Statistica Neerlandica (2003) Vol. 57, nr. 2, pp. 177–206

The Candy model: properties and inference

M. N. M. van Lieshout* and R. S. Stoica**

CWI, P.O. Box 94079, 1090 GB Amsterdam, The Netherlands

In this paper we study the Candy model, a marked point process introduced by Stoica et al. (2000). We prove Ruelle and local stability, investigate its Markov properties, and discuss how the model may be sampled. Finally, we consider estimation of the model parameters and present a simulation study.

Key Words and Phrases: Candy model, Markov chain Monte Carlo simulation, Markov marked point process, maximum likelihood estimation, stability.

1 Set-up and notation

High resolution images found in fields such as microscopy, remote sensing and medicine pose interpretation problems that can often be formulated in geometric terms. For example, a microscopist may want to classify cells, and a doctor is interested in detecting abnormalities in scans. Thus, it is of prime interest to develop statistical models for object scenes to replace the more commonly used image models that operate on the pixel level. Convenient theoretical tools in this context are provided by the theory of marked point processes, in which an image is naturally regarded as a collection of objects, and inference focuses on locating and characterizing them. However, realistic models for the complicated objects encountered in practice are still scarce. Indeed, it is the (very ambitious) goal of the scientific community involved in this area to build and analyze rigorous mathematical models which can deal with the complex reality of high resolution images.

Stoica et al. (2000) and Stoica (2001) introduced a marked point process model for line segments – dubbed *Candy* – as prior distribution for the image analysis problem of extracting linear networks such as roads or rivers from images (usually obtained by aerial photography or satellites). In this paper we investigate the analytical properties of the model, focusing on the Ruelle condition, local stability and the interaction structure. We also study statistical aspects, including simulation by Markov chain Monte Carlo and parameter estimation.

We shall represent a line segment as a point in some compact subset $K \subset \mathbb{R}^2$ of strictly positive volume $0 < \nu(K) < \infty$ with an attached mark taking values in the

*M.N.M.van.Lieshout@cwi.nl
**Radu.Stoica@cwi.nl

Cartesian product $[l_{\min}, l_{\max}] \times [0, \pi)$ for some $0 < l_{\min} < l_{\max} < \infty$. Each marked point (k, l, θ) can be interpreted as a line segment with midpoint k, length l, and orientation θ. If required, an extra mark for the width of the segment may be added. Note that in the original formulation in (STOICA et al., 2000, and STOICA, 2001) the mark space for orientations is $[0, 2\pi]$.

A configuration of line segments is a finite set of marked points. Thus, for $n \in \mathbb{N}_0$, write S_n for the set of all (unordered) configurations $\mathbf{s} = \{s_1, \ldots, s_n\}$ that consist of n, not necessarily distinct, marked points $s_i \in S = K \times [l_{\min}, l_{\max}] \times [0, \pi)$. Hence, the configuration space can be written as $\Omega = \cup_{n=0}^{\infty} S_n$, which may be equipped with the σ-algebra $\mathcal{F}$ generated by the mappings $\{s_1, \ldots, s_n\} \mapsto \sum_{i=1}^{n} 1\{s_i \in A\}$ that count the number of marked points in Borel sets $A \subseteq S$. If the marks are discarded, the configuration space of midpoints is $\Omega_K = \cup_{n=0}^{\infty} K_n$, where K_n is the set of all configurations $\mathbf{x} = \{k_1, \ldots, k_n\}$ that consist of n, not necessarily distinct, points $k_i \in K$; the associated σ-algebra $\mathcal{F}_K$ is generated by the mappings counting the number of points falling in Borel subsets of K.

A point process on K is a measurable mapping from some probability space into $(\Omega_K, \mathcal{F}_K)$; a marked point process with points in K and marks in $[l_{\min}, l_{\max}] \times [0, \pi)$ is a point process on the product space $K \times [l_{\min}, l_{\max}] \times [0, \pi)$ with the additional property that the marginal process of segment centers is a point process on K. For further details, see (DALEY and VERE-JONES, 1988, Section 7.1).

Perhaps the simplest marked point process model is the Poisson process defined by the probability measure on $(\Omega, \mathcal{F})$ assigning mass

$$\mu(F) = \sum_{n=0}^{\infty} \frac{e^{-\nu(K)}}{n!} \{\pi(l_{\max} - l_{\min})\}^{-n} \int_S \cdots \int_S 1_F(\{(k_1, l_1, \theta_1), \ldots, (k_n, l_n, \theta_n)\}) \times \mathrm{d}\nu(k_1) \cdots \mathrm{d}\nu(k_n) \mathrm{d}l_1 \cdots \mathrm{d}l_n \mathrm{d}\theta_1 \cdots \mathrm{d}\theta_n$$

to $F \in \mathcal{F}$. In other words, under μ, midpoints are placed in K according to a Poisson process with intensity measure ν, to which points independent, uniformly distributed marks are assigned to determine the length and orientation. As it exhibits no interactions, the above Poisson marked point process is the ideal reference process. Indeed, one may define more complicated models by specifying a Radon–Nikodym derivative p with respect to μ.

The plan of this paper is as follows. In section 2, we define the Candy model in terms of its density (Radon-Nikodym derivative) with respect to μ. We establish the Ruelle condition and local stability. Furthermore, we define several relations on S, and investigate the Markov behavior of the Candy model. In section 3, a Metropolis–Hastings algorithm based on births and deaths is suggested for sampling from the Candy model. We discuss the convergence of the algorithm, and prove geometric ergodicity. More sophisticated updates including non-uniform births and deaths, and changes in the marks are discussed subsequently. Section 4 builds on the results obtained in previous sections to perform maximum likelihood based inference. The paper is concluded by simulation examples.

2 The Candy model: stability and Markov properties

2.1 *Model specification*

The Candy model was developed in the context of a concrete image analysis problem (Stoica et al., 2000), where, in order to decide whether two line segments were connected, discretization effects had to be taken into account. From a theoretical point of view, under the reference Poisson process almost surely no exact join between a pair of segments occurs. Such considerations motivate the following definition.

Definition 1. *Let $x = (k_x, l_x, \theta_x)$ and $y = (k_y, l_y, \theta_y)$ be two marked points. Then, x and y are said to be connected, $x \sim_c y$, if at least one of the following holds*

$$\begin{aligned}
&\left\| k_x + \tfrac{1}{2} l_x(\cos\theta_x, \sin\theta_x) - k_y - \tfrac{1}{2} l_y(\cos\theta_y, \sin\theta_y) \right\| \leqslant r_c \\
&\left\| k_x + \tfrac{1}{2} l_x(\cos\theta_x, \sin\theta_x) - k_y + \tfrac{1}{2} l_y(\cos\theta_y, \sin\theta_y) \right\| \leqslant r_c \\
&\left\| k_x - \tfrac{1}{2} l_x(\cos\theta_x, \sin\theta_x) - k_y - \tfrac{1}{2} l_y(\cos\theta_y, \sin\theta_y) \right\| \leqslant r_c \\
&\left\| k_x - \tfrac{1}{2} l_x(\cos\theta_x, \sin\theta_x) - k_y + \tfrac{1}{2} l_y(\cos\theta_y, \sin\theta_y) \right\| \leqslant r_c
\end{aligned}$$

for some given $r_c < l_{\min}$.

The relation of definition 1 is reflexive, that is, any $x \in S$ is connected to itself, and symmetric. Similarly, an endpoint e of a segment x is said to be connected in the configuration $\mathbf{s}$ if another segment in $\mathbf{s}$ can be found with at least one endpoint closer than r_c to e. Following Stoica et al. (2000) and Stoica (2001), we distinguish between *singly connected* segments with exactly one connected endpoint and *doubly connected* ones for which both endpoints are connected. A segment that is not connected to any other segment is said to be *free*.

Lemma 1. *The mappings n_f and n_c assigning to a configuration $\mathbf{s} \in \Omega$ the number of free, respectively singly connected segments are measurable with respect to $\mathcal{F}$.*

Proof: First, consider n_f. By its very nature, the mapping that counts the number of free segments in a configuration is a symmetric function of its argument. Thus, it is sufficient (Reiss, 1993, Section 3.1) to check that the function $f: S^n \to \mathbb{R}$ defined by

$$f(s_1, \ldots, s_n) = \sum_{i=1}^{n} 1\{s_i \text{ is free}\}$$

is Borel measurable for each $n \in \mathbb{N}_0$. Now, for fixed $i \neq j \in \{1, \ldots, n\}$, the function $f_{i,j}^1(s_1, \ldots, s_n)$ defined by

$$1\left\{ \left\| k_{s_i} + \frac{1}{2} l_{s_i}(\cos\theta_{s_i}, \sin\theta_{s_i}) - k_{s_j} - \frac{1}{2} l_{s_j}(\cos\theta_{s_j}, \sin\theta_{s_j}) \right\| > r_c \right\}$$

is Borel measurable as a mapping on S^n. Here, we use the notation $s_i = (k_{s_i}, l_{s_i}, \theta_{s_i})$.

Analogously, $f_{i,j}^2$, $f_{i,j}^3$ and $f_{i,j}^4$ defined similar to $f_{i,j}^1$ but using the second up to fourth condition of definition 1 instead of the first are Borel measurable. Consequently,

$$1\{s_i \text{ is free}\} = \prod_{j \neq i} \prod_{m=1}^{4} f_{i,j}^m(s_1, \ldots, s_n)$$

is Borel measurable, and so is the sum of these functions over i. A similar argument implies that n_c is measurable with respect to $\mathcal{F}$. □

Next, define two neighborhood relations on S.

DEFINITION 2. *Let $\delta > 0$. The relation $\sim_r$ on S is defined by*

$$x \sim_r y \Leftrightarrow \|k_x - k_y\| \leqslant \max\{l_x, l_y\}/2 \text{ and } \; | \, |\theta_x - \theta_y| - \pi/2 \, | > \delta$$

for any pair of marked points $x = (k_x, l_x, \theta_x)$ and $y = (k_y, l_y, \theta_y)$.

The relation $\sim_r$ is symmetric, and reflexive if $\delta < \pi/2$.

DEFINITION 3. *The influence zone $Z(s)$ of a marked point $s = (k, l, \theta) \in S$ is given by*

$$Z(s) = b\left(k + \frac{1}{2} l(\cos\theta, \sin\theta), \frac{1}{4} l\right) \cup b\left(k - \frac{1}{2} l(\cos\theta, \sin\theta), \frac{1}{4} l\right),$$

*the union of balls with radius $l/4$ around the endpoints. The relation $\sim_o$ on S is defined by $x \sim_o y \Leftrightarrow \|k_x - k_y\| > \frac{1}{2}\max\{l_x, l_y\}$ **and either** exactly one endpoint $k_x \pm \frac{1}{2} l_x(\cos\theta_x, \sin\theta_x)$ of x is a member of $Z(y)$ **or** exactly one endpoint $k_y \pm \frac{1}{2} l_y(\cos\theta_y, \sin\theta_y)$ of y is a member of $Z(x)$. Here $x = (k_x, l_x, \theta_x)$ and $y = (k_y, l_y, \theta_y)$ are elements of S.*

For a given configuration $\mathbf{s}$, write $n_r(\mathbf{s})$ for the number of unordered $\sim_r$ neighbor pairs in $\mathbf{s}$; similarly $n_o(\mathbf{s})$ denotes the number of $\sim_o$ neighbor pairs $\{x,y\}$ in $\mathbf{s}$ with the extra property that

$$\min\{|\theta_x - \theta_y|, \pi - |\theta_x - \theta_y|\} > \tau \qquad (*)$$

for some threshold value $\tau > 0$.

Note that $\sim_o$ is symmetric but not reflexive.

LEMMA 2. *The mappings n_r and n_o assigning to a configuration $\mathbf{s} \in \Omega$ the number of its $\sim_r$ neighbor pairs, respectively the number of its $\sim_o$ neighbor pairs satisfying $(*)$ are measurable with respect to $\mathcal{F}$.*

PROOF. The counting of marked point pairs satisfying the conditions mentioned above is a symmetric operation. Regarding n_r, for each (x, y), $1\{x \sim_r y\}$ is a Borel measurable function on S^2, from which observation the result follows as in the proof of lemma 1. A similar, slightly more involved, argument applies to n_o.

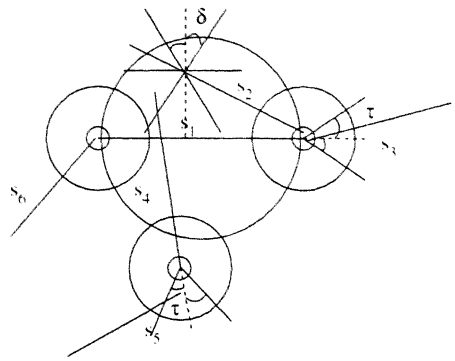


Fig. 1. Geometrical representation of interactions between segments.

In figure 1 we give a geometrical representation of some interactions between segments. Note that $s_1 \sim_c s_2$, $s_1 \sim_c s_3$ and $s_2 \sim_c s_3$; the segments s_4 and s_5 are free, s_2 and s_3 are singly connected and s_1 is doubly connected. When δ is small, $s_1 \sim_r s_2$ whereas $s_1 \not\sim_r s_4$. Furthermore, $s_1 \sim_o s_3$ as well as $s_4 \sim_o s_5$, but, for small τ, the pair $\{s_4, s_5\}$ contributes to n_o whereas $\{s_1, s_3\}$ does not.

We are now ready to give a definition of the Candy model by specifying its density p with respect to μ. Let $\mathbf{s} = \{s_1, \ldots, s_n\}$, $i = 1, \ldots, n = n(\mathbf{s})$, be a configuration of line segments. Then

$$p(\mathbf{s}) = \alpha\beta^{n(\mathbf{s})}\left\{\prod_{i=1}^{n(\mathbf{s})} \exp\left[\frac{l_{s_i} - l_{\max}}{l_{\max}}\right]\right\} \times \gamma_1^{n_f(\mathbf{s})}\gamma_2^{n_c(\mathbf{s})}\gamma_3^{n_r(\mathbf{s})}\gamma_4^{n_o(\mathbf{s})} \tag{1}$$

where $\gamma_1, \gamma_2, \gamma_3, \gamma_4 \in (0,1]$ and $\beta > 0$ are the model parameters, and $\alpha = p(\emptyset) > 0$ is the normalizing constant. STOICA et al. (2000) recommend $\gamma_1 < \gamma_2$. Thus, the model discourages free and singly connected segments, as well as sharp crossings and disagreements in orientation of close segments. In theorem 1 of section 2.2, we show that the model is well-defined.

2.2 *Stability*

The existence of any point process specified in terms of an unnormalized, measurable density p with respect to a Poisson point process is ensured by Ruelle's stability condition (RUELLE, 1969, Chapter 3, and GATES and WESTCOTT, 1986). This condition requires the energy $E(\mathbf{s}) = -\log(p(\mathbf{s})/p(\emptyset))$ to be bounded from below by a linear term in the number of marked points in $\mathbf{s}$, i.e. $E(\mathbf{s}) \geqslant -cn(\mathbf{s})$ for some $c > 0$, in which case the density (or the corresponding energy) is called *stable*. For the Candy model (1),

$$\begin{aligned} E(\mathbf{s}) &= -n(\mathbf{s})\log\beta - \sum_{i=1}^{n(\mathbf{s})} \frac{l_{s_i} - l_{\max}}{l_{\max}} \\ &\quad - n_f(\mathbf{s})\log\gamma_1 - n_c(\mathbf{s})\log\gamma_2 - n_r(\mathbf{s})\log\gamma_3 - n_o(\mathbf{s})\log\gamma_4 \\ &\geqslant -n(\mathbf{s})\log\beta. \end{aligned}$$

If $\beta > 1$, $E(\mathbf{s})$ is stable with $c = \log \beta$; otherwise $E(\mathbf{s}) \geqslant 0 \geqslant -cn(\mathbf{s})$ for any $c > 0$.

THEOREM 1. *The unnormalized Candy density* (1) *is* $(\Omega, \mathcal{F})$*-measurable and integrable, hence specifies a well-defined marked point process.*

PROOF: Measurability follows from lemmata 1–2, integrability is implied by the Ruelle condition. Indeed write X for the line segment process with density p. Then

$$E_\mu\left[\frac{p(X)}{p(\emptyset)}\right] \leqslant \sum_{n=0}^{\infty} \frac{e^{-v(K)}}{n!} (\max\{\beta, 1\})^n v(K)^n = \exp[\max\{\beta - 1, 0\} v(K)] < \infty. \square$$

A stronger stability condition is that of *local stability*, which requires the ratio $p(\mathbf{s} \cup \{\eta\})/p(\mathbf{s})$ to be uniformly bounded from above, both in $\mathbf{s} \in \Omega$ and $S \ni \eta \notin \mathbf{s}$, whenever $p(\mathbf{s}) > 0$.

LEMMA 3. *The Candy model* (1) *is locally stable.*

PROOF: Let $\mathbf{s} \in \Omega$, and $\eta = (k, l, \theta) \in S$. Since $p(\mathbf{s}) > 0$, the ratio $p(\mathbf{s} \cup \{\eta\})/p(\mathbf{s})$ is well-defined. Clearly, the addition of η results in an extra term $\beta \exp\{(l - l_{\max})/l_{\max}\} \leqslant \beta$ regardless of the position of η with respect to $\mathbf{s}$. The effect on the other four terms does depend on the type of connections introduced by η, which we investigate separately below.

First consider $n_f(\mathbf{s} \cup \{\eta\}) - n_f(\mathbf{s})$. If η is not connected to any segment in $\mathbf{s}$, the difference in free segments is 1. If η is singly connected, say through its endpoint e, by the addition of η to $\mathbf{s}$ the number of free segments decreases by the number of segments connected to e that were free in $\mathbf{s}$; since at most 6 segment endpoints separated by at least a distance r_c can be placed in a ball of radius r_c centered at e, in this case $n_f(\mathbf{s} \cup \{\eta\}) - n_f(\mathbf{s}) \geqslant -6$. Analogously, for doubly connected segments η, $n_f(\mathbf{s} \cup \{\eta\}) - n_f(\mathbf{s}) \geqslant -12$.

Next, turn to $n_c(\mathbf{s} \cup \{\eta\}) - n_c(\mathbf{s})$. If η is free, the number of singly connected segments does not change. If η is singly connected through its endpoint e, since the status of segments not connected to η is not affected, we have to examine segments connected to e. Now, segments that were free with respect to $\mathbf{s}$ might get singly connected in $\mathbf{s} \cup \{\eta\}$; if both endpoints of a segment were connected in $\mathbf{s}$, so are they in $\mathbf{s} \cup \{\eta\}$. Segments for which an endpoint connected to e was also connected in $\mathbf{s}$ and the other endpoint was free in $\mathbf{s}$ remain singly or become doubly connected in the new configuration $\mathbf{s} \cup \{\eta\}$. On the other hand, a segment that was singly connected in $\mathbf{s}$ but whose $\mathbf{s}$-free endpoint is connected to e becomes doubly connected after the addition of η. Hence $n_c(\mathbf{s} \cup \{\eta\}) - n_c(\mathbf{s})$ increases by 1 at least, and decreases by the number of segments that were singly connected in $\mathbf{s}$ with the free endpoint connected to e. Since there could be at most 6 segments of the latter type, $n_c(\mathbf{s} \cup \{\eta\}) - n_c(\mathbf{s}) \geqslant -5$. In the case where η is doubly connected, again we may restrict ourselves to considering the status of segments connected to η. As before, $n_c(\mathbf{s} \cup \{\eta\}) - n_c(\mathbf{s})$ decreases by at most

the number of segments that were singly connected in $\mathbf{s}$ with the free endpoint connected to η, a number that is bounded by 12.

Finally, note that $n_r(\mathbf{s} \cup \{\eta\}) \geqslant n_r(\mathbf{s})$ and $n_o(\mathbf{s} \cup \{\eta\}) \geqslant n_o(\mathbf{s})$. If we collect all the terms examined above, we obtain

$$\frac{p(\mathbf{s} \cup \{\eta\})}{p(\mathbf{s})} \leqslant \beta(\gamma_1, \gamma_2)^{-12}$$

and the proof is complete. □

2.3 *Markov properties*

A marked point process is said to be *(Ripley–Kelly) Markov* (RIPLEY and KELLY, 1977) with respect to some symmetric relation $\sim$ on S if its density is hereditary (that is $p(\mathbf{s}) > 0$ implies $p(\mathbf{s}') > 0$ for all $\mathbf{s}' \subseteq \mathbf{s}$), and if for all $\mathbf{s}$ such that $p(\mathbf{s}) > 0$, and all $\eta \notin \mathbf{s}$, the ratio $p(\mathbf{s} \cup \{\eta\})/p(\mathbf{s})$ depends only on η and those $s \in \mathbf{s}$ satisfying $s \sim \eta$. In physical terms, the energy required to add η to $\mathbf{s}$ depends only on η and its $\sim$-neighbors in $\mathbf{s}$. Equivalently, p is a Markov density if it can be written as a product of interactions associated with *cliques*: configurations in which each pair of elements consists of neighbors. By convention, the empty set and singletons are cliques as well. See the monograph (LIESHOUT, 2000) for further details.

PROPOSITION 1. *For $\gamma \in (0,1]$, the Candy model with probability density*

$$p(\mathbf{s}) \propto \gamma^{n_r(\mathbf{s})}, \qquad \mathbf{s} \in \Omega,$$

with respect to μ is Markov with respect to the relation $\sim_r$.

PROOF: The density is strictly positive, hence hereditary. Furthermore, for $\eta \notin \mathbf{s} \in \Omega$,

$$\frac{p(\mathbf{s} \cup \{\eta\})}{p(\mathbf{s})} = \gamma^{n_r(\mathbf{s} \cup \{\eta\}) - n_r(\mathbf{s})} = \gamma^{n(\{s \in \mathbf{s} : s \sim_r \eta\})}$$

depends only on the number of $\sim_r$-neighbors of η in $\mathbf{s}$. □

PROPOSITION 2. *For $\gamma \in (0,1]$, the Candy model with probability density*

$$p(\mathbf{s}) \propto \gamma^{n_o(\mathbf{s})}, \qquad \mathbf{s} \in \Omega,$$

with respect to μ is Markov with respect to the relation $\sim_o$.

PROOF: The density is strictly positive, hence hereditary. Furthermore, for $\eta = (k_\eta, l_\eta, \theta_\eta) \notin \mathbf{s} \in \Omega$,

$$\frac{p(\mathbf{s} \cup \{\eta\})}{p(\mathbf{s})} = \gamma^{n_o(\mathbf{s} \cup \{\eta\}) - n_o(\mathbf{s})} = \gamma^{n(\{s = (k_s, l_s, \theta_s) \in \mathbf{s} : s \sim_o \eta : \min\{|\theta_\eta - \theta_s|, \pi - |\theta_\eta - \theta_s|\} > \tau\})}$$

depends only on η and its $\sim_o$-neighbors in $\mathbf{s}$. □

Both the model of proposition 1 and that of proposition 2 exhibit pairwise interactions only, that is, factorize according to

$$p(\mathbf{s}) = p(\emptyset) \prod_{s_i \sim s_j} \phi(s_i, s_j)$$

where $\sim$ is either $\sim_r$ or $\sim_o$ and ϕ is a symmetric, measurable interaction function. It takes the constant value γ for the Candy model of proposition 1; for the model in proposition 2,

$$\phi(s_i, s_j) = \gamma^{1\{\min\{|\theta_i - \theta_j|, \pi - |\theta_i - \theta_j|\} > \tau\}}.$$

The statistics n_f and n_c call for a configuration dependent Markov property (BADDELEY and MØLLER, 1989), since in order to decide whether a given segment is free, singly, or doubly connected, one needs to examine the segments connected to it, and the segments connected to these, as well. The two-step iterated neighbors relation, also studied in (HAYAT and GUBNER, 1996, and GRABARNIK and SÄRKKÄ, 2001), is defined as follows. Based on the relation $\sim_c$ on S (definition 1), set

$$x \sim_{\mathbf{s}}^2 y \Leftrightarrow x \sim_c y \text{ or } \exists z \in \mathbf{s} : x \sim_c z \sim_c y$$

for $x, y \in \mathbf{s}$ and $\mathbf{s} \in \Omega$. A point process with density p is said to be *nearest-neighbor Markov* (BADDELEY and MØLLER, 1989) with respect to $\sim^2$ in the sense of Baddeley and Moller if p is hereditary, and if for any configuration $\mathbf{s}$ such that $p(\mathbf{s}) > 0$, and all $\eta \notin \mathbf{s}$, the ratio $p(\mathbf{s} \cup \{\eta\})/p(\mathbf{s})$ depends only on η, its two-step iterated neighbors in $\mathbf{s} \cup \{\eta\}$, and the relations $\sim_{\mathbf{s}}^2$ and $\sim_{\mathbf{s}\cup\{\eta\}}^2$ restricted to this neighborhood.

THEOREM 2. *For $\gamma_1, \gamma_2 \in (0,1]$, the Candy model with probability density*

$$p(\mathbf{s}) \propto \gamma_1^{n_f(\mathbf{s})} \gamma_2^{n_c(\mathbf{s})}, \quad \mathbf{s} \in \Omega,$$

with respect to μ is two-step iterated neighbors Markov with respect to the relation $\sim_c$ on S.

PROOF: By the proof of lemma 3,

$$\gamma_1^{n_f(\mathbf{s}\cup\{\eta\}) - n_f(\mathbf{s})}$$

depends only on the status of η and that of the segments connected to it. To decide the status of η, knowledge of its $\sim_c$-neighbors suffices; to assess the connection type of these neighbors, their neighbors have to be taken into account. The same is true for

$$\gamma_2^{n_c(\mathbf{s}\cup\{\eta\}) - n_c(\mathbf{s})}.$$

Consequently, p is a two-step iterated neighbors Markov point process with respect to the connection relation $\sim_c$. □

As a consequence of propositions 1–2 and theorem 2, we have the following result.

COROLLARY 1. *The Candy model is Ripley–Kelly Markov at fixed range* $2(l_{\max}+r_c)$ *regardless of the marks, i.e. with respect to the relation* $\sim$ *defined by*

$$x \sim y \Leftrightarrow \| k_x - k_y \| \leqslant 2(l_{\max} + r_c).$$

3 Metropolis–Hastings algorithms

3.1 *Review*

The Candy model (1) is too complicated to sample from directly. Rather, we apply Markov chain Monte Carlo techniques (GILKS et al., 1996, GEYER, 1999, and MØLLER, 1999) to construct a Markov chain that has the Candy model as its equilibrium distribution π, i.e.

$$\pi(F) = \int_F p(\mathbf{s}) \mathrm{d}\mu(\mathbf{s}) \tag{2}$$

for all $F \in \mathcal{F}$; as before, μ denotes the distribution of the reference Poisson process. An example of such a Markov chain is the Metropolis–Hastings sampler, originally introduced in statistical physics (METROPOLIS et al., 1953, and BARKER, 1965). It is a flexible proposal–acceptance technique (HASTINGS, 1970, and PESKUN, 1973) that is well adapted to point processes (OGATA and TANEMURA, 1981, GEYER and MØLLER, 1994, and GREEN, 1995). In that context, transitions must at least include births and deaths in order to jump between configurations containing different numbers of segments. The generic choice is as follows. Suppose a birth is proposed with probability p_b, and a death with the complementary probability $p_d = 1 - p_b$. In the case of a birth, a new segment is sampled uniformly, so that the birth proposal density can be written as

$$b(\mathbf{s}, \eta) = \frac{1}{v(K)}, \quad s \in \Omega, \ \eta \in S, \tag{3}$$

with respect to the product $\mathrm{d}\sigma(\eta) = (\mathrm{d}v(k)\mathrm{d}l\mathrm{d}\theta)/(\pi\,(l_{\max}-l_{\min}))$ of Lebesgue measure on K and uniform distributions on $[l_{\min}, l_{\max}]$ and $[0, \pi)$. It should be noted that (3) does not depend on the current configuration $\mathbf{s}$. The probability mass function of death proposals for points $\eta \in \mathbf{s}$ is given by

$$d(\mathbf{s}, \eta) = \frac{1}{n(\mathbf{s})} \tag{4}$$

for $\mathbf{s} \neq \emptyset$, i.e. each point η has the same probability of being removed. In the case $\mathbf{s} = \emptyset$, the new state is empty too.

A transition from $\mathbf{s}$ to $\mathbf{s}'$ is subsequently accepted with probability $\alpha(\mathbf{s},\mathbf{s}')$. The detailed balance equations require that, under the target equilibrium density p, the addition of $\eta \in S$ to $\mathbf{s} \in \Omega$ is matched by a death of η from $\mathbf{s} \cup \{\eta\}$, that is,

$$p_b b(\mathbf{s}, \eta)\alpha(\mathbf{s}, \mathbf{s} \cup \{\eta\})p(\mathbf{s}) = p_d d(\mathbf{s} \cup \{\eta\}, \eta)\alpha(\mathbf{s} \cup \{\eta\}, \mathbf{s})p(\mathbf{s} \cup \{\eta\}). \tag{5}$$

A solution is

$$\alpha(\mathbf{s}, \mathbf{s} \cup \{\eta\}) = \min\left\{1, \frac{p_d d(\mathbf{s} \cup \{\eta\}, \eta)p(\mathbf{s} \cup \{\eta\})}{p_b b(\mathbf{s}, \eta)p(\mathbf{s})}\right\} \tag{6}$$

with $\alpha(\mathbf{s} \cup \{\eta\}, \mathbf{s})$ given by substitution of (6) into (5). By the results in (GEYER and MØLLER, 1994, Section 4), the algorithm converges in total variation to π for π-almost all initial configurations provided $p_b \in (0, 1)$. The theorem applies equally to any pair of strictly positive proposal distributions, not necessarily equal to (3)–(4).

3.2 *Tailor-made updates*

3.2.1 *Connection-dependent transitions*

STOICA et al. (2000) (see also STOICA, 2001) used the following updates:

- birth and death of a free segment;
- birth and death of a singly connected segment with a single $\sim_c$-neighbor;
- birth and death of a singly connected segment with at least two $\sim_c$-neighbors;
- birth and death of a doubly connected segment.

Clearly, such moves are tailored to obtain connected configurations, but the subsets of S to which new segments of a given type must belong are quite complicated. Thus, STOICA et al. (2000) and STOICA (2001) felt forced to use approximations (both of the proposal density and the acceptance ratio) that jeopardize the convergence of the Markov chain to the correct target distribution.

A more tractable alternative is to design a probability density that tends to propose segments near to and aligned with the current network. The idea is that preference should be given to positions that 'fit' the current configuration. More specifically, a new segment might be positioned in such a way that it extends the current configuration.

Let us consider an endpoint e of a segment η, cf. figure 2. To sample a segment connected to e, we begin by choosing an orientation θ, say according to a probability density f with respect to the uniform distribution on $[0, \pi)$. Let $H(e, \eta)$ be the half-open half plane at e orthogonal to η that does not contain η. Now, since the center of the new segment must be an element of the set $K \cap H(e, \eta)$, the segment length cannot exceed twice the distance $l_{e,\eta}(\theta)$ of e to K^c along the line through e with orientation θ restricted to the half plane $H(e, \eta)$. Consequently, conditional on θ, we assume the length law to possess a density $g(\cdot|e, \eta, \theta)$ with respect to the uniform distribution on $[l_{\min}, l_{\max}]$ that is concentrated on $[l_{\min}, \min\{2l_{e,\eta}(\theta), l_{\max}\}]$. The update is completed by generating a midpoint k, uniformly or otherwise, on $M(e, \eta, \theta, l) = b(e \pm l(\cos\theta, \sin\theta)/2, r_c) \cap K$, the sign chosen so as to belong to $H(e,\eta)$. We will denote the probability density with respect to ν by $h(k|e, \eta, \theta, l)$. Clearly, the birth is possible only if the interval $[l_{\min}, \min\{2l_{e,\eta}(\theta), l_{\max}\}]$ and the set $M(e, \eta, \theta, l)$ both have strictly positive Lebesgue measure. In that case, the proposal density at endpoint e of segment η is given by

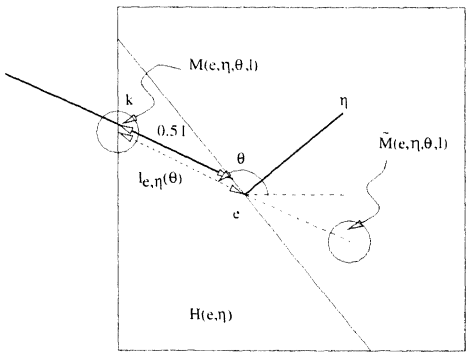


Fig. 2. How to extend the network.

$$\tilde{b}(e, \eta, (k, l, \theta)) = h(k|e, \eta, \theta, l) g(l|e, \eta, \theta) f(\theta) \tag{7}$$

where $\theta \in [0, \pi)$, $l \in [l_{\min}, \min\{2l_{e,\eta}(\theta), l_{\max}\}]$, and $k \in M(e, \eta, \theta, l)$; otherwise $\tilde{b}(e, \eta, (k, l, \theta)) = 0$. In summary, provided $A(\mathbf{s}) \neq \emptyset$ for $\mathbf{s} \in \Omega$, the proposal density for prolonging the segment configuration $\mathbf{s}$ is given by the average

$$b_p(\mathbf{s}, (k, l, \theta)) = \frac{1}{n(A(\mathbf{s}))} \sum_{(e,\eta) \in A(\mathbf{s})} \tilde{b}(e, \eta, (k, l, \theta)) \tag{8}$$

of (7) over $A(\mathbf{s})$, the set of endpoint-segment pairs (e, η), $\eta \in \mathbf{s}$, allowing addition of a new segment to e. If $n(A(\mathbf{s})) = 0$, a uniformly distributed birth is proposed as in (3).

Examples of (7) include uniform updates

$$\begin{aligned} f(\theta) &= 1; \\ g(l|e, \eta, \theta) &= \frac{(l_{\max} - l_{\min}) 1\{l \in [l_{\min}, \min\{2l_{e,\eta}(\theta), l_{\max}\}]\}}{\min\{2l_{e,\eta}(\theta), l_{\max}\} - l_{\min}}; \\ h(k|e, \eta, \theta, l) &= \frac{1\{k \in M(e, \eta, \theta, l)\}}{\nu(M(e, \eta, \theta, l) \cap K)}, \end{aligned} \tag{9}$$

again assuming non-zero denominators. Alternatively, the orientation could be centered around that of η, for example by means of a Beta distribution, to favor a better alignment.

In the simulations of section 5 we connect only to segment endpoints $e \in \eta$ further than $\frac{1}{2} l_{\max} + r_c$ away from K^c; the current connections to e may be taken into account as well, as illustrated in figure 3. With this convention, for any θ, $g(\cdot|e, \eta, \theta)$ may be positive on the maximal interval $[l_{\min}, l_{\max}]$ and the putative midpoint is sampled on a full ball of area πr_c^2.

Back bends, although penalized by the model for most values of γ_4 and τ (see (1)), may be formalized by sampling a new center in the subset $\tilde{M}(e, \eta, \theta, l)$ of $H(e, \eta)^c$ for given l and θ, as indicated in figure 2. Note that the two directed distances to the boundary of K along a line through e with orientation θ may well be different,

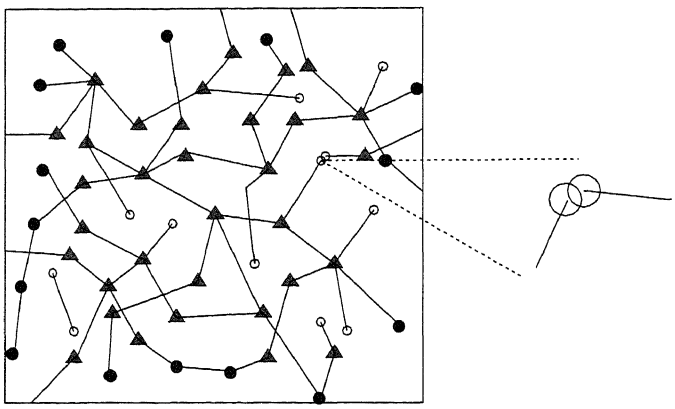


Fig. 3. Extremities marked by triangles are connected and further than $\frac{1}{2}l_{\max} + r_c$ to the boundary, those labeled by a black disk are closer than $\frac{1}{2}l_{\max} + r_c$ to the boundary of K.

leading to conditional length distributions that are concentrated on different supports. In practice, we restrict ourselves to extremities that are far away from the boundary, hence both distributions are concentrated on the full support $[l_{\min}, l_{\max}]$. Thus, a mixture proposal distribution for prolongation and back bends could take the following form. Choose an orientation θ according to a probability density f with respect to the uniform distribution on $[0, \pi)$. Conditionally given θ, the length is sampled according to a density $g(\cdot|e, \eta, \theta)$ with respect to the uniform distribution on $[l_{\min}, l_{\max}]$. Finally, with probability p_M a midpoint is sampled on $M(e, \eta, \theta, l)$, say uniformly; with the complementary probability $p_{\tilde{M}} = 1 - p_M$, a center is generated on $\tilde{M}(e, \eta, \theta, l)$ (see figure 2).

3.2.2 *Modifying the segment characteristics*

To improve the mixing of the Markov chain, apart from adapting the birth proposal density to the target density, a common strategy is to include transition types other than births and deaths. Thus, in (Stoica et al., 2000, and Stoica, 2001), the following updates are considered:

- changing the orientation of a singly connected segment;
- changing the length of a singly connected segment;
- changing the position of a singly connected segment with a single $\sim_c$-neighbor;
- changing the position of a singly connected segment with at least two $\sim_c$-neighbors.

The classification according to connection types has the same drawback as for the birth and death moves of section 3.2.1. Here we present some alternatives that are easier to implement.

In the set-up described in (Ogata and Tanemura, 1981, and Geyer and Møller, 1994), transitions from $\mathbf{s} \neq \emptyset$ to $\mathbf{s}' = (\mathbf{s} \backslash \{\eta\}) \cup \{\zeta\}$ for $\eta \in \mathbf{s}$ and $\zeta \in S$ are governed

by the proposal kernel $c(\mathbf{s}, \eta, \zeta)$ and acceptance probabilities $\alpha(\mathbf{s}, (\mathbf{s} \setminus \{\eta\}) \cup \{\zeta\})$. Thus, for each choice of $\mathbf{s} \in \Omega$ and $\eta \in \mathbf{s}$, $c(\mathbf{s}, \eta, \cdot)$ is a probability density (with respect to the intensity measure of the reference Poisson process law μ) governing the change of $\eta \in \mathbf{s}$, and the proposal to replace η by ζ is accepted with probability $\alpha(\mathbf{s},(\mathbf{s} \setminus \{\eta\}) \cup \{\zeta\})$. If a member η of configuration $\mathbf{s}$ is selected for modification with probability $q(\mathbf{s},\eta)$, the detailed balance equations require that $p(\mathbf{s})q(\mathbf{s}, \eta)c(\mathbf{s}, \eta, \zeta)\alpha(\mathbf{s}, (\mathbf{s} \setminus \{\eta\}) \cup \{\zeta\}) = p((\mathbf{s} \setminus \{\eta\}) \cup \{\zeta\})q((\mathbf{s} \setminus \{\eta\}) \cup \{\zeta\}, \zeta)c((\mathbf{s} \setminus \{\eta\}) \cup \{\zeta\}, \zeta, \eta)\alpha((\mathbf{s} \setminus \{\eta\}) \cup \{\zeta\}, \mathbf{s})$ whenever $p(\mathbf{s})$, $p((\mathbf{s} \setminus \{\eta\}) \cup \{\zeta\}) > 0$. We assume the selection probabilities are strictly positive, and impose the condition that $c(\mathbf{s}, \eta, \zeta) > 0$ if and only if $c((\mathbf{s} \setminus \{\eta\}) \cup \{\zeta\}, \zeta, \eta) > 0$. In other words, if $\eta \in \mathbf{s}$ may be changed into ζ, the reverse update is also possible. Then,

$$\alpha(\mathbf{s},(\mathbf{s}\setminus\{\eta\})\cup\{\zeta\}) := \min\left\{1, \frac{p((\mathbf{s}\setminus\{\eta\})\cup\{\zeta\})q((\mathbf{s}\setminus\{\eta\})\cup\{\zeta\},\zeta)c((\mathbf{s}\setminus\{\eta\})\cup\{\zeta\},\zeta,\eta)}{p(\mathbf{s})q(\mathbf{s},\eta)c(\mathbf{s},\eta,\zeta)}\right\} \tag{10}$$

is well-defined and solves the detailed balance equations.

Within the general context described above, there are many valid choices for the proposal kernel. To implement uniformly distributed joint 'local' changes, let $C(\eta) = C_k(k_\eta) \times C_m(l_\eta, \theta_\eta) \subseteq K \times ([l_{\min}, l_{\max}] \times [0, \pi))$ be a neighborhood of the segment $\eta = (k_\eta, l_\eta, \theta_\eta)$ such that $\nu(C_k(k_\eta))$ and $\nu(C_m(l_\eta, \theta_\eta))$ are both strictly positive, and set

$$c(\mathbf{s}, \eta, (k, l, \theta)) = \frac{1\{k \in C_k(k_\eta)\}}{\nu(C_k(k_\eta))} \frac{(l_{\max} - l_{\min})\pi 1\{(l, \theta) \in C_m(l_\eta, \theta_\eta)\}}{\nu(C_m(l_\eta, \theta_\eta))}.$$

In order to ensure reversibility, we have to require that $\zeta \in C(\eta)$ whenever $\eta \in C(\zeta)$. Typically, $C(\eta)$ will be relatively small and centered at η. If $C(\eta) = S$, the local character is lost, and a new segment is proposed uniformly over the whole space. The latter has the potential advantage of moving faster through the state space, the former of fine tuning likely configurations without destroying the overall appearance of the network. If C_m is of Cartesian product form, and the proposal density factorizes with respect to its position, length and orientation component, the modification may be implemented sequentially.

Change transitions are also useful for performing a death followed by a birth in one step, especially if the acceptance probability for the death is low. Thus, as in section 3, let $b(\cdot, \cdot)$ and $d(\cdot, \cdot)$ be strictly positive, and set

$$q(\mathbf{s}, \eta) = d(\mathbf{s}, \eta); \qquad c(\mathbf{s}, \eta, \zeta) = b(\mathbf{s} \setminus \{\eta\}, \zeta) \tag{11}$$

for the proposal to move from $\mathbf{s} \in \Omega$ to $(\mathbf{s} \setminus \{\eta\}) \cup \{\zeta\}$ for some $\eta \in \mathbf{s}$, $\zeta \in S$.

A second type of update is to change a single segment component, say the orientation. Thus, for each $\eta = (k_\eta, l_\eta, \theta_\eta) \in \mathbf{s} \in \Omega$, we define a probability density $c_\theta(\mathbf{s}, \eta, \cdot)$ with respect to the uniform distribution on $[0, \pi)$. Given a neighborhood $C_\theta(\theta_\eta) \subseteq [0, \pi)$ of θ_η with positive length, one might set

$$c_\theta(\mathbf{s}, \eta, \theta) = \frac{\pi 1\{\theta \in C_\theta(\theta_\eta)\}}{\text{length}(C_\theta(\theta_\eta))}. \tag{12}$$

If $C_\theta(\theta_\eta) = [0, \pi)$, the new orientation is sampled uniformly over its full range; more commonly, a value in some small neighborhood of the current one is proposed. Again, we denote the probability that $\eta \in \mathbf{s}$ is selected for modification by $q(\mathbf{s}, \eta)$, and assume positivity. Then, the detailed balance equations read

$$\begin{aligned} &p(\mathbf{s})q(\mathbf{s},\eta)c_\theta(\mathbf{s},\eta,\theta)\alpha(\mathbf{s},(\mathbf{s}\backslash\{\eta\})\cup\{(k_\eta,l_\eta,\theta)\}) \\ &\quad = p((\mathbf{s}\backslash\{\eta\})\cup\{(k_\eta,l_\eta,\theta)\})q((\mathbf{s}\backslash\{\eta\})\cup\{(k_\eta,l_\eta,\theta)\},(k_\eta,l_\eta,\theta)) \\ &\qquad \times c_\theta((\mathbf{s}\backslash\{\eta\})\cup\{(k_\eta,l_\eta,\theta)\},(k_\eta,l_\eta,\theta),\theta_\eta)\alpha((\mathbf{s}\backslash\{\eta\})\cup\{(k_\eta,l_\eta,\theta)\},\mathbf{s}) \end{aligned}$$

whenever $p(\mathbf{s})$, $p((\mathbf{s}\backslash\{\eta\}) \cup \{(k_\eta, l_\eta, \theta)\})$ are positive. We assume that $c_\theta(\mathbf{s}, \eta, \theta_\zeta) > 0$ if and only if $c_\theta((\mathbf{s}\backslash\{\eta\}) \cup \{\zeta\},\zeta,\theta_\eta) > 0$ whenever η and ζ differ only in their orientation component. Then,

$$\begin{aligned} &\alpha(\mathbf{s},(\mathbf{s}\backslash\{\eta\})\cup\{\zeta\}) \\ &\quad := \min\left\{1, \frac{p((\mathbf{s}\backslash\{\eta\})\cup\{\zeta\})q((\mathbf{s}\backslash\{\eta\})\cup\{\zeta\},\zeta)c_\theta((\mathbf{s}\backslash\{\eta\})\cup\{\zeta\},\zeta,\theta_\eta)}{p(\mathbf{s})q(\mathbf{s},\eta)c_\theta(\mathbf{s},\eta,\theta_\zeta)}\right\} \end{aligned} \tag{13}$$

is well-defined and solves the detailed balance equations.

Similarly, one may define a proposal density $c_k(\mathbf{s}, \eta, \cdot)$ with respect to ν on K for modifying the position of a segment, or $c_l(\mathbf{s}, \eta, \cdot)$ with respect to the uniform distribution on $[l_{\min}, l_{\max}]$ for the length.

3.3 *Convergence*

In this section, we investigate the limit behavior of the Metropolis–Hastings algorithm with transitions as described in sections 3.1 and 3.2. As in (2), write π for the law of the Candy model and denote the product measure on S by σ. The transition kernel is

$$\begin{aligned} P(\emptyset, F) &= p_b \int_S b(\emptyset,\eta)\alpha(\emptyset,\{\eta\})1_F(\{\eta\})\mathrm{d}\sigma(\eta) \\ &\quad + 1_F(\emptyset)\left[1 - p_b\int_S b(\emptyset,\eta)\alpha(\emptyset,\{\eta\})\mathrm{d}\sigma(\eta)\right] \end{aligned} \tag{14}$$

for $\mathbf{s} = \emptyset$, $F \in \mathcal{F}$, and $P(\mathbf{s}, F)$ equals

$$\begin{aligned} &p_b\int_S b(\mathbf{s},\eta)\alpha(\mathbf{s},\mathbf{s}' := \mathbf{s}\cup\{\eta\})1_F(\mathbf{s}')d\sigma(\eta) + p_d\sum_{s_i\in\mathbf{s}} d(\mathbf{s},s_i)\alpha(\mathbf{s},\mathbf{s}' := \mathbf{s}\backslash\{s_i\})1_F(\mathbf{s}') \\ &\quad + p_c\sum_{s_i\in\mathbf{s}} q(\mathbf{s},s_i)\int_S c(\mathbf{s},s_i,\eta)\alpha(\mathbf{s},\mathbf{s}' := (\mathbf{s}\backslash\{s_i\})\cup\{\eta\})1_F(\mathbf{s}')\mathrm{d}\sigma(\eta) \\ &\quad + p_\theta\sum_{s_i\in\mathbf{s}} q(\mathbf{s},s_i)\int_0^\pi c_\theta(\mathbf{s},s_i,\theta)\alpha(\mathbf{s},\mathbf{s}' := (\mathbf{s}\backslash\{s_i\})\cup\{(k_{s_i},l_{s_i},\theta)\})1_F(\mathbf{s}')\frac{\mathrm{d}\theta}{\pi} \end{aligned}$$

$$+p_k \sum_{s_i \in \mathbf{s}} q(\mathbf{s}, s_i) \int_K c_k(\mathbf{s}, s_i, k)\alpha(\mathbf{s}, \mathbf{s}' := (\mathbf{s}\backslash\{s_i\}) \cup \{(k, l_{s_i}, \theta_{s_i})\}) 1_F(\mathbf{s}')\mathrm{d}\nu(k)$$

$$+ 1_F(\mathbf{s}) \Bigg[1 - p_b \int_S b(\mathbf{s}, \eta)\alpha(\mathbf{s}, \mathbf{s} \cup \{\eta\})\mathrm{d}\sigma(\eta) - p_d \sum_{s_i \in \mathbf{s}} \mathrm{d}(\mathbf{s}, s_i)\alpha(\mathbf{s}, \mathbf{s}\backslash\{s_i\})$$

$$- p_c \sum_{s_i \in \mathbf{s}} q(\mathbf{s}, s_i) \int_S c(\mathbf{s}, s_i, \eta)\alpha(\mathbf{s}, (\mathbf{s}\backslash\{s_i\}) \cup \{\eta\})\mathrm{d}\sigma(\eta)$$

$$- p_\theta \sum_{s_i \in \mathbf{s}} q(\mathbf{s}, s_i) \int_0^\pi c_\theta(\mathbf{s}, s_i, \theta)\alpha(\mathbf{s}, (\mathbf{s}\backslash\{s_i\}) \cup \{(k_{s_i}, l_{s_i}, \theta)\}) \frac{\mathrm{d}\theta}{\pi}$$

$$- p_k \sum_{s_i \in \mathbf{s}} q(\mathbf{s}, s_i) \int_K c_k(\mathbf{s}, s_i, k)\alpha(\mathbf{s}, (\mathbf{s}\backslash\{s_i\}) \cup \{(k, l_{s_i}, \theta_{s_i})\})\mathrm{d}\nu(k) \Bigg] \quad (15)$$

otherwise. Here, p_c, p_θ and p_k are the probabilities of performing a change update, a modification of orientation and position respectively. The densities associated with the various transition proposals and the acceptance probabilities are as described in sections 3.1 and 3.2.

Let $L(\mathbf{s}, F)$ be the probability that the Markov chain started at $\mathbf{s} \in \Omega$ ever hits the set $F \in \mathcal{F}$. The chain is said to be *Harris recurrent* (MEYN and TWEEDIE, 1993, Chapter 9, and GEYER, 1999) if $L(\mathbf{s}, F) = 1$ for all $\mathbf{s} \in \Omega$ and all $F \in \mathcal{F}$ with $\pi(F) > 0$. In words, all π-positive sets F are almost surely reached eventually from every initial state. Moreover, such sets will be visited infinitely often (MØLLER, 1999). The weaker condition of π-irreducibility requires only $L(\mathbf{s}, F) > 0$ for all $\mathbf{s} \in \Omega$ and all π-positive F, or equivalently $P^n(\mathbf{s}, F) > 0$ for some $n \in \mathbb{N}_0$.

An even stronger property than Harris recurrence is *geometric ergodicity*, that is geometric convergence in total variation:

$$||P^n(\mathbf{s}, \cdot) - \pi||_{TV} \leqslant c(\mathbf{s})\gamma^n$$

for some constant $\gamma < 1$ and some π-integrable, non-negative function c. This property is important in establishing a central limit theorem for the sample path average of certain π-integrable functions (MEYN and TWEEDIE, 1993, Chapter 17, GEYER, 1999, and MØLLER, 1999). Geometric ergodicity can be proved by means of the so-called geometric drift condition (MEYN and TWEEDIE, 1993, Theorem 15.0.1). In order to state this condition, we need the concept of a small set. A set $C \in \mathcal{F}$ is *small* if $\pi(C) > 0$ and there exists a probability measure φ on $\mathcal{F}$, a constant $\epsilon > 0$, and an integer $n \in \mathbb{N}_0$ such that

$$P^n(\mathbf{s}, F) \geqslant \epsilon\varphi(F)$$

for all $\mathbf{s} \in C$ and all $F \in \mathcal{F}$. Now, the geometric drift condition entails the existence of a function $V : \Omega \to [1, \infty)$, constants $a < 1$ and $b < \infty$, and a small set $C \in \mathcal{F}$ such that

$$\int_\Omega V(\mathbf{s}')P(\mathbf{s}, d\mathbf{s}') \leqslant aV(\mathbf{s}) + b1\{\mathbf{s} \in C\} \quad (16)$$

for all $\mathbf{s} \in \Omega$.

For further details on Markov chains on general state spaces, see e.g. the textbook by MEYN and TWEEDIE (1993).

THEOREM 3. *Let the functions b, d, c, c_k, c_0 and α be as described in sections* 3.1–3.2.2, *and in particular suppose that the birth proposal density and the death proposal probabilities are strictly positive. Assume that*

$$u_n = \sup_{\eta \in S, \mathbf{s} \in S_n} \frac{d(\mathbf{s} \cup \{\eta\}, \eta)}{b(\mathbf{s}, \eta)} \to 0$$

as $n \to \infty$, and that $p_b + p_d + p_c + p_k + p_0 = 1$ with $p_b, p_d \in (0,1)$ and $p_c, p_k, p_0 \in [0,1)$. Then the Metropolis–Hastings sampler for the Candy model (1) defined by (14)–(15) is geometrically ergodic.

The proof is an adaptation to the Candy model of the proof of (GEYER, 1999, Proposition 3.3).

PROOF: By lemma 3, the Candy model is locally stable. Let $\lambda > 0$ be an upper bound to the likelihood ratio, and set $V(\mathbf{s}) = A^{n(\mathbf{s})}$ for some $A > 1$.

The acceptance probability (6) for adding $\eta \notin \mathbf{s}$ to $\mathbf{s}$ is

$$\min\left\{1, \frac{p_d d(\mathbf{s} \cup \{\eta\}, \eta) p(\mathbf{s} \cup \{\eta\})}{p_b b(\mathbf{s}, \eta) p(\mathbf{s})}\right\} \leqslant \frac{p_d \lambda}{p_b} \times u_{n(\mathbf{s})},$$

which, as $u_{n(\mathbf{s})}$ tends to 0, does not exceed a prefixed constant $\epsilon > 0$ if $n(\mathbf{s})$ is sufficiently large. Similarly, the acceptance probability for removing $\eta \notin \mathbf{s}$ from $\mathbf{s} \cup \eta$ equals

$$\min\left\{1, \frac{p_b b(\mathbf{s}, \eta) p(\mathbf{s})}{p_d d(\mathbf{s} \cup \{\eta\}, \eta) p(\mathbf{s} \cup \{\eta\})}\right\} \geqslant \min\left\{1, \inf\left[\frac{b(\mathbf{s}, \eta)}{d(\mathbf{s} \cup \{\eta\}, \eta)}\right] \times \frac{p_b}{p_d \lambda}\right\}$$

which reduces to 1 since the assumptions of the theorem imply

$$\inf_{\eta \in S, \mathbf{s} \in S_n} \frac{b(\mathbf{s}, \eta)}{d(\mathbf{s} \cup \{\eta\}, \eta)} \to \infty$$

as $n(\mathbf{s})$ tends to infinity.

For the Metropolis–Hastings transition kernel (15),

$$\begin{aligned} \int_\Omega V(\mathbf{s}') P(\mathbf{s}, d\mathbf{s}') = p_b A^{n(\mathbf{s})} \int_S b(\mathbf{s}, \eta)(A - 1)\alpha(\mathbf{s}, \mathbf{s} \cup \{\eta\}) \mathrm{d}\sigma(\eta) \\ + p_d A^{n(\mathbf{s})} \sum_{\eta \in \mathbf{s}} d(\mathbf{s}, \eta)(A^{-1} - 1)\alpha(\mathbf{s}, \mathbf{s} \backslash \{\eta\}) + A^{n(\mathbf{s})}. \end{aligned} \tag{17}$$

For line segment configurations $\mathbf{s}$ of sufficiently large cardinality, say $n(\mathbf{s}) > N_\epsilon$, $\alpha(\mathbf{s}, \mathbf{s} \setminus \{\eta\}) = 1$ and $\alpha(\mathbf{s}, \mathbf{s} \cup \{\eta\}) \leqslant \epsilon$, hence, recalling $A > 1$, (17) is less than or equal to

$$\left[p_b(A - 1)\epsilon + p_d(A^{-1} - 1) + 1\right] V(\mathbf{s}).$$

Since we have not yet specified ϵ, and the multiplier of $V(\mathbf{s})$ in the right hand side converges to $1+p_d(A^{-1}-1) = p_b+p_c+p_k+p_0+p_d/A < 1$ as ϵ tends to zero, we can pick ϵ such that $\int_\Omega V(\mathbf{s}')P(\mathbf{s}, d\mathbf{s}') \leqslant aV(\mathbf{s})$ for some $a<1$.

Now, the set $C = \{\mathbf{s} \in \Omega : n(\mathbf{s}) \leqslant N_\epsilon\}$ is small. Indeed, the acceptance probability of a down step exceeds $\Delta := \min\{p_b/(u_n p_d \lambda) : n \leqslant N_\epsilon\}$. Without loss of generality, Δ is strictly less than 1. Moreover, $P(\emptyset, S_0) \geqslant p_d$. Hence,

$$P^{N_\epsilon}(\mathbf{s}, S_0) \geqslant P^{n(\mathbf{s})}(\mathbf{s}, S_0) P^{N_\epsilon - n(\mathbf{s})}(\emptyset, S_0) \geqslant (p_d\Delta)^{N_\epsilon}$$

for any configuration $\mathbf{s}$ consisting of at most N_ϵ segments. Hence, C is small with scalar multiplier $(p_d\Delta)^{N_\epsilon}$ to the Dirac measure on $\emptyset$.

We have seen that (16) is satisfied for $\mathbf{s} \notin C$. For $\mathbf{s} \in C$, the geometric drift condition holds if we take $b = A^{N_\epsilon + 1}$. □

Since self-transitions occur with positive probability, the Metropolis–Hastings chain is aperiodic, and the proof of theorem 3 implies the chain is Harris recurrent (GEYER, 1999).

For the uniform birth and death kernels (3)–(4),

$$u_n = \frac{v(K)}{n+1} \to 0$$

as $n \to \infty$. In the simulation study described below in section 5, we use a uniform death kernel (4) and the mixture

$$b(\mathbf{s}, \eta) = p_{1b}\frac{1}{v(K)} + p_{2b}b_p(\mathbf{s}, \eta) \tag{18}$$

with $b_p(\cdot, \cdot)$ given by (8) to govern births. Here $p_{1b} = 1 - p_{2b} \in [0, 1]$. If $p_{1b} = 0$, when choosing uniform updates (9), it is readily verified that $u_n \to 0$ still. However, if another density is preferred in constructing the updates, for example a Beta distribution, then $u_n \to \infty$ might diverge. To avoid this problem we shall always take p_{1b} strictly positive. Then, by arguments similar to the ones above for $p_{1b} = 1$, u_n tends to zero if n increases to infinity. Hence, our choice of the mixture (18) is justified both by theoretical reasoning and by the fact that it improves the convergence properties of our Markov chain (cf. section 5).

3.4 *Discussion*

In the preceding sections we discussed a range of updates that may be used as ingredients for a Metropolis–Hastings sampler. Although we tried to be rather general, yet other types of moves can be envisaged. For instance, it is possible to merge two close segments into one, or reversely to split a large one in two (GRENANDER and MILLER, 1994, RUE and HUSBY, 1998, RUE and SYVERSVEEN, 1998, or RUE and HURN, 1999). However, one would have to be careful in order to guarantee that the length of the new segment is in the interval $[l_{\min}, l_{\max}]$. It would also be possible to update several segments at the same time.

It is important to stress that a uniformly optimal sampler does not exist. For $\gamma_i = 1$ for $i = 1, \ldots, 4$, the Candy model reduces to a Poisson line segment process, and simple uniform birth and death proposals will suffice. For stronger interaction, more weight should be given to updates that result in more likely patterns. In practice, in order to build a sampler that convergences in a reasonable time, some experimentation is needed to find a balance between the various moves that accomplishes these objectives.

Finally, note that in order to assess whether the algorithm has converged, diagnostic tests based on the sufficient statistics of the model are widely used, see e.g. (Stoica, 2001). However, such tests only serve to falsify, that is, to indicate convergence is not yet reached. Theoretically, since the Candy model is locally stable (cf. lemma 3), coupling into and from the past (Propp and Wilson, 1996, and Kendall and Møller, 2000) can be used to obtain exact samples from (1), but due to the lack of monotonicity, it seems to be rather cumbersome in practice, especially in case of strong interaction between the segments.

4 Maximum likelihood estimation

The Candy model (1) is a five-parameter exponential family

$$p_\theta(\mathbf{s}) = \alpha(\theta) h(\mathbf{s}) \exp\left[t(\mathbf{s})^T \log \theta\right]$$

with normalizing constant $\alpha(\theta)$, $h(\mathbf{s}) = \prod_{i=1}^{n(\mathbf{s})} \exp\left[\frac{l_i - l_{\max}}{l_{\max}}\right]$, canonical sufficient statistic $t(\cdot) = (n(\cdot), n_A(\cdot), n_+(\cdot), n_x(\cdot), n_0(\cdot))^T$, and parameter vector $\theta = (\beta, \gamma_1, \gamma_2, \gamma_3, \gamma_4)^T$. Upon observing a pattern $\mathbf{s}$, consider the log likelihood ratio

$$l(\theta) = \log \frac{p_\theta(\mathbf{s})}{p_{\theta_0}(\mathbf{s})} = \log \frac{\alpha(\theta)}{\alpha(\theta_0)} + t(\mathbf{s})^T (\log \theta - \log \theta_0)$$

with respect to some reference value $\theta_0 \in (0, \infty) \times (0,1]^4$. For notational convenience, from now on we shall write $\omega = \log \theta$ component wise and express the log likelihood ratio as a function of ω. It is well known (Geyer and Thompson, 1992, and Geyer, 1999) that $\alpha(\omega_0)/\alpha(\omega) = E_{\omega_0} \exp\left[t(X)^T(\omega - \omega_0)\right]$. Hence, the log likelihood ratio can be rewritten as follows

$$l(\omega) = t(\mathbf{s})^T (\omega - \omega_0) - \log E_{\omega_0} \exp\left[t(X)^T (\omega - \omega_0)\right] \tag{19}$$

from which it is easy to derive the score equations $\nabla l(\omega) = t(\mathbf{s}) - E_\omega t(X)$ and Fisher information matrix $-\nabla^2 l(\omega) = \mathrm{Var}_\omega t(X)$. In summary, the maximum likelihood equations

$$E_{\hat{\omega}} t(X) = t(\mathbf{s}) \tag{20}$$

state that under $\hat{\omega}$, the expected values of the sufficient statistics must be equal to the observed values. Now, since the covariance matrix of $t(X)$ is positive definite, (19) is concave in ω. Therefore, provided the score equations have a solution $\hat{\omega}$ in $\mathbb{R} \times \mathbb{R}_-^4$,

a unique maximum likelihood estimator exists and equals $\hat{\omega}$. Otherwise, a maximum may be found on the boundary of the parameter space.

To solve (20), GEYER and THOMPSON (1992), GEYER (1994) and GEYER (1999) suggested approximating the expectation in (19) by its Monte Carlo counterpart

$$\sum_{i=1}^{m} \exp\left[t(X_i)^T(\omega - \omega_0)\right]/m$$

based on a sample $X_1, \ldots, X_m$ from p_{ω_0}. If we write $\hat{\omega}_m$ for the Monte Carlo approximation to the true maximum likelihood estimator $\hat{\omega}$, under mild regularity conditions (GEYER, 1994, Theorem 7), this Monte Carlo maximum likelihood estimator is consistent and satisfies the following central limit theorem

$$\sqrt{m}(\hat{\omega}_m - \hat{\omega}) \to \mathcal{N}(0, I(\hat{\omega})^{-1}\Sigma I(\hat{\omega})^{-1})$$

where $I(\hat{\omega}) = \mathrm{Var}_{\hat{\omega}} t(X) = -\nabla^2 l(\hat{\omega})$ denotes the Fisher information matrix at the maximum likelihood estimator, and Σ is the asymptotic covariance matrix of the normalized Monte Carlo score $\sqrt{m}\nabla l_m(\hat{\omega})$. Clearly, $I(\hat{\omega})$ can be estimated by $-\nabla^2 l_m(\hat{\omega}_m)$.

An estimator for Σ is given by

$$\frac{C_m}{\left\{\frac{1}{m}\sum_{i=1}^{m} \exp\left[t(X_i)^T(\hat{\omega}_m - \omega_0)\right]\right\}^2}.$$

where C_m is the empirical covariance matrix of $(t(\mathbf{s}) - t(X))e^{t(X)^T(\hat{\omega}_m - \omega_0)}$ based on a sample $X_1, \ldots, X_m$ from p_{ω_0}.

Importance sampling (19) relies on a reference value ω_0 that is not too far from the maximum likelihood estimator. One could use a grid of such values, with linear interpolation, or use a preliminary iteration. The Monte Carlo Newton–Raphson method (PENTTINEN, 1984) iteratively updates the parameters by

$$\omega_{k+1} = \omega_k - \nabla^2 l_m(\omega_k)^{-1}\nabla l_m(\omega_k)$$

$k = 1, 2, \ldots$, where $l_m(\cdot)$ denotes the Monte Carlo approximation to the log likelihood ratio (19) based on a sample of size m from p_{ω_k}. Since $\nabla l(\omega_k) = t(\mathbf{s}) - E_{\omega_k} t(X)$, another possibility is to set

$$\omega_{k+1} = \omega_k + \epsilon_k[t(\mathbf{s}) - t(X_k)]$$

for decreasing step sizes $\epsilon_k > 0$ and single realizations X_k from p_{ω_k}, a technique known as stochastic approximation (YOUNES, 1988, and MOYEED and BADDELEY, 1991). As k tends to infinity, under regularity conditions, ω_k approaches the maximum likelihood estimator, but no central limit theorem appears to be known for either method, although recent hybrid stochastic approximation techniques seem promising (DELYON et al., 1999, and GU and ZHU, 2001). Here we use the iterative gradient method, a variation on Newton–Raphson that guarantees convergence towards the

local optimum in the vicinity of the initial point ω_1 (CIARLET, 1994, Chapter 8, and PRESS et al., 1988, Chapter 10), i.e.

$$\begin{cases} l_m(\omega_k + \rho(\omega_k)\nabla l_m(\omega_k)) = \max_{r\in\mathbb{R}} l_m(\omega_k + r\nabla l_m(\omega_k)) \\ \omega_{k+1} = \omega_k + \rho(\omega_k)\nabla l_m(\omega_k) \end{cases} \tag{21}$$

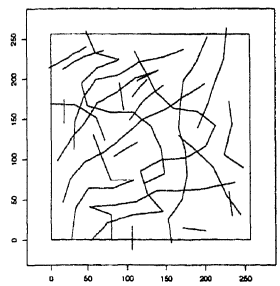
where $\rho(\omega_k)$ is computed using a one-dimensional optimization of the log likelihood ratio. With occasional re-sampling to avoid numerical instability, the following algorithm (DESCOMBES et al., 1999. and STOICA, 2001) was used.

1. *Initialize* ω_1 *and* $k = 1$;
2. *Generate a sample of size* m *from* p_{ω_k} *and compute* $\nabla l_m(\omega_k)$;
3. *For every component* $i = \{1,\ldots,5\}$ *and gradient component* $\triangle_i$, *compute the intervals* $I_k^i = [\omega_k^i - \lambda\triangle_i, \omega_k^i + \lambda\triangle_i]$ *with scalar precision parameter* $\lambda > 0$, *and maximize the log likelihood ratio in every such interval by golden section search to obtain a new value* ω_{k+1};
4. *If* $\|\omega_{k+1} - \omega_k\| > T_1$, *then* $k = k+1$ *and go to the step* 2. T_1 *is a fixed threshold;*
5. *If* $\|\nabla l_m(\omega_{k+1}) - \nabla l_m(\omega_k)\| > T_2$, *then* $k = k+1$ *and go to the step* 3, *else stop the algorithm.* T_2 *is a fixed threshold.*

5 Examples

This section is devoted to a simulation study of the Candy model, a realization of which is shown in figure 4. The parameters are given in the figure, writing $\omega_t = \log \beta$, and n_t for the total number of points. We suppress the dependence of the sufficient statistics on the realization for brevity. Throughout, the point space $K = [0, 256] \times [0, 256]$, and marks take values in $[30, 40] \times [0, \pi)$. The connection radius is $r_c = 1/\sqrt{\pi}$. The threshold values δ and τ are $0.05\ \pi$ and $0.2\ \pi$ respectively.

In our first experiment, we ran the Metropolis–Hastings algorithm defined by the kernel (15) with $p_b = 0.6$, $p_d = 0.2$, $p_c = 0.1$, $p_0 = 0.1$ and $p_k = 0.0$ from an empty

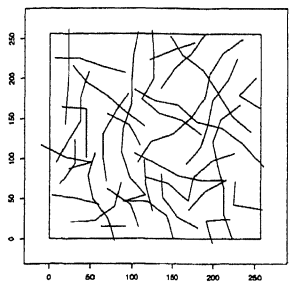


Model parameters
$\omega_t = 2.5$
$\omega_f = -11.0$
$\omega_c = -5.5$
$\omega_r = -2.5$
$\omega_o = -2.5$

Sufficient statistics
$n_t = 98$
$n_f = 7$
$n_c = 32$
$n_r = 11$
$n_o = 12$

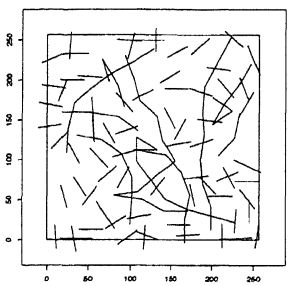
Fig. 4. Realization (left) of the Candy model with parameter values as listed in the center table. The observed values of the sufficient statistics are listed in the right hand table.

initial configuration for 2×10^7 iterations, sub sampling the sufficient statistics every 10^3 steps. The birth proposal density $b(\mathbf{s}, \eta)$ was a mixture of (3) and (8) with respective weights $p_{1b} = 0.2$ and $p_{2b} = 0.8$; for the network extension, we used the uniform laws (9) and $p_M = p_{\tilde{M}} = 0.5$. For a configuration $\mathbf{s}$ the set $A(\mathbf{s})$ was the



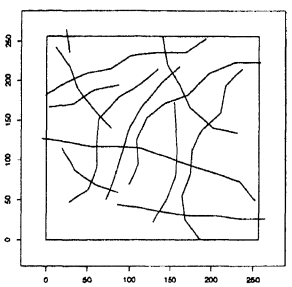
Model parameters
$\omega_t = 2.5$
$\omega_f = -11.0$
$\omega_c = -4.5$
$\omega_r = -2.5$
$\omega_o = -2.5$

Sufficient statistics
$n_t = 114$
$n_f = 1$
$n_c = 58$
$n_r = 11$
$n_o = 7$



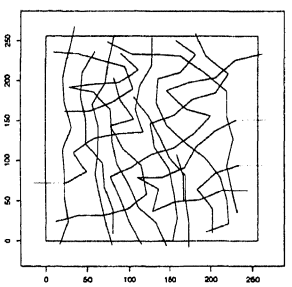
Model parameters
$\omega_t = 2.5$
$\omega_f = -7.5$
$\omega_c = -5.5$
$\omega_r = -2.5$
$\omega_o = -2.5$

Sufficient statistics
$n_t = 127$
$n_f = 68$
$n_c = 20$
$n_r = 15$
$n_o = 4$



Model parameters
$\omega_t = 2.5$
$\omega_f = -11.0$
$\omega_c = -5.5$
$\omega_r = -6.5$
$\omega_o = -6.5$

Sufficient statistics
$n_t = 72$
$n_f = 1$
$n_c = 24$
$n_r = 0$
$n_o = 0$



Model parameters
$\omega_t = 4.0$
$\omega_f = -12.5$
$\omega_c = -7.0$
$\omega_r = -2.5$
$\omega_o = -2.5$

Sufficient statistics
$n_t = 137$
$n_f = 0$
$n_c = 30$
$n_r = 14$
$n_o = 23$

Fig. 5. Realizations (left plot) of the Candy model for a range of parameter values (center table) with observed values of the sufficient statistics (right hand table).

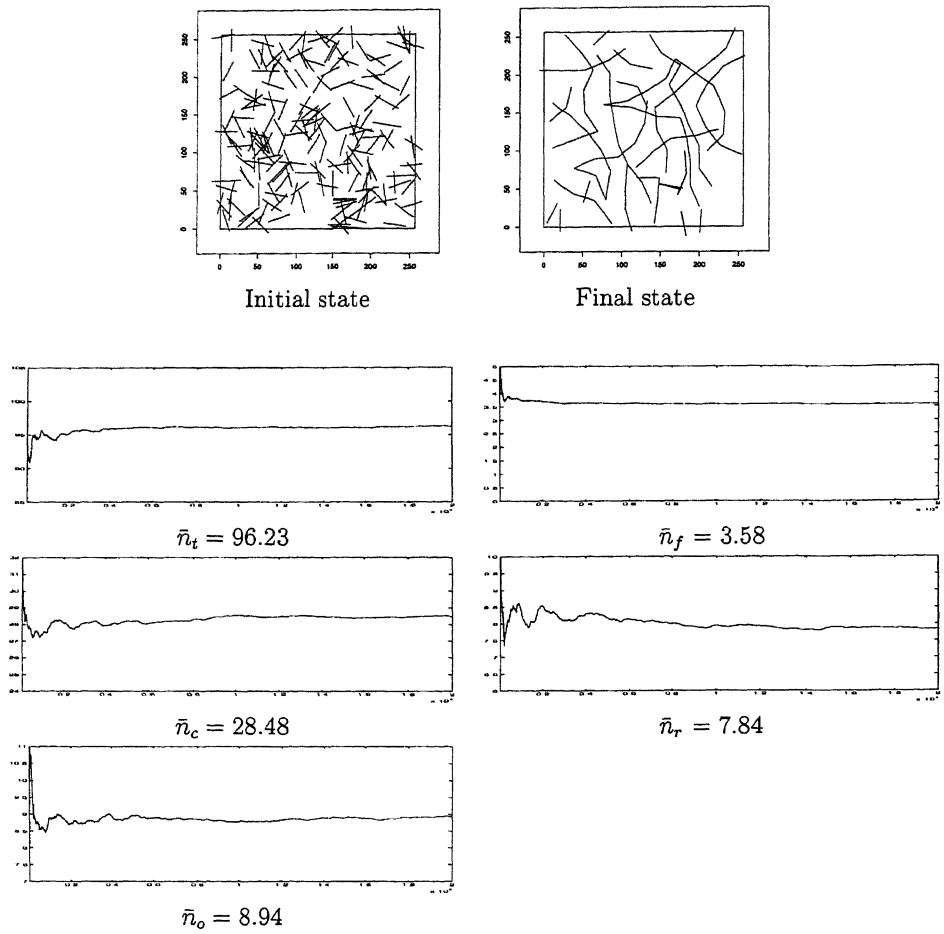


Fig. 6. Time series of the cumulative means of the sufficient statistics during a run of the Metropolis–Hastings sampler described in the text. The initial state (a realization of a binomial process of 200 segments) is shown in the top left plot, the final configuration in the top right figure.

union of all the extremities of the segments which are not connected and which are further than $\frac{1}{2}l_{\max} + r_c$ to the boundary of K.

The death proposal probabilities were as in (4). Regarding the change updates, in all cases $q(\mathbf{s}, s_i) = 1/n(\mathbf{s})$, while $c(\mathbf{s}, s_i, \eta) = b(\mathbf{s} \setminus \{s_i\}, \eta)$ and $c_\theta(\mathbf{s}, s_i, \theta)$ was as in (12) with $C_\theta(\theta_{s_i}) = [0, \pi)$.

Figure 5 gives an idea of how the topology of typical configurations depends upon the model parameters. It can be seen that the connectivity of the network can be controlled by the parameters ω_c and ω_f, the curvature by ω_o, ω_r, and the density by ω_t.

Our second experiment aimed to assess the performance of the Metropolis–Hastings algorithm by investigating the effect of the initial configuration and the various move types on the convergence speed. Figures 6 and 7 show realizations

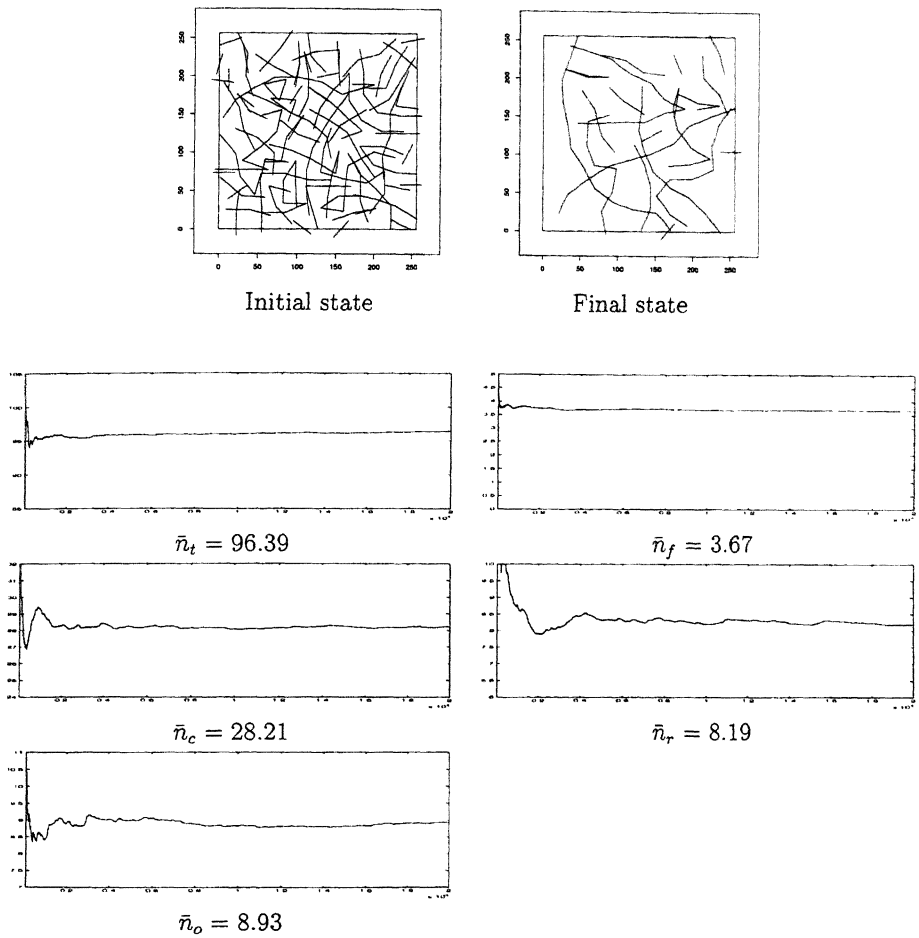


Fig. 7. Time series of the cumulative means of the sufficient statistics during a run of the Metropolis–Hastings sampler described in the text. The initial state is shown in the top left plot, the final configuration in the top right figure.

of the reference Candy model (parameters as in figure 4) obtained by the sampler described above, but initialized respectively with a realization of a binomial process consisting of 200 line segments and a random network rather than an empty configuration. To obtain the random network, we ran the Metropolis–Hastings sampler using change moves only, i.e. $p_b = p_d = p_k = 0.0$, $p_c = 0.5$, $p_\theta = 0.5$ with $q(\mathbf{s}, s_i)$, $c(\mathbf{s}, s_i, \eta)$ and $c_\theta(\mathbf{s}, s_i, \eta)$ as before and a realization of a binomial process of 200 points as the initial state. As for figure 4, we carried out 2×10^7 iterations; the sufficient statistics were sub sampled every 10^3 steps. The estimated means $\bar{n}_t, \ldots, \bar{n}_o$ of the sufficient statistics based on the three runs are close, and their evolution during the simulation does not seem to evoke doubts about convergence.

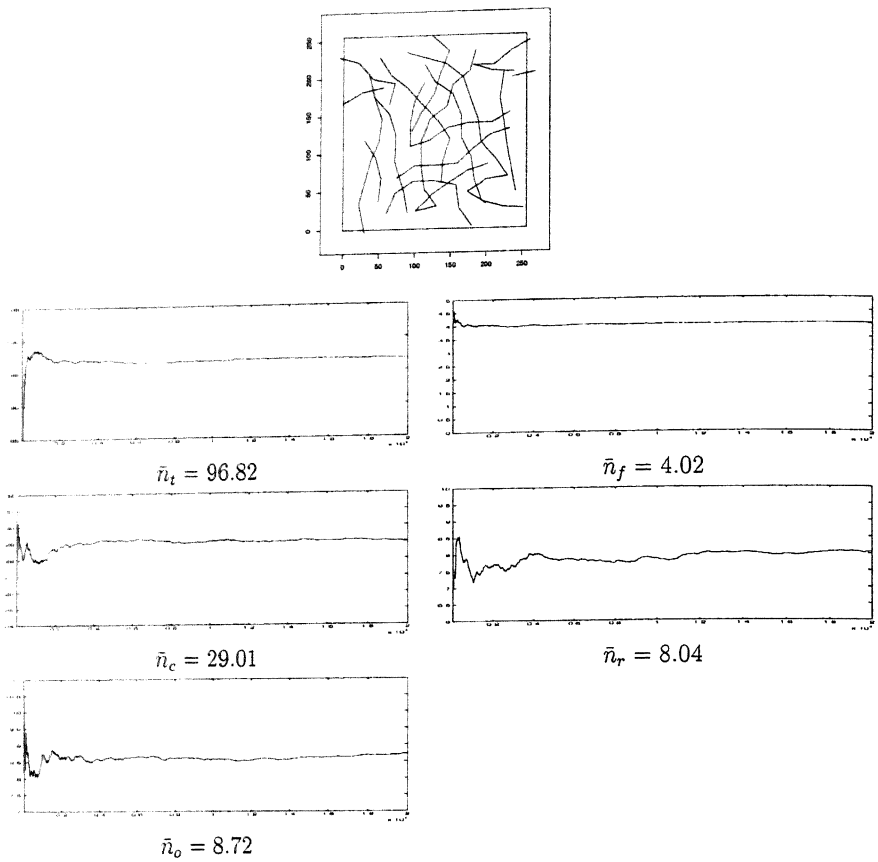


Fig. 8. Time series of the cumulative means of the sufficient statistics during a run of the Metropolis–Hastings sampler with mixture weights $p_b = 0.45$, $p_d = 0.15$, $p_c = 0.3$, $p_\theta = 0.1$. The initial state is the empty configuration, the final configuration is plotted in the top figure.

Next, we varied the mixture weights of the various moves. Figure 8 shows a realization and time series of the cumulative means for the weights $p_b = 0.45$, $p_d = 0.15$, $p_c = 0.3$, $p_\theta = 0.1$. In figure 9 the modified weights were $p_b = 0.7$, $p_d = 0.1$, $p_c = 0.1$, and $p_\theta = 0.1$. In both cases, $p_k = 0.0$, $p_{1b} = 0.2, p_{2b} = 0.8$, and $p_M = p_{\tilde{M}} = 0.5$.

The results indicate that neither the choice of initial state nor that of the mixture weights is crucial in the investigated range. However, p_{2b} should not drop so far as to effectively exclude the tailored moves, as we show in figure 10, a simulation in which only uniform birth and death moves were used (i.e. $p_b = 0.75$, $p_d = 0.25$ and $p_{1b} = 1.0$, $p_{2b} = 0.0 = p_c = p_\theta = p_k$).

From the plots, it can be observed that after a large number of iterations a connected network emerges, but that the evolution of the sufficient statistics still indicates non-stationarity, in contrast to the previous examples.

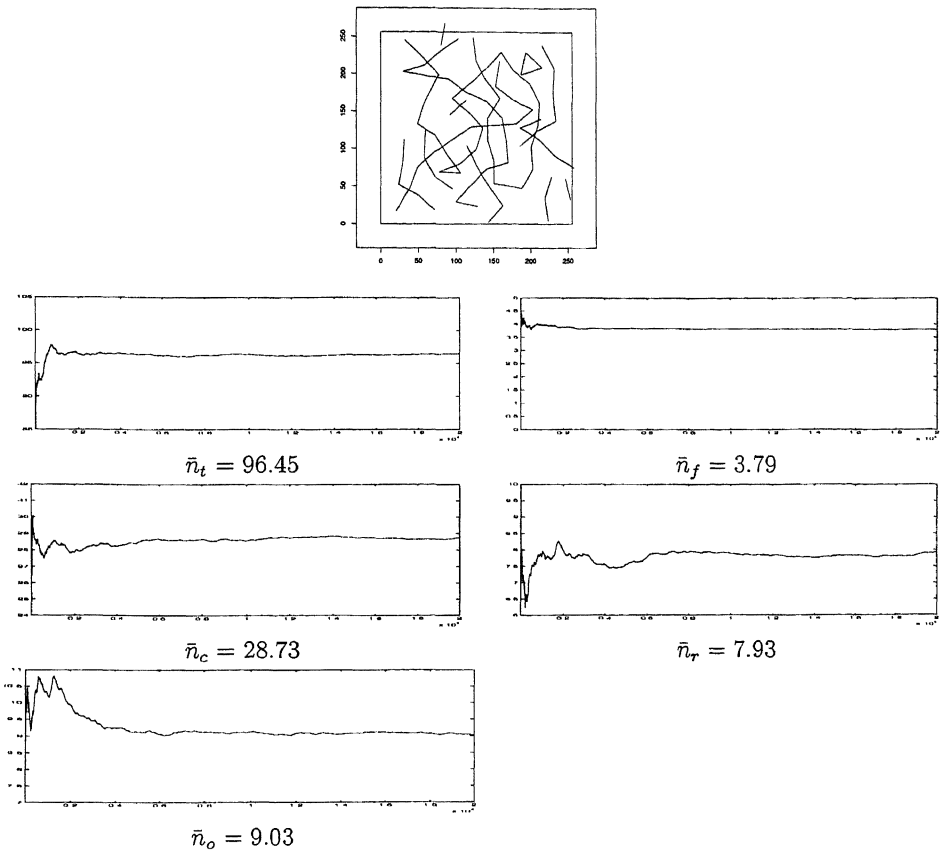


Fig. 9. Time series of the cumulative means of the sufficient statistics during a run of the Metropolis–Hastings sampler with mixture weights $p_b = 0.7$, $p_d = 0.1$, $p_c = 0.1$, $p_0 = 0.1$. The initial state is the empty configuration, the final configuration is plotted in the top figure.

To illustrate parameter estimation (section 4), suppose the data consist of the segment pattern shown in figure 4. We implemented the procedure explained in section 4, and initialized the iterative gradient algorithm (21) with arbitrary initial values listed in the first column of table 1. For the fixed thresholds $\lambda = 10^{-3}$, $T_1 = 3.0$ and $T_2 = 10^{-6}$, we obtained the output shown in table 1 (second column). Taking these values as reference parameter, we computed the Monte Carlo log likelihood ratio based on a Metropolis–Hastings run of 2×10^7 iterations, sub sampling the sufficient statistics every 10^3 steps. The weights of the various moves were the same as in the simulation of the reference model in figure 4. Cross sections of the Monte Carlo log likelihood ratio thus obtained are presented in figure 11. The maximum of $l_m(\omega)$ is located at $\hat{\omega}_m$, which vector is listed in the third column of table 1. The asymptotic standard deviation of the unknown maximum likelihood estimator $\hat{\omega}$, and the Monte Carlo standard error (MCSE) are tabulated in table 2.

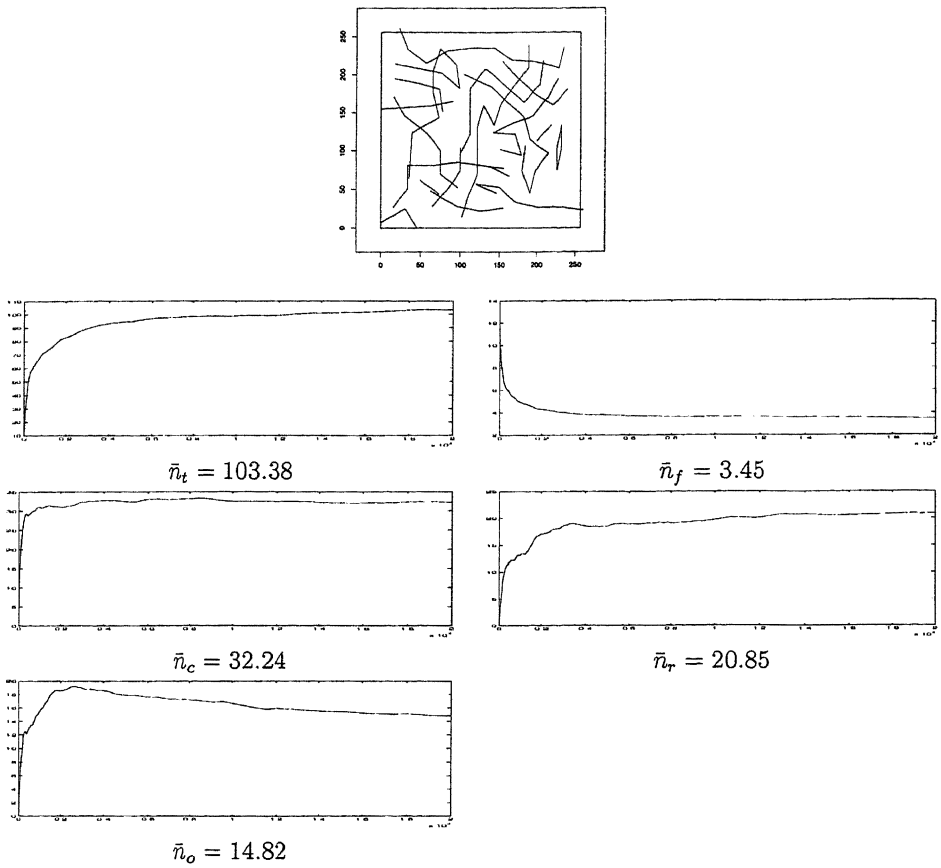


Fig. 10. Time series of the cumulative means of the sufficient statistics during a run of the Metropolis–Hastings sampler with mixture weights $p_b = 0.75$, $p_d = 0.25$ and $p_{1b} = 1.0$, $p_{2b} = p_c = p_0 = 0.0$. The initial state is the empty configuration, the final configuration is plotted in the top figure.

Table 1. Estimating the parameters for the data of figure 4.

Iterative method	Iterative method	Monte Carlo MLE
$\omega_t^i = 1.5$	$\hat{\omega}_t^0 = 2.28$	$\hat{\omega}_t^m = 2.24$
$\omega_f^i = -11.00$	$\hat{\omega}_f^0 = -10.11$	$\hat{\omega}_f^m = -10.08$
$\omega_c^i = -5.5$	$\hat{\omega}_c^0 = -5.18$	$\hat{\omega}_c^m = -5.09$
$\omega_r^i = -3.5$	$\hat{\omega}_r^0 = -2.22$	$\hat{\omega}_r^m = -2.23$
$\omega_o^i = -3.5$	$\hat{\omega}_o^0 = -2.00$	$\hat{\omega}_o^m = -2.06$

Table 2. Estimation errors.

Asymptotic standard deviation	Monte Carlo standard deviation
0.17	0.004
0.39	0.002
0.25	0.003
0.33	0.002
0.29	0.004

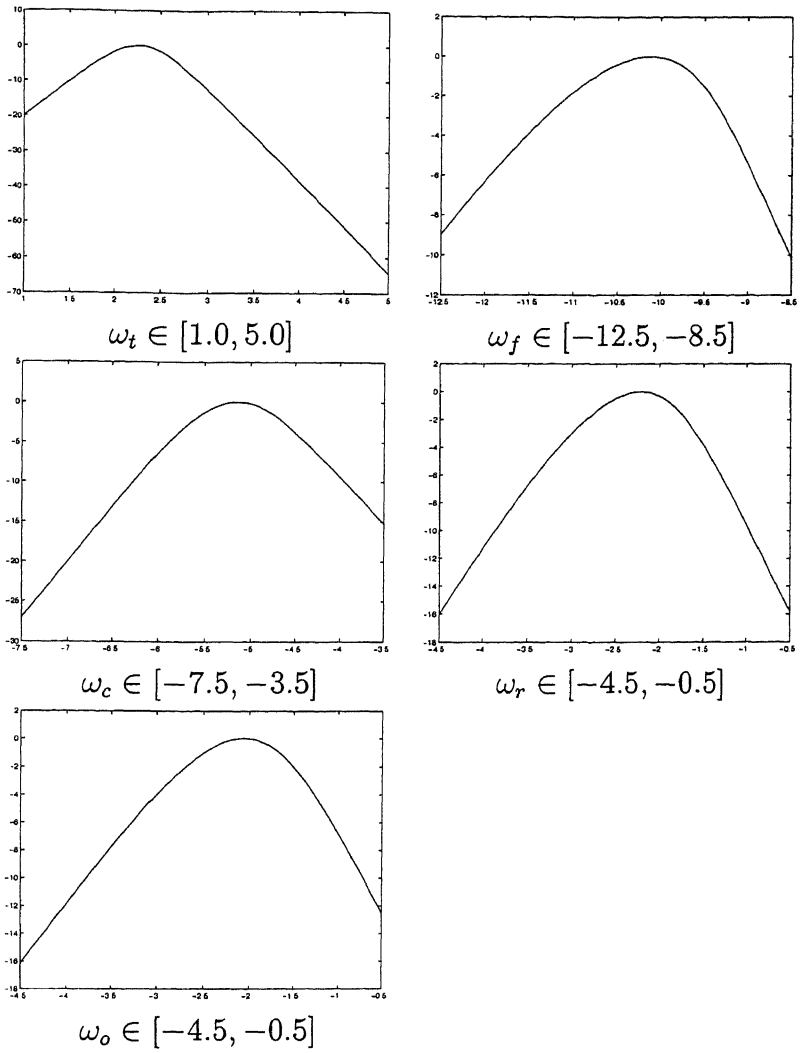


Fig. 11. Monte Carlo approximation of the log likelihood ratio for the data of figure 4. The x axis represents the variation of a single component. The y axis represents the values of the Monte Carlo log likelihood ratio with all other components of $\hat{\omega}^0$ fixed.

6 Conclusion

In the first part of this paper, we recalled the definition of the Candy model, and studied its analytical properties, concentrating on the Ruelle condition, local stability and Markovianity. The second part was devoted to statistical inference by Markov chain Monte Carlo. We suggested a variety of tailor-made updates, and proved convergence of the resulting transition kernel. Finally, we applied the sampler in a parameter estimation scheme, and performed a simulation study which shows the importance of a reasonable mix of updates that balance quick moves through the

state space with tailor made ones for fine tuning and enhancement. The relative weights of the moves may be adapted to the model parameters. Simple statistics, such as the number of free segments, converge faster than more complex ones like the average fraction of doubly connected segments.

Since the Candy model was conceived in the context of road extraction from satellite images, we expect the results presented in this paper to be a starting point in unsupervised network extraction. This can be done by adding to the Candy model a term (STOICA et al., 2000, and STOICA, 2001) which adapts the location of the road network to the data. As road density depends on geographical location, we expect to be able to improve the detection by defining a Candy model with respect to a non-homogeneous Poisson point process (STOYAN and STOYAN, 1998). Another important point is to study the feasibility of exact simulation algorithms for the Candy model.

Acknowledgements

This research was supported by NWO grant 'Inference for random sets' (613-03-045). The authors are grateful to Xavier Descombes and Josiane Zerubia for helpful discussions and to the anonymous referees for their detailed comments.

References

BADDELEY, A. J. and J. MØLLER (1989), Nearest-neighbour Markov point processes and random sets, *International Statistical Review* **57**, 89–121.

BARKER, A. A. (1965), Monte Carlo calculation of the radial distribution functions for a proton-electron plasma, *Australian Journal of Physics* **18**, 119–133.

CIARLET, P. G. (1994), *Introduction à l'analyse numérique matricielle et à l'optimisation*, Masson, Paris.

DALEY, D. J. and D. VERE-JONES (1988), *An introduction to the theory of point processes*, Springer Verlag, New York.

DELYON, B., M. LAVIELLE and E. MOULINES (1999), Convergence of a stochastic approximation version of the EM algorithm, *The Annals of Statistics* **27**, 94–128.

DESCOMBES, X., R. D. MORRIS, J. ZERUBIA and M. BERTHOD (1999), Estimation of Markov random field prior parameters using Markov chain Monte Carlo maximum likelihood, *IEEE Transactions on Image Processing* **8**, 954–963.

GATES, D. J. and M. WESTCOTT (1986), Clustering estimates for spatial point distributions with unstable potentials, *Annals of the Institute of Statistical Mathematics* **38**, 123–135.

GEYER, C. J. (1994), On the convergence of Monte Carlo maximum likelihood calculations, *Journal of the Royal Statistical Society, Series B* **56**, 261–274.

GEYER, C. J. (1999), Likelihood inference for spatial point processes, in: O. BARNDORFF-NIELSEN, W. S. KENDALL and M. N. M. VAN LIESHOUT (eds.), *Stochastic geometry, likelihood and computation*, CRC Press/Chapman and Hall, Boca Raton.

GEYER, C. J. and J. MØLLER (1994), Simulation procedures and likelihood inference for spatial point processes, *Scandinavian Journal of Statistics* **21**, 359–373.

GEYER, C. J. and E. A. THOMPSON (1992), Constrained Monte Carlo maximum likelihood for dependent data. *Journal of the Royal Statistical Society, Series B* **54**, 657–699.

GILKS, W. R., S. RICHARDSON and D. J. SPIEGELHALTER (1996), *Markov chain Monte Carlo in practice*, Chapman and Hall, London.

GRABARNIK, P. and A. SÄRKKÄ (2001), Interacting neighbour point processes: some models for clustering, *Journal of Statistical Computation and Simulation* **68**, 103–126.

GREEN, P. J. (1995), Reversible jump MCMC computation and Bayesian model determination. *Biometrika* **82**, 711–732.

GRENANDER, U. and M. I. MILLER (1994), Representations of knowledge in complex systems, *Journal of the Royal Statistical Society, Series B* **56**, 549–603.

GU, M. G. and H-T. ZHU (2001), Maximum likelihood estimation for spatial models by Markov chain Monte Carlo stochastic approximation, *Journal of the Royal Statistical Society, Series B* **63**, 339–355.

HASTINGS, W. (1970), Monte Carlo sampling methods using Markov chains and their application, *Biometrika* **57**, 97–109.

HAYAT, M. M. and J. A. GUBNER (1996), A two-step Markov point process, Technical report ECE-96-2, Department of Electrical Computer Engineering, University of Wisconsin-Madison.

KENDALL, W. S. and J. MØLLER (2000), Perfect Metropolis–Hastings simulation of locally stable spatial point processes, *Advances in Applied Probability (SGSA)* **32**, 844–865.

LIESHOUT, M. N. M. VAN (2000), *Markov point processes and their applications*, Imperial College Press/World Scientific Publishing, London/Singapore.

METROPOLIS, N., A. W. ROSENBLUTH, M. N. ROSENBLUTH, A. H. TELLER and E. TELLER (1953), Equation of state calculations by fast computing machines, *Journal of Chemical Physics* **21**, 1087–1092.

MEYN, S. P. and R. L. TWEEDIE (1993), *Markov chains and stochastic stability*, Springer-Verlag, London.

MOYEED R. A. and A. J. BADDELEY (1991), Stochastic approximation of the MLE for a spatial point pattern, *Scandinavian Journal of Statistics* **18**, 39–50.

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*, CRC Press/Chapman and Hall, Boca Raton.

OGATA, Y. and M. TANEMURA (1981), Estimation for interaction potentials of spatial point patterns through the maximum likelihood procedure, *Annals of the Institute of Statistical Mathematics* **33**, 315–338.

PENTTINEN, A. (1984), *Modelling interaction in spatial point patterns: parameter estimation by the maximum likelihood method*, Jyväskylä Studies in Computer Science, Economics and Statistics (Vol. 7), University of Jyväskylä.

PESKUN, P. (1973), Optimum Monte Carlo sampling using Markov chains, *Biometrika* **60**, 607–612.

PRESS, W. H., B. P. FLANNERY and S. A. TEUKOLSKY (1988), *Numerical recipes in C: the art of scientific computing*, Cambridge University Press.

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

REISS, R.-D. (1993), *A course on point processes*, Springer, New York.

RIPLEY, B. D. and F. P. KELLY (1977), Markov point processes, *Journal of the London Mathematical Society* **15**, 188–192.

RUE, H. and M. A. HURN (1999), Bayesian object identification, *Biometrika* **86**, 649–660.

RUE, H. and O. K. HUSBY (1998), Identification of partly destroyed objects using dynamic polygons, *Statistics and Computing* **8**, 221–228.

RUE, H. and A. R. SYVERSVEEN (1998), Bayesian object recognition with Baddeley's Delta loss, *Advances in Applied Probability (SGSA)* **30**, 64–84.

RUELLE, D. (1969), *Statistical mechanics*, Wiley, New York.

STOICA, R. (2001), *Processus ponctuels pour l'extraction des réseaux linéiques dans les images satellitaires et aériennes*. PhD thesis, University of Nice–Sophia Antipolis.

STOICA, R., X. DESCOMBES and J. ZERUBIA (2000), A Gibbs point process for road extraction in remotely sensed images, Research report 3923, INRIA Sophia Antipolis.

STOYAN, D. and H. STOYAN (1998), Non-homogeneous Gibbs process models for forestry – a case study, *Biometrical Journal* **40**, 521–531.

YOUNES, L. (1988), Estimation and annealing for Gibbsian fields, *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques* **24**, 269–294.

Received: September 2001. Revised: July 2002.