top) and $R = 0.1$ (bottom). The dotted lines denote the 0.05, 0.25, 0.75 and 0.95 quantiles

The algorithm works for the important symmetric case $\beta_1 = \beta_2 = \beta$ of the mixture model (and hence also for the area-interaction model with $\gamma = e^\beta$ and for the continuum random-cluster process with $\gamma = \frac{1}{2}$). Initializing with any (\mathbf{x}, \mathbf{y}) , an iteration of the algorithm consists of the following steps:

1. Choose a new value of (\mathbf{x}, \mathbf{y}) according to its conditional distribution given $\mathbf{x} \cup \mathbf{y}$, that is, flip a fair coin independently for each connected component of $U_{\mathbf{x} \cup \mathbf{y}}$ to determine whether the points should be of the first or the second type.
2. Replace \mathbf{y} by a Poisson(β) process on $A \setminus U_{\mathbf{x}}$.

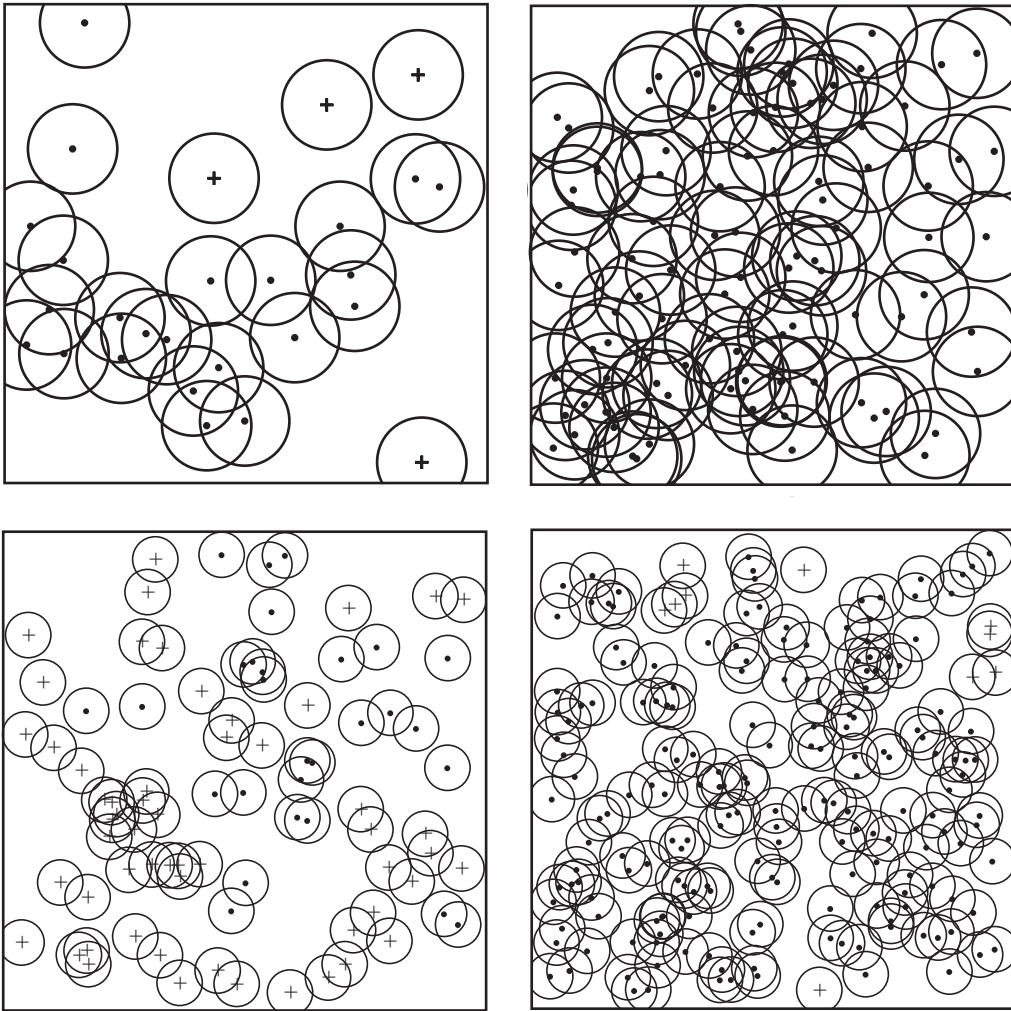


Figure 2. Realization of the mixture model with $\beta = 30$, $R = 0.2$ (top left), $\beta = 80$, $R = 0.2$ (top right), $\beta = 100$, $R = 0.1$ (bottom left) and $\beta = 200$, $R = 0.1$ (bottom right)

It is immediate that the mixture measure given by (3.1) is invariant under step 1 of this algorithm, and we have already seen in Section 3 that it is invariant under step 2. The algorithm can also be extended to the case of a multitype point process (3.2), and also to the $\gamma < 1$ case of the continuum random-cluster model, even when γ^{-1} is not an integer. In the latter case, the algorithm goes as follows. First paint the points of each connected component red with probability γ (independently for different connected components), and then replace all the red connected components by a $\text{Poisson}(\beta)$ process on the part of A not occupied by the remaining connected components.

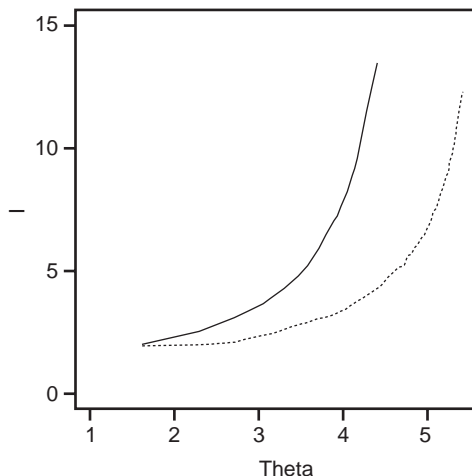


Figure 3. Monte Carlo estimates of $E_\theta I$ as a function of θ at interaction radius $R = 0.2$ (solid) and $R = 0.1$ (dashed)

We believe (although we have no rigorous justification) that this algorithm approaches stationarity much faster and mixes better than the two-component Gibbs sampler when β is large (that is, in the phase transition region of the parameter space). The reason should be that when the Gibbs sampler starts with a configuration with mostly points of the first (or second) type, then it will tend to stay in this state for an astronomical amount of time provided that β is large and A is large compared to a disc with radius R , while the Swendsen–Wang type algorithm will jump back and forth between the two states. One might suggest that the slowness of the Gibbs sampler could be solved by allowing \mathbf{x} and \mathbf{y} to change places occasionally, but this is presumably not the case, the reason being the following. Suppose β is large and $A = [-M, M]^2$, where $M \gg R$, and we start with a ‘dense’ crowd of points of type 1 in $[-M, M] \times [-M, 0]$ and a similar crowd of points of type 2 in $[-M, M] \times [0, M]$. This highly improbable type of configuration will remain for a long time using the modified Gibbs sampler (although the two types will sometimes interchange regions), while on the other hand the Swendsen–Wang type algorithm does not seem to exhibit such a phenomenon.

It would be very nice if the Swendsen–Wang type algorithm could be combined with the ideas of the previous section in order to obtain exact samples, but unfortunately it seems very difficult to find any useful monotonicity property of the algorithm. Propp and Wilson (1996) make a similar remark about the original Swendsen–Wang algorithm. They are still able to obtain exact samples in reasonable time also in the phase transition regime by simulating the Fortuin–Kasteleyn random-cluster model rather than the Ising–Potts models directly. The corresponding thing to do here would be first to simulate the continuum random-cluster model and then to update the components as in step 1 in our Swendsen–Wang type algorithm but, in the absence of any simple monotonicity property in the continuum random-cluster model, we do not at present see any suitable way of doing it.

Acknowledgements

This work was initiated in the autumn of 1995 when the second author visited the third author at the Department of Theoretical Statistics in Aarhus; in February 1996 the first author visited the third author at the Department of Mathematical Sciences in Aalborg. The hospitality and support of these departments is gratefully acknowledged. The research has been funded by the Danish Informatics Network in Agricultural Sciences, the Danish Natural Science Research Council, the Dutch Mathematical Research Institute, the Swedish Natural Science Research Council and grant SCI/180/94/103 (Applications of stochastic geometry in the analysis of spatial data) of the Nuffield Foundation.

References

- Baddeley, A.J. and Møller, J. (1989) Nearest-neighbour Markov point processes and random sets. *Internat. Statist. Rev.*, **57**, 89–121.
- Baddeley, A.J. and van Lieshout, M.N.M. (1992) ICM for object recognition. In Y. Dodge and J. Whittaker (eds), *Computational Statistics, Vol. 2*, pp. 271–286. Heidelberg: Physica-Verlag.
- Baddeley, A.J. and van Lieshout, M.N.M. (1995) Area-interaction point processes. *Ann. Inst. Statist. Math.*, **46**, 601–619.
- Baddeley, A.J., van Lieshout, M.N.M. and Møller, J. (1996) Markov properties of cluster processes. *Adv. Appl. Probab. (Stochastic Geom. Statist. Appl.)*, **28**, 346–355.
- Besag, J.E. and Green, P.J. (1993) Spatial statistics and Bayesian computation (with discussion). *J. Roy. Statist. Soc. Ser. B*, **55**, 25–37.
- Borgs, C. and Chayes, J.T. (1996) The covariance matrix of the Potts model: a random cluster analysis. *J. Statist. Phys.*, **82**, 1235–1297.
- Chayes, J.T., Chayes, L. and Kotecky, R. (1995) The analysis of the Widom–Rowlinson model by stochastic geometric methods. *Comm. Math. Phys.*, **172**, 551–569.
- Chayes, L. and Machta, J. (1998) Graphical representations and cluster algorithms part II. *Phys. A.*, **254**, 477–516.
- Daley, D.J. and Vere-Jones, D. (1988) *An Introduction to the Theory of Point Processes*. New York: Springer-Verlag.
- Diggle, P.J., Fiksel, T., Grabarnik, P., Ogata, Y., Stoyan, D. and Tanemura, M. (1994) On parameter estimation for pairwise interaction processes. *Internat. Statist. Rev.*, **62**, 99–117.
- Fortuin, C.M. and Kasteleyn, P.W. (1972) On the random-cluster model. I. Introduction and relation to other models. *Phys.*, **57**, 535–559.
- Gates, D.J. and Westcott, M. (1986) Clustering estimates for spatial point distributions with unstable potentials. *Ann. Inst. Statist. Math.*, **38**, 123–135.
- Georgii, H.-O. and Häggström, O. (1996) Phase transition in continuum Potts models. *Comm. Math. Phys.*, **181**, 507–528.
- Geyer, C.J. (1999) Likelihood inference for spatial point processes. In O. Barndorff-Nielsen, W.S. Kendall, and M.N.M. van Lieshout, (eds), *Proceedings of the Séminaire Européen de Statistique: Stochastic Geometry: Likelihood, and Computation*, pp. 79–140. Boca Raton: Chapman & Hall/CRC.
- Geyer, C.J. and Møller, J. (1994) Simulation procedures and likelihood inference for spatial point processes. *Scand. J. Statist.*, **21**, 359–373.

- Grimmett, G. (1995) The stochastic random-cluster process, and the uniqueness of random-cluster measures. *Ann. Probab.*, **23**, 1461–1510.
- Häggström, O. (1998) Random-cluster analysis of a class of binary lattice gases. *J. Statist. Phys.*, **91**, 47–74.
- Hamilton, W.D. (1971) Geometry for the selfish herd. *J. Theoret. Biol.*, **31**, 295–311.
- Hammersley, J.M., Lewis, J.W.E. and Rowlinson, J.S. (1975) Relationships between the multinomial and Poisson models of stochastic processes, and between the canonical and grand canonical ensembles in statistical mechanics, with illustrations and Monte Carlo methods for the penetrable sphere model of liquid–vapour equilibrium. *Sankhyā, Ser. A*, **37**, 457–491.
- Kallenberg, O. (1984) An informal guide to the theory of conditioning in point processes. *Internat. Statist. Rev.*, **52**, 151–164.
- Kelly, F.P. and Ripley, B.D. (1976) On Strauss’s model for clustering. *Biometrika*, **63**, 357–360.
- Kendall, W.S. (1998) Perfect simulation for the area-interaction point process. In L. Accardi and C.C. Heyde (eds), *Probability Towards 2000*, pp. 218–234. New York: Springer-Verlag.
- Kendall, W.S. and Möller, J. (1999) Perfect Metropolis–Hastings simulation of locally stable spatial point processes. Research Report 99–2001, Department of Mathematical Sciences, Aalborg University.
- Kendall, W.S., van Lieshout, M.N.M. and Baddeley, A.J. (1999) Quermass-interaction processes: stability properties. *Adv. Appl. Probab. (Stochastic Geom. Statist. Appl.)*, **31**. To appear.
- Klein, W. (1982) Potts-model formulation of continuum percolation. *Phys. Rev. B*, **26**, 2677–2678.
- Lindvall, T. (1992) *Lectures on the Coupling Method*. New York: Wiley.
- Møller, J. (1994) Discussion contribution. *Scand. J. Statist.*, **21**, 346–349.
- Møller, J. (1999) Markov chain Monte Carlo and spatial point processes. In O. Barndorff–Nielsen, W.S. Kendall, and M.N.M. van Lieshout (eds), *Stochastic Geometry: Likelihood, and Computation*, pp. 141–172. Boca Raton: Chapman & Hall/CRC.
- Møller, J. and Waagepetersen, R. (1996) Markov connected component fields. Research Report 96–2009, Department of Mathematics and Computer Science, Aalborg University. *Adv. in Appl. Probab. (Stochastic. Geom. Statist. Appl.)*, **30**, 1–35.
- Propp, J.G. and Wilson, D.B. (1996) Exact sampling with coupled Markov chains and applications to statistical mechanics. *Random Structures Algorithms*, **9**, 223–252.
- Ripley, B.D. and Kelly, F.P. (1977) Markov point processes. *J. London Math. Soc.*, **15**, 188–192.
- Rowlinson, J.S. (1980) Penetrable sphere models of liquid–vapor equilibrium. *Adv. in Chem. Phys.*, **41**, 1–57.
- Rowlinson, J.S. (1990) Probability densities for some one-dimensional problems in statistical mechanics. In G.R. Grimmett and D.J.A. Welsh (eds), *Disorder in Physical Systems*, pp. 261–276. Oxford: Clarendon Press.
- Ruelle, D. (1969) *Statistical Mechanics*. New York: Wiley.
- Ruelle, D. (1971) Existence of a phase transition in a continuous classical system. *Phys. Rev. Lett.*, **27**, 1040–1041.
- Strauss, D.J. (1975) A model for clustering. *Biometrika*, **63**, 467–475.
- Swendsen, R.H. and Wang, J.-S. (1987) Nonuniversal critical dynamics in Monte Carlo simulations. *Phys. Rev. Lett.*, **58**, 86–88.
- van Lieshout, M.N.M. and Molchanov, I.S. (1998) Shot-noise-weighted processes: a new family of spatial point processes. *Comm. Statist. Stochastic Models*, **14**, 715–734.
- Widom, B. and Rowlinson, J.S. (1970) A new model for the study of liquid–vapor phase transitions. *J. Chem. Phys.*, **52**, 1670–1684.

Received October 1996 and revised April 1998.