Ann Inst Stat Math (2009) 61:371–390
DOI 10.1007/s10463-007-0147-z

On a model of sequential point patterns

V. Shcherbakov

Received: 30 January 2006 / Revised: 8 November 2006 / Published online: 10 August 2007

Abstract A finite point process motivated by the cooperative sequential adsorption model is proposed. Analytical properties of the point process are considered in details. It is shown that the introduced point process is useful for modeling both aggregated and regular point patterns. A possible scheme of maximum likelihood estimation of the process parameters is briefly discussed.

Keywords Cooperative sequential adsorption model · Finite point process · Local stability · Maximum likelihood estimation · MCMC methods

1 Introduction

This paper is devoted to studying a finite point process motivated by the cooperative sequential adsorption (CSA) model. CSA models are widely used in physics and chemistry for modeling adsorption processes like chemisorption on single-crystal surfaces, adsorption in colloidal systems and other similar processes. For the physics–chemistry background and for surveys of the relevant literature we refer to Evans (1993), Privman (2000) and references therein. It should be noted that another area of applicability of such models is biological, ecological and sociological systems (Evans 1993). To imagine the phenomena one can think of some material that attracts particles from the space around. The main peculiarity of all cooperative sequential adsorption models is that adsorbed particles change adsorption properties of the material. The dependence on

V. Shcherbakov is on leave from Laboratory of Large Random Systems, Faculty of Mathematics and Mechanics, Moscow State University, Moscow.

V. Shcherbakov (✉)
Department of Mathematical Sciences, University of Bath, Bath BA2 7AY, UK
e-mail: V.Shcherbakov@bath.ac.uk

previously adsorbed particles can be modeled in various ways. A relatively simple variant is the one when the adsorption rates at a point depend on a number of previously adsorbed particles in some vicinity of the point. In lattice setting it is known in statistical physics as a model of monomer filling with cooperative effects (Evans 1993). The proposed point process directly relates to this CSA model and we call it CSA point process.

CSA point process might be useful for modeling point patterns in a particular situation where an interaction between points mimics the inter-particle interaction in physical CSA models. By this we mean point patterns formed by marked points, where a point mark is a real non-negative number interpreted as an "arrival time" and a point can interact only with those points that have lower marks. In this case it might be technically convenient to represent a point process state by a sequence of random points. The process distribution is then specified by a density with respect to a reference measure which corresponds to random point sequences formed by a Poisson's number of points, such that each point is uniformly distributed in a compact set $D \in \mathbf{R}^d$. We use here this approach following recent research by Lieshout (2006a) on sequential Markov point processes, see also Lieshout (2006b, 2006c) for more details and examples.

We prove sufficient conditions for existence of a well-defined (integrable) process density and show that in most interesting (non-degenerated) cases these conditions are also necessary. Then we briefly discuss Markov properties of the process. In particular, an explicit formula for a clique interaction function is obtained. It is interesting to note that the clique interaction functions can be expressed in terms of the well-known Pascal triangle. For simulation of the process we use the Metropolis-Hastings and spatial birth-death algorithms adopted in Lieshout (2006a) for sequential Markov point processes. To ensure the convergence of the algorithms a point process should be locally stable. In general CSA point process is not locally stable and we construct an example of that. It is proved that the process is locally stable in all practically important cases. A formula for the local stability bounds is given in these cases. Moreover, it is shown that the obtained local stability bounds cannot be improved in general.

We argue that the proposed point process provides a simple and flexible choice for modeling both aggregated and regular point patterns. The mechanism of clusters formation is explained and examples of computer simulations are given.

In Sect. 5 we describe a scheme of maximum likelihood estimation for a class of CSA point processes which seems to be useful in applications.

2 CSA point process

Let D be a compact subset of $\mathbf{R}^d$ with positive Lebesgue measure. Denote $F_D = \cup_{n=0}^{\infty} D^n$ the set of finite sequences $\mathbf{x} = (x_1, \ldots, x_n)$, $x_i \in D$, $i = 1, \ldots, n$, assuming $D^0 = \emptyset$, and let $\mathcal{F}_D$ be the standard σ-algebra on F_D. By $t(\mathbf{x})$ we denote a number of points in $\mathbf{x}$. Point $y \in D$ is called neighbor of point $x \in D$ if the distance between y and x is not greater than some fixed positive constant R. For any point $x \in D$ and sequence $\mathbf{x} = (x_1, \ldots, x_k)$ let $n(x, \mathbf{x})$ be a number of neighbors of point x in the sequence $\mathbf{x}$. By definition $n(x, \emptyset) = 0$, where $\emptyset$ is an empty sequence.

Given the sequence of points $\mathbf{x} = (x_1, \ldots, x_k)$, $k \geq 1$ we will denote for short $\mathbf{x}_{<k} = (x_1, \ldots, x_{k-1})$, for $k \geq 2$ and $\mathbf{x}_{<1} = \emptyset$.

Define a reference measure ν on $(F_D, \mathcal{F}_D)$ corresponding to random sequences of Poisson length with mean 1 and with independent uniformly distributed coordinates in D. Thus,

$$\nu(F) = \sum_{n=0}^{\infty} \frac{e^{-|D|}}{n!} \int_{D^n} 1_{\{(x_1,\ldots,x_n)\in F\}} \mathrm{d}x_1 \cdots \mathrm{d}x_n, \quad F \in \mathcal{F}_D, \tag{1}$$

where by $|D|$ we denoted the Lebesgue measure of D. Let $\{\beta_k,\ k \geq 0\}$ be a set non-negative numbers. Consider a finite point process X on D with probability distribution specified by the following density with respect to the reference measure ν

$$f(x_1, \ldots, x_n) = Z^{-1} \prod_{k=1}^{n} \beta_{n(x_k, \mathbf{x}_{<k})}, \tag{2}$$

where Z is a normalizing constant (partition function)

$$Z = \sum_{n=0}^{\infty} \frac{e^{-|D|}}{n!} \int_{D^n} \prod_{k=1}^{n} \beta_{n(x_k, \mathbf{x}_{<k})} \mathrm{d}x_1 \cdots \mathrm{d}x_n.$$

Note that, the unnormalized process density h can be written as follows

$$h(\mathbf{x}) = h(x_1, \ldots, x_n) = \prod_{k=0}^{\widehat{N}(\mathbf{x})} \beta_k^{t_k(\mathbf{x})}.$$

where

$$t_k(\mathbf{x}) = \sum_{x_i \in \mathbf{x}} 1_{\{n(x_i, \mathbf{x}_{<i}) = k\}}, \quad k \geq 0, \tag{3}$$

and

$$\widehat{N}(\mathbf{x}) = \max_{x_i \in \mathbf{x}} n(x_i, \mathbf{x}_{<i}) = \max\{k : t_k(\mathbf{x}) > 0\}. \tag{4}$$

The point process we have just defined is motivated by physical CSA lattice models where single lattice sites are filled by particles with intensities depending on states of neighboring sites. The dependence of reaction rates on existing configuration can be modeled in various ways. We have chosen a particular case when a reaction rate at a point depends on a number of the point neighbors. Certainly, other forms of dependence can be taken as well. We call the introduced point process CSA point process by analogy with statistical physics. Also, we call number R an interaction radius and the parameters $\beta_k,\ k \geq 0$ are called intensities.

2.1 Existence of CSA point process

We need to check that the partition function is finite to get a well-defined (integrable) density (2). It is easy to see that if the intensities are uniformly bounded

$$\sup_{n\geq 0} \beta_n < \infty,$$

then the partition function is finite. But if the numbers β_n, $n \geq 0$, increase sufficiently fast as $n \to \infty$, then the partition function might become infinite. Proposition 1 shows that if the domain D is not degenerate, then the linear growth of intensities is the critical one separating these two cases.

Proposition 1 1) *Assume that a non-decreasing sequence β_n, $n \geq 0$, is such that $\beta_n/n \to 0$ as $n \to \infty$, then the partition function is finite.*

2) *Assume that $\beta_n = \beta(n+1)$, $n \geq 0$, for some $\beta > 0$ and D is a domain with non-empty interior. If β is sufficiently large, then the partition function is infinite.*

Proof 1). Fix $0 < \varepsilon < |D|^{-1}$ and let $n(\varepsilon)$ be such that $\beta_n/n \leq \varepsilon$ for any $n > n(\varepsilon)$. Represent the partition function as follows

$$Z = \sum_{n\leq n(\varepsilon)} \frac{e^{-|D|}}{n!} \int_{D^n} \prod_{k=1}^{n} \beta_{n(x_k,\mathbf{x}_{<k})} \mathrm{d}x_1 \cdots \mathrm{d}x_n + \sum_{n>n(\varepsilon)} \frac{e^{-|D|}}{n!} \int_{D^n} \prod_{k=1}^{n} \beta_{n(x_k,\mathbf{x}_{<k})} \mathrm{d}x_1 \cdots \mathrm{d}x_n.$$

The first term in the preceding display is finite. Let us bound the second term. Note that $n(x_k, \mathbf{x}_{<k}) \leq k-1$ for any $k \geq 1$. Hence $\beta_{n(x_k,\mathbf{x}_{<k})} \leq \beta_{k-1}$, since the sequence β_n, $n \geq 0$, is non-decreasing. Therefore the product in the integral can be bounded as follows

$$\prod_{k=1}^{n} \beta_{n(x_k,\mathbf{x}_{<k})} = \prod_{k=1}^{n(\varepsilon)} \beta_{n(x_k,\mathbf{x}_{<k})} \prod_{k=n(\varepsilon)+1}^{n} \beta_{n(x_k,\mathbf{x}_{<k})} \leq \beta_{n(\varepsilon)}^{n(\varepsilon)} \frac{\varepsilon^{n-n(\varepsilon)} n!}{n(\varepsilon)!},$$

where we have used that $\beta_k \leq \beta_{n(\varepsilon)}$, if $k \leq n(\varepsilon)$. Using the bound in the preceding display we obtain

$$\sum_{n\geq n(\varepsilon)} \frac{e^{-|D|}}{n!} \int_{D^n} \prod_{k=1}^{n} \beta_{n(x_k,\mathbf{x}_{<k})} \mathrm{d}x_1 \cdots \mathrm{d}x_n \leq \left(\frac{\beta_{n(\varepsilon)}}{\varepsilon}\right)^{n(\varepsilon)} \frac{e^{-|D|}}{n(\varepsilon)!} \sum_{n\geq n(\varepsilon)} (\varepsilon|D|)^n,$$

and the right hand side is finite by choice of ε. Therefore $Z < \infty$.

2). In this case D contains a ball $B(y,r)$ of positive radius $r < R/2$, so one can bound

$$Z \geq \sum_{n=0}^{\infty} \frac{e^{-|D|}\beta^n}{n!} \int_{B^n(y,r)} \prod_{k=1}^{n} n(x_k, \mathbf{x}_{<k}) \mathrm{d}x_1 \cdots \mathrm{d}x_n \geq e^{-|D|} \sum_{n=0}^{\infty} |B(y,r)|^n \beta^n,$$

and the last series is infinite provided that $|B(y,r)|\beta \geq 1$. □

2.2 Examples of CSA point processes

A wide class of CSA point processes that might be interesting in applications is formed by CSA processes whose densities are specified by a finite number of different intensities, i.e., β_k, $k \leq N$ and $\beta_k = \gamma \geq 0$, $k > N$ for some $N \geq 0$. The unnormalized process density in this case is

$$h(x_1, \ldots, x_n) = \gamma^{s(\mathbf{x})} \prod_{k=0}^{N} \beta_k^{t_k(\mathbf{x})}, \tag{5}$$

where

$$s(\mathbf{x}) = t(\mathbf{x}) - \sum_{k=0}^{N} t_k(\mathbf{x}), \tag{6}$$

is the total number of points that have at least $N+1$ neighbors.

An important subclass of the CSA point processes with a finite number of different intensities is formed by CSA processes with *hard core interaction*. The last means that only a finite number of intensities $\beta_0, \ldots, \beta_N$ is non-zero. The unnormalized process density in this case is

$$h(x_1, \ldots, x_n) = 1_{\{\widehat{N}(\mathbf{x}) \leq N\}} \prod_{k=0}^{N} \beta_k^{t_k(\mathbf{x})}. \tag{7}$$

2.3 Markov property of CSA point process

It is easy to see that $f(\mathbf{y}) > 0$ implies $f(\mathbf{z}) > 0$ for all subsequences $\mathbf{z} \subset \mathbf{y}$ and that for all sequences $\mathbf{y}$ such that $f(\mathbf{y}) > 0$, the ratio

$$\frac{f((\mathbf{y}, u))}{f(\mathbf{y})} = \lambda(u|\mathbf{y}) = \beta_{n(u,\mathbf{y})} \tag{8}$$

depends only on its neighbors in sequence $\mathbf{y}$. According to the definition in Lieshout (2006a), the point process X defined by the density (2) is a sequential Markov point process with respect to the following neighborhood relation $\sim$ between points: $y \sim x$,

if Euclidean distance between x and y is not greater than R. Similar to the case of classical Markov point processes Hammersley–Clifford factorization formula holds for density f of a sequential Markov point process (see Lieshout 2006a)

$$f(\mathbf{x}) = \prod_{j=1}^{t(\mathbf{x})} \prod_{\mathbf{y} \subseteq \mathbf{x}_{<j}} \varphi(x_j, \mathbf{y}), \tag{9}$$

where the inner products are over all x_j-cliques $\mathbf{y}$, that is all components $y \in \mathbf{y}$ satisfy $x_j \sim y$, and $\varphi : D \times F_D \to \mathbf{R}_+$ is some function, such that $\varphi(x_j, \mathbf{y}) = 1$, if $\mathbf{y}$ is not an x_j-clique. The function $\varphi(x_j, \mathbf{y})$ is called a clique interaction function.

Knowing the clique interaction function is necessary if one needs the conditional distribution of a Markov point process restricted on a subset $D' \subset D$ given the process configuration in $D \setminus D'$. Computing a formula for the clique interaction function might be difficult, see, for example, Baddeley et al. (1996). Therefore it is interesting to note that a clique interaction function can be computed explicitly in the case of CSA point process.

Theorem 1 *The clique interaction function $\varphi(x, \mathbf{y})$ of CSA point process is defined as follows*

$$\varphi(x, \mathbf{y}) = \prod_{m=0}^{t(\mathbf{y})} \beta_m^{C(n(\mathbf{y}),m)}, \tag{10}$$

if $\mathbf{y} = (y_1, \ldots, y_n)$ is an x-clique, where

$$C(n, m) = (-1)^{n+m} \binom{n}{m}, \quad n \geq m \geq 0. \tag{11}$$

Proof Denote $\psi_n = \log(\varphi(x, \{y_1, \ldots, y_n\}))$, $n \geq 0$, where $n = t(\mathbf{y})$ is the number of points in the x-clique $\mathbf{y}$. Using (8) and (9), one can get that $\psi_0 = \log(\beta_0)$ and

$$\psi_n = \log(\beta_n) - \sum_{k=0}^{n-1} \binom{n}{k} \psi_k, \quad n \geq 1.$$

It can be shown by induction that

$$\psi_n = \sum_{m=0}^{n} C(n, m) a_m, \tag{12}$$

where $a_m = \log(\beta_m)$, $m = 0, 1, \ldots, n$ and

$$C(n, m) = \binom{n}{m} \sum_{k=1}^{n-m} (-1)^k k! S(n - m, k), \tag{13}$$

and $S(n,m)$, $n \geq 1, 1 \leq m \leq n$, are the so-called Stirling numbers of the second kind, i.e., the number $S(i,j)$ is the number of ways to partition a set of i objects into j non-empty groups (Comtet 1974). It remains to note that for the Stirling numbers of the second kind the following classical identity holds

$$\sum_{k=1}^{M}(-1)^k k!S(M,k) = (-1)^M, \tag{14}$$

for any $M \geq 1$ (Abramowitz and Stegun 1972, Sect. 24.1.4, p. 825). The theorem is proved. □

Remark 1 It is easy to see that the coefficients $C(n,m)$ form a table which is nothing else but the alternating Pascal triangle.

3 Local stability and simulation of CSA point processes

To simulate CSA point process we use Metropolis–Hastings and birth–death samplers adopted for the sequential Markov point processes in Lieshout (2006a). For the reader's convenience we briefly recall the basic choices for both of them and their properties. Consider a sequential Markov process given by its density $g(\mathbf{x})$ with respect to the reference measure ν. The generic variant of the Metropolis-Hastings sampler is the following one. Given the current state $\mathbf{x}$ the birth and death are proposed equally likely. If birth is proposed, then we choose a position $i \in \{0,\ldots,t(\mathbf{x})\}$ at random, sample a uniformly distributed point $\xi \in D$ and accept the new state $s_i(\mathbf{x},\xi) = (\ldots, x_{i-1}, \xi, x_i, \ldots)$ with probability

$$\min\left\{1, \frac{g(s_i(\mathbf{x},\xi))}{g(\mathbf{x})(t(\mathbf{x})+1)}\right\}. \tag{15}$$

If death is proposed, then a position $i \in \{1,\ldots,t(\mathbf{x})\}$ is chosen at random and the point x_i is removed with probability

$$\min\left\{1, \frac{g(x_1,\ldots,x_{i-1},x_{i+1},\ldots,x_{n(\mathbf{x})})t(\mathbf{x})}{g(\mathbf{x})|D|}\right\}. \tag{16}$$

The generic choice of the birth-death sampler is determined by the birth rate

$$b_i(\mathbf{x},\xi) = \frac{g(s_i(\mathbf{x},\xi))}{(t(\mathbf{x})+1)g(\mathbf{x})}, \tag{17}$$

for the birth of a new point ξ at position i given the current configuration $\mathbf{x}$, and $d_i(\mathbf{x}) = 1$ for the death rate at position i in the current configuration $\mathbf{x}$.

Note that the Markov chains generated by both Metropolis–Hastings and birth–death algorithm are reversible with respect to the probability distribution specified by the density $g(\mathbf{x})$. Therefore this distribution is an invariant one for the Markov chains.

A sufficient condition for geometric ergodicity for both the Metropolis–Hastings and birth-and-death samplers for a sequential Markov process with density g is the following local stability condition

$$\frac{g(s_i(\mathbf{x}, \xi))}{g(\mathbf{x})} \le C, \tag{18}$$

uniformly in $\xi \in D$, $\mathbf{x} = (x_1, \ldots, x_n)$, $g(\mathbf{x}) > 0$, and $i = 0, \ldots, n$ for some $C > 0$.

3.1 Local stability of CSA point process

The next theorem provides us with local stability bounds for CSA point processes mentioned in Sect. 2.2.

Theorem 2 *Consider a CSA point process with a finite number of different intensities.*

1) *If $\beta_k > 0$, $k = 0, 1, \ldots, N$, $\beta_k = 0$, $k > N$ for some N, then the density (2) is locally stable with a constant*

$$C_1(N, \beta) = \max(\beta_0, \ldots, \beta_N) \left(\frac{\max(\beta_0, \ldots, \beta_N)}{\min(\beta_0, \ldots, \beta_N)} \right)^{T(d)N}, \tag{19}$$

where $T(d)$ is the maximal number of points which can be allocated on d-dimensional sphere with unit radius in such a way that the distance between any two of them is not less than 1.

2) *If $\beta_k > 0$, $k = 0, 1, \ldots, N$, $\beta_k = \gamma > 0$, $k > N$ for some N and γ, then the density (2) is locally stable with a constant*

$$C_2(N, \beta) = \max(\beta_0, \ldots, \beta_N, \gamma) \left(\frac{\max(\beta_0, \ldots, \beta_N, \gamma)}{\min(\beta_0, \ldots, \beta_N)} \right)^{T(d)N}, \tag{20}$$

where $T(d)$ is the constant defined in part 1).

Proof Let us prove part 1) of Theorem 2. We assume that $d = 2$, for higher dimensions the modification is obvious. Note that $T(2) = 6$. Denote for short

$$r_i(\mathbf{x}, \xi) = \beta_{n(\xi, \mathbf{x}_{<i})} \prod_{k=i}^{n} \frac{\beta_{n(x_k, (\mathbf{x}_{<k}, \xi))}}{\beta_{n(x_k, \mathbf{x}_{<k})}}. \tag{21}$$

For any ξ and $\mathbf{x}_{<i}$ we have that $\beta_{n(\xi, \mathbf{x}_{<i})} \le \max(\beta_0, \ldots, \beta_N)$. Let us bound the product. Note that $n(x_k, (\mathbf{x}_{<k}, \xi)) - n(x_k, \mathbf{x}_{<k})$ can be 0 or 1. If $n(x_k, \mathbf{x}_{<k}) = 0$ and $\|\xi - x_k\| > R$, then $n(x_k, (\mathbf{x}_{<k}, \xi)) - n(x_k, \mathbf{x}_{<k}) = 0$ and the corresponding ratio in the product is

$$\frac{\beta_{n(x_k, (\mathbf{x}_{<k}, \xi))}}{\beta_{n(x_k, \mathbf{x}_{<k})}} = 1.$$

If $\|\xi - x_k\| \le R$, but $n(x_k, \mathbf{x}_{<k}) = N$, then $n(x_k, (\mathbf{x}_{<k}, \xi)) = N + 1$ and the corresponding ratio is

$$\frac{\beta_{N+1}}{\beta_N} = 0,$$

since $\beta_{N+1} = 0$. So, the point ξ can be inserted as the ith coordinate of the vector $\mathbf{x}$ if and only if there are no points in the configuration $\mathbf{x} \cap B(\xi, R)$ that already have N neighbors. If $n(x_k, \mathbf{x}_{<k}) < N$ and $\|\xi - x_k\| \le R$, then the product gains the factor $\beta_{n(x_k,\mathbf{x}_{<k})+1}/\beta_{n(x_k,\mathbf{x}_{<k})}$ that we bound by $\max(\beta_0, \ldots, \beta_N)/\min(\beta_0, \ldots, \beta_N)$. If $\beta_{j+1}/\beta_j \le 1$ for any $j = 0, \ldots, N-1$, then the bound (18) is obvious and the proof is finished. In the general case we bound

$$r_i(\mathbf{x}, \xi) \le \max(\beta_0, \ldots, \beta_N) \left(\frac{\max(\beta_0, \ldots, \beta_N)}{\min(\beta_0, \ldots, \beta_N)} \right)^{m_i(\xi,\mathbf{x})},$$

where $m_i(\mathbf{x}, \xi)$ is the number of $\mathbf{x}$ components x_k, $k \ge i$, such that $n(x_k, \mathbf{x}_{<k}) < N$ and $\|\xi - x_k\| \le R$. A simple geometric argument shows that we can bound $m_i(\mathbf{x}, \xi) \le 6N$. Indeed, represent the ball $B(\xi, R)$ as a union of 6 equal disjoint sectors. It is easy to see that the distance between any two points belonging to the same sector is not more than R. Therefore there cannot be more than N points of the vector $\mathbf{x}$ in any of these sectors. Indeed, assume there was a sector that contains at least $N + 1$ points of the vector $\mathbf{x} = (x_1, \ldots, x_n)$, $n \ge N + 1$. Let x_i be the last arrival in this sector, i.e., it is the point that appeared there later than the others. Then x_i has already (at least) N neighbors (other vector points belonging to the sector) and it would have at least $N + 1$ neighbors if we inserted ξ at the beginning of the vector but this situation is forbidden. So, finally

$$r_i(\mathbf{x}, \xi) \le \max(\beta_0, \ldots, \beta_N) \left(\frac{\max(\beta_0, \ldots, \beta_N)}{\min(\beta_0, \ldots, \beta_N)} \right)^{6N}.$$

Part 1) of the theorem is proved.

Let us prove part 2) of Theorem 2. Assume again for definiteness and without loss of generality that $d = 2$. For any ξ and $\mathbf{x}_{<i}$ one can estimate

$$\beta_{n(\xi,\mathbf{x}_{<i})} \le \max(\beta_0, \ldots, \beta_N, \gamma).$$

Bounding the product (21) we consider two cases. If $\|\xi - x_k\| \le R$ and $n(x_k, \mathbf{x}_{<k}) > N$, then the corresponding ratio is trivial

$$\frac{\beta_{n(x_k,\mathbf{x}_{<k})}}{\beta_{n(x_k,(\mathbf{x}_{<k},\xi))}} = 1.$$

If $\|\xi - x_k\| \le R$, $n(x_k, (\mathbf{x}_{<k}, \xi)) - n(x_k, \mathbf{x}_{<k}) = 1$ and $n(x_k, \mathbf{x}_{<k}) \le N$, then the product gains the factor that might be not equal to 1 and we bound it as follows

$$\frac{\beta_{n(x_k,\mathbf{x}_{<k})+1}}{\beta_{n(x_k,\mathbf{x}_{<k})}} \le \frac{\max(\beta_0, \ldots, \beta_N, \gamma)}{\min(\beta_0, \ldots, \beta_N)}.$$

So

$$r_i(\mathbf{x}, \xi) \le \max(\beta_0, \ldots, \beta_N, \gamma) \left(\frac{\max(\beta_0, \ldots, \beta_N, \gamma)}{\min(\beta_0, \ldots, \beta_N)} \right)^{l_i(\xi,\mathbf{x})},$$

where $l_i(\mathbf{x}, \xi)$ is the number of components $x_k \in \mathbf{x}$, $k \ge i$, such that $n(x_k, \mathbf{x}_{<k}) \le N$ and $\|\xi - x_k\| \le R$. The geometric argument we used in the proof of part 1) gives that $l_i(\mathbf{x}, \xi) \le 6N$, so we get the bound (20). The theorem is proved. □

Remark 2 It is well known that the practical implementation of birth-and-death samplers in the locally stable case is based on dependent thinning of transitions of a Poisson process with unit death rate and constant birth rate C that dominates any $r_i(\mathbf{x}, \xi)$: $r_i(\mathbf{x}, \xi) \le C$. If C is overestimated, then it leads to very low acceptance probabilities. Therefore it is desirable to get the bound (18) as small as possible. From this point of view the bounds we have obtained in Theorem 2 are not satisfactory. We obviously overestimated the values of $m_i(\xi, \mathbf{x})$ and $l_i(\xi, \mathbf{x})$, since the points of any sector might have neighbors from other sectors. But in general both bounds (19) and (20) are close to optimal. Indeed, consider at first the bound (19) in the case when $\beta_0, \beta_1 > 0$ and $\beta_k = 0$ for $k \ge 2$. Then following the proof, one gets that the bound for $m_i(\xi, \mathbf{x})$ (the maximal number of points of the configuration $\mathbf{x}$ which can contribute the non-trivial factor in the product (21)) is 6. One would get this number if the configuration shown in Fig. 1 occurred, where 6 points of the vector $\mathbf{x}$ are located at the boundary of the ball $B(\xi, R)$ such that the distance between the nearest of them is R. Obviously, this set of configurations has probability 0, since the ball boundary has Lebesgue measure 0 in dimension 2. But the set of configurations shown in Fig. 2 has already a non-zero probability and gives the value $m_i(\xi, \mathbf{x}) = 5$.

Consider now the process with intensities $\beta_0 > 0, \beta_1 > 0, \beta_k = \gamma > 0, \ k \ge 2$. It is easy to see that the same "ideal" configuration in Fig. 1 gives the bound (20) with $T(2) = 6$ and the configuration in Fig. 2 provides the "lower bound" for the bound (20) and it is 5 since $l_i(\xi, \mathbf{x}) = 5$ in this case. The only difference is that now the points of configuration $\mathbf{x}$ which we locate near the ball boundary, cannot have more than 1 neighbor outside the ball (otherwise the corresponding factor in the product (21) would be trivial, i.e., equal to 1).

The obtained bounds can be very large even for small N. For instance, if $N = 2$ and $\max(\beta_0, \ldots, \beta_N)/\min(\beta_0, \ldots, \beta_N) = 10$, then the bound is of order 10^{12}. Therefore we can conclude that the birth-and-death sampler based on thinning is not a good practical choice for simulating CSA point processes.

Remark 3 It is easy to see that the condition (18) holds for any decreasing sequence of intensities $\beta_0 \ge \cdots \ge \beta_k \ge \beta_{k+1} \ge \cdots$, $k \ge 0$, which is not necessarily finite, since in this case the factor $\beta_{n(x_k,(\mathbf{x}_{<k},\xi))}/\beta_{n(x_k,\mathbf{x}_{<k})}$ is always bounded by 1.

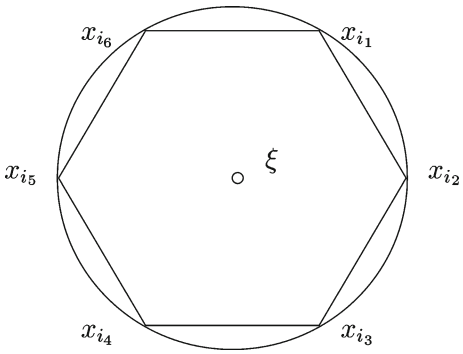

Fig. 1 Six points at the surface with the minimal distance R between them, R is radius of the ball. It is assumed that the points do not have any neighbors outside the ball

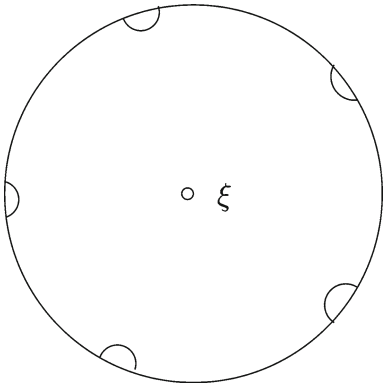

Fig. 2 Five domains with positive measures near the ball boundary. The distances between the domains are not less than the radius of the ball. A vector $\mathbf{x}$ has a point of intersection with each of the domains, these points do not have any neighbors outside the ball

Remark 4 If the set of intensities is unbounded, then the density (2) may not be locally stable. Indeed, assume that $\beta_i \to \infty$ as $i \to \infty$. Let the area D contain at least one ball $B(y, r)$ with radius $r < R/2$. Consider a configuration $\mathbf{x} = (x_1, \ldots, x_n) \in B(y, r)$. Then, for $\xi \in B(y, r)$

$$r_0(\mathbf{x}, \xi) = \beta_{n(\xi,\emptyset)} \prod_{k=1}^{n} \frac{\beta_{n(x_k,(\mathbf{x}_{<k},\xi))}}{\beta_{n(x_k,\mathbf{x}_{<k})}} = \beta_0 \prod_{k=0}^{n-1} \frac{\beta_{k+1}}{\beta_k} = \beta_n,$$

which is unbounded in n.

4 Modeling repulsion and attraction in point patterns by CSA point processes

In physical CSA models reaction rates might be either enhanced or inhibited by the reacted neighbors. The former leads to clustering and the latter leads to repulsion between particles. We are going to demonstrate that these properties are inherited by CSA point process. An intuitive picture that one may expect is that if $\beta_0 \geq \cdots \geq \beta_k \geq \beta_{k+1} \geq \cdots$, then it might result in repulsion between points, and if $\cdots \leq \beta_k \leq \beta_{k+1} \leq \cdots$, then it is reasonable to expect clustering in a point pattern. Basically, these predictions hold though there are certain limitations which will be mentioned. All images

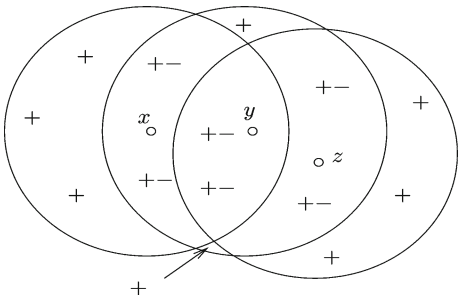

Fig. 3 Formation of the "linear cluster"

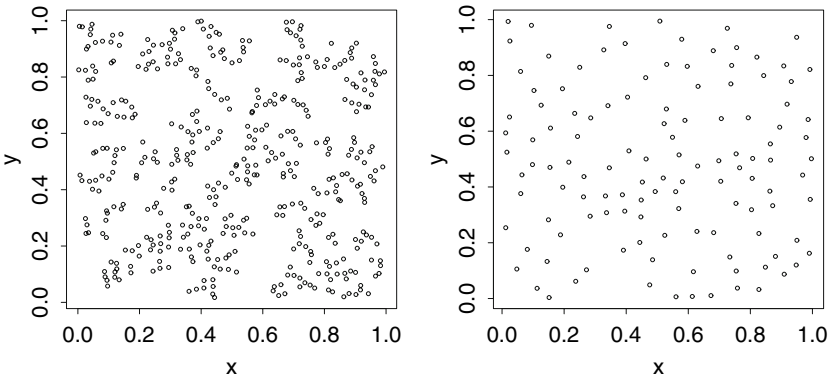

Fig. 4 CSA patterns with two non-zero intensities. *Left*: $R = 0.03$ and $\beta_0 = 10, \beta_1 = 2{,}000$. *Right*: $R = 0.05$ and $\beta_0 = 500,\ \beta_1 = 5$. Statistics $(t_0, t_1) = (76, 367)$ and $(t_0, t_1) = (121, 1)$ respectively

we refer in this section were obtained by the Metropolis-Hastings algorithm. Every image is supplied with the corresponding process parameters $R, \beta_k,\ k = 0, \ldots, N$, and with values of statistics $t_k,\ k = 0, \ldots, N$ (see (3) for the definition). As an example we consider the simplest non-trivial situation, when there are only two non-zero intensities β_0 and β_1, i.e., a point can have at most one neighbor.

If $\beta_0 \gg \beta_1$ and β_1 is small, then patterns with repulsion between points are observed, e.g., the right image provided by Fig. 4. Such patterns are similar to patterns produced by a CSA point process with the same intensity β_0 and $\beta_k = 0,\ k \geq 1$.

If $\beta_0 \ll \beta_1$, then aggregated point patterns appear, e.g., the left image in Fig. 4. It is visible that the point clusters are formed by "directed local chains" of points which approximately have "a local linear structure". Figure 3 illustrates the mechanism of a chain formation. Areas labeled by $+$ are the areas where the intensity of point appearance is high, $+-$ areas are the ones which were active but are blocked at the moment. Point x is the first one, it activates the area in the ball $B(x, R)$. Then point y appears and immediately the area $B(x, R) \cap B(y, R)$ is frozen. Point z is the third one, its appearance blocks the intersection $B(z, R) \cap B(y, R)$.

It is important to note that if β_1 is taken sufficiently large, then it might spoil the aggregated structure, resulting in high density homogeneous point patterns formed by pairs of close points. This effect might occur even if β_0 is rather small. Without going

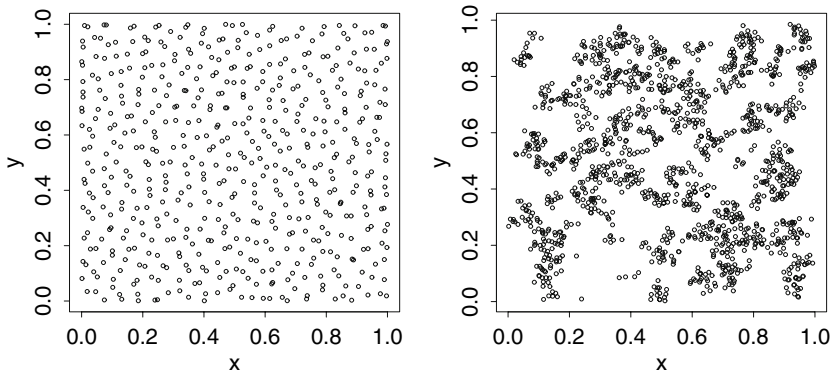

Fig. 5 *Left*: example of degenerated cluster structure, $R = 0.05$, $\beta_0 = 0.0005$, $\beta_1 = 500{,}000$, $\beta_k = 0$, $k \geq 2$, statistics $(t_0, t_1) = (1{,}544)$. *Right*: $R = 0.03, \beta_0 = 10, \beta_1 = 510, \beta_2 = 1{,}010, \beta_3 = 1{,}510, \beta_4 = 2{,}010, \beta_5 = 2{,}510, \beta_6 = 3{,}010, \beta_7 = 3{,}510, \beta_8 = 4{,}010, \beta_9 = 4{,}510$, statistics $(t_0, \ldots, t_9) = (79, 254, 236, 179, 146, 134, 111, 86, 72, 43)$

into details we briefly sketch an idea of how this effect can be explained analytically. It is easy to see that in the case of two non-zero intensities the partition function Z is just a finite linear combination of terms $\beta_0^i \beta_1^j$, $i > 0$. Let $j_{\max}$ be a maximal power of β_1 in the expansion for the partition function. If $\beta_0 \beta_1^{j_{\max}} \to 0$ as $\beta_0 \to 0$, $\beta_1 \to \infty$, then obviously the limit process is trivial, i.e., its distribution is concentrated on the empty configuration. But if $\beta_0 \beta_1^{j_{\max}} \to \infty$, or is just bounded below, and $\beta_0 \beta_1^{j_{\max}-1} \to 0$ as $\beta_0 \to 0$, $\beta_1 \to \infty$, then the limit process is non-trivial and its distribution is concentrated on the dense packing configurations $\mathbf{x}$ (no more points can be placed) with $t_0(\mathbf{x}) = 1$. Apparently the left image in Fig. 5 illustrates this effect.

Similar mechanism of cluster formation works in the case of any CSA point process with a finite number of non-zero intensities and the same mixed attraction-repulsion effects can be observed. An image corresponding to a CSA point process with 10 non-zero intensities is provided by the right panel of Fig. 5. The aggregated structure of this image is apparently stipulated by sufficiently small value of the interaction radius and substantial growth of intensities.

Another easy to handle choice for modeling cluster structures is provided by CSA point processes with a finite number of different intensities in case when all of them are positive. The mechanism of cluster formation is the same as the one explained above. The only difference is that there are no restrictions on the number of neighbors in this situation. Density and sizes of clusters can be controlled by making at least one of the intensities smaller or bigger. By varying the interaction radius one can also reduce or intensify aggregating effects in point patterns. For example, compare the images provided in Fig. 6. The aggregated structure can be lost if intensities β_k, $k \geq 1$, are taken too large. For instance, the images in Fig. 6 have visible cluster structure, point clusters are separated one from each other. In the left image of Fig. 7 the clusters begin coalescing though they still can be recognized. The right image in Fig. 7 gives an example when clusters have coalesced forming one "connected component".

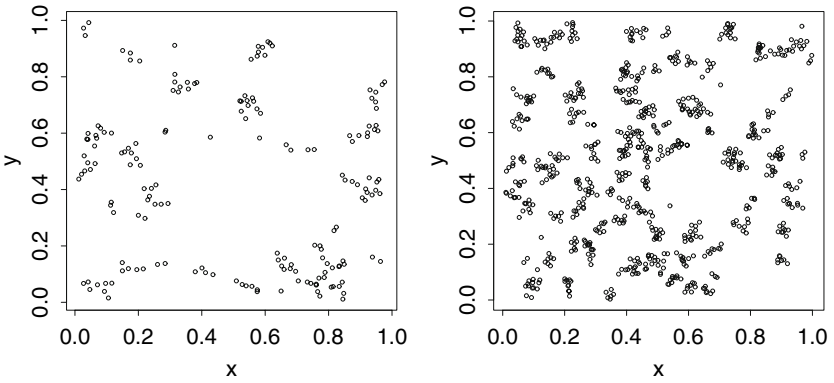

Fig. 6 *Left*: $R = 0.03$ and $\beta_0 = 5$, $\beta_1 = 1{,}000$, $\beta_k = 500$, $k \geq 2$. *Right*: $R = 0.02$ and $\beta_0 = 10$, $\beta_k = 2{,}000$, $k \geq 1$

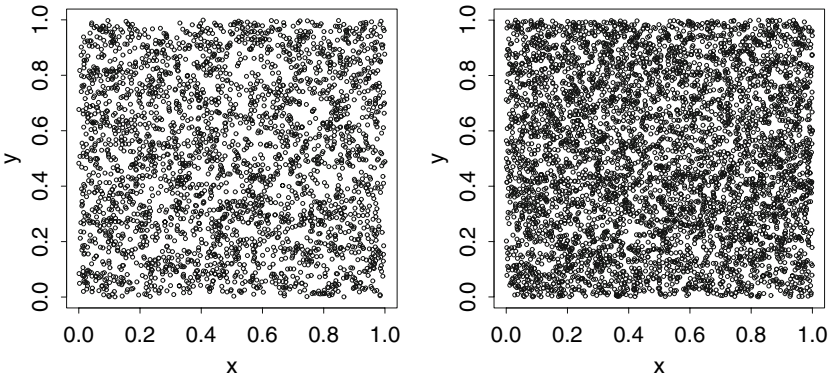

Fig. 7 *Left*: $R = 0.03$ and $\beta_0 = 5$, $\beta_k = 3{,}000$, $k \geq 2$, the total number of points 2971, the maximal number of neighbors 18. *Right:* $R = 0.07$ and $\beta_0 = 5$, $\beta_k = 5{,}000$, $k \geq 1$, the total number of points 5,057, the maximal number of neighbors 97

5 MLE for CSA point processes with a finite number of non-zero intensities

In this section we are going to describe a possible scheme of maximum likelihood estimation (MLE) for CSA point processes with a finite number of non-zero intensities. This class of CSA point processes seems to be convenient for applications.

In case of a finite number of non-zero intensities the process density is characterized by the following parameters: R the interaction radius, N the maximal allowed number of neighbors (the number of non-zero intensities minus one) and $N+1$ intensities $\beta_0, \ldots, \beta_N$. Denoting $\theta = \{\theta_k = \log(\beta_k),\ k = 0, \ldots, N\}$ one can write the unnormalized process density in exponential form

$$h_{R,N,\theta}(\mathbf{x}) = e^{\sum_{k=0}^{N} \theta_k t_k(\mathbf{x})} 1_{\{\widehat{N}(\mathbf{x}) \leq N\}}, \tag{22}$$

where the statistic $t_k(\mathbf{x})$, $k = 0, \ldots, N$, is a number of $\mathbf{x}$-coordinates which have exactly k neighbors and $\widehat{N}(\mathbf{x})$ is a maximal number of neighbors observed in

$\mathbf{x} = (x_1, \ldots, x_n)$ (see formulas (3) and (4)). Given observation $\mathbf{x}$ we look for a set of parameters $\{R, N, \theta_k,\ k = 0, \ldots, N\}$, that maximize the log-likelihood

$$L(R, N, \theta, \mathbf{x}) = \sum_{k=0}^{N} \theta_k t_k(\mathbf{x}) - \log(Z_{R,N}(\theta)), \tag{23}$$

$Z_{R,N}(\theta)$ is the normalizing constant corresponding to the parameters $\{R, N, \theta\}$.

It is easy to see that if the radius R is known, then $\widehat{N}$ is a maximum likelihood estimator for N. $\widehat{N}$ is a biased estimator of N, since $\widehat{N}(\mathbf{x}) \leq N$ for any realization $\mathbf{x}$. Assume that a true parameter is $N = \widehat{N} + n$, where $n \geq 1$, which means that the number of non-zero intensities is $\widehat{N} + n + 1$. It is easy to show that if $\widehat{\theta}_k,\ k = 0, \ldots, \widehat{N} + n$, are maximum likelihood estimators, then $\widehat{\theta}_k = -\infty$, or, equivalently $\widehat{\beta}_k = 0$, for $k = \widehat{N} + 1, \ldots, \widehat{N} + n - 1$. Therefore, given the interaction radius R one can stick with the model with $\widehat{N} + 1$ non-zero intensities.

Assuming that the true value of the radius is in a range $[R_{\min}, R_{\max}]$ one can approximate this continuous set by a finite discrete set $\{R_k,\ k = 1, \ldots, M\}$, determine $\widehat{N} = \widehat{N}(R_k, \mathbf{x})$ for any $k = 1, \ldots, M$, and compute maximum profile likelihood estimates

$$\theta(\mathbf{x}, R_k, \widehat{N}) = \arg\max_{\theta} L(R_k, \widehat{N}, \theta, \mathbf{x}),$$

depending on R_k and $\widehat{N}$.

Denote now for short $h_\theta = h_{R,N,\theta}$ the unnormalized density, P_θ the probability distribution of the process with density h_θ and E_θ expectation with respect to P_θ. Since the log-likelihood is a concave function of θ there exists a unique maximum in the interior of the parameter space $\mathbf{R}^{N+1}$ (recall that $\theta_k = \log(\beta_k) \in \mathbf{R}$). Therefore the point of maximum $\widehat{\theta}$ can be found as a unique solution of the maximum log-likelihood equation

$$\nabla_\theta L(R, N, \theta, \mathbf{x}) = 0,$$

which is the following set of equations

$$t_k(\mathbf{x}) = \mathsf{E}_\theta t_k(X), \quad k = 0, 1, \ldots, N. \tag{24}$$

We cannot solve the system of equations (24) directly, since the partition function $Z(\theta) = Z_{R,N}(\theta)$ in (23) cannot be computed. Therefore MCMC approximation to MLE (MCMCMLE) is required. For detailed description, discussions and applications of MCMCMLE see Geyer and Thompson (1992), Geyer (1994, 1999) and references therein. The idea consists in fixing a so-called reference parameter $\psi = (\psi_0, \ldots, \psi_N)$, and maximizing the function

$$l_\psi(R, N, \theta, \mathbf{x}) = L(R, N, \theta, \mathbf{x}) - L(R, N, \psi, \mathbf{x}). \tag{25}$$

It should be noted that the ratio $h_\theta(\mathbf{x})/h_\psi(\mathbf{x})$ is well defined, since the densities h_θ and h_ψ have a common support. The ratio $Z_{R,N}(\theta)/Z_{R,N}(\psi)$ is replaced by its MCMC approximation and one needs to maximize the function

$$l_{\psi,m}(R, N, \theta, \mathbf{x}) = \log\left(\frac{h_\theta(\mathbf{x})}{h_\psi(\mathbf{x})}\right) - \log\left(\frac{1}{m}\sum_{i=1}^{m} e^{\sum_{k=0}^{N}(\theta_k-\psi_k)t_k(X_i)}\right), \qquad (26)$$

where $X_1, \ldots, X_m$ is a sample from a reversible Markov chain generated by Metropolis-Hastings algorithm and which has P_ψ as its stationary distribution. In the case of a CSA point process with a finite number of non-zero intensities this Markov chain is uniformly geometrically ergodic due to the hard-core character of interaction between points (any point in configuration can have not more than N neighbors). By the ergodic theorem for Markov chains we get that the MCMC approximation (26) converges to $l_\psi(R, N, \theta, \mathbf{x})$ as $m \to \infty$. By the general results for exponential families the maximizers $\widehat{\theta}_m = (\widehat{\theta}_{m,0}, \ldots, \widehat{\theta}_{m,N})$ of $l_{\psi,m}(\theta, \mathbf{x})$ converges to the true maximum likelihood estimators $\widehat{\theta} = (\widehat{\theta}_0, \ldots, \widehat{\theta}_N)$ almost surely as $m \to \infty$.

In addition to reversibility and geometric ergodicity of Metropolis–Hastings sampler one needs to verify some other conditions in order to use MCMCMLE. There are some smoothness conditions which are easy to check in our case. Another important condition is positive definiteness of the Fisher information matrix of the model with known R and N. Note that the set of densities with known R and N is an exponential family in minimal form (linear independence of sufficient statistics $t_k,\ k = 0, \ldots, N$ can be verified by a direct computation in this case), therefore positive definiteness of the Fisher information matrix follows from concavity of the log-likelihood function.

Having obtained $\{\theta(\mathbf{x}, R_k, \widehat{N}),\ k = 1, \ldots, M\}$ it remains to determine the point of a global maximum

$$\widehat{R} = \widehat{R}(\mathbf{x}) = \arg\max_{R_k} L(R_k, \theta(\mathbf{x}, R_k, \widehat{N}), \mathbf{x}). \qquad (27)$$

Finally, MLE estimators are $\{\widehat{R}, \widehat{N}, \widehat{\beta}_k = \exp(\widehat{\theta}_k),\ k = 0, \ldots, N\}$.

Consider a numerical example with the real life data in order to mention some specific simulation details. It should be noted that we are not going to perform a detailed statistical analysis of the data or to give any interpretation of the model in relation with the data, these issues are beyond the scope of the paper.

The real life data is provided by Dr. R. Huele, Leiden University, Institute of Environmental Sciences (CML) and is given by a sequence of geographical coordinates of the North Sea points where appearances of harbor porpoises were observed. The observation window is a subregion of the North Sea which after an appropriate scaling is given by the interior of a polygon $A_1A_2A_3A_4A_5A_6$ in Fig. 8. The image of the data in the transformed window is given by a set of labeled points in Fig. 8. The point coordinates are stored in Table 1.

We decide that the true value of the radius is somewhere in the range $[0.001, 0.1]$ and we approximate this continuous set by the discrete set $\{R_k = 0.001 \cdot k,\ k = 1, \ldots, 100\}$. Table 2 gives the values of canonical statistics and the number of non-zero intensities for a given value of the interaction radius.

To find the maximizers of the function (26) we simply computed its values for a discrete subset of a parameters θ and took the point where the maximum is taken. We were making sure that this subset contains the point of maximum (in cases when we got

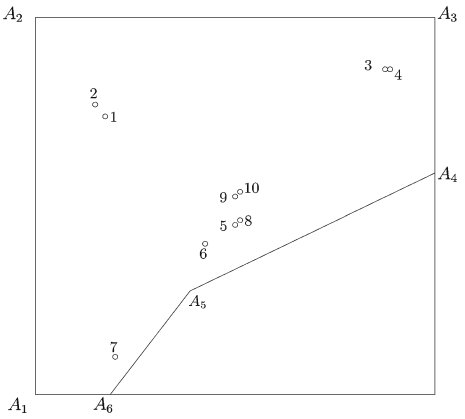

Fig. 8 The observation window and data

Table 1 Data

	1	2	3	4	5	6	7	8	9	10
x	0.175	0.152	0.883	0.889	0.507	0.434	0.226	0.524	0.493	0.518
y	0.740	0.777	0.856	0.856	0.454	0.392	0.110	0.460	0.541	0.548

Table 2 Statistics t_k and $\widehat{N}$ (formulas (3) and (4)) for different values of radius

R	Statistics $t_k = t_k(R), k = 0, \ldots$	$\widehat{N}(R)$
$0.001, \ldots, 0.005$	$(10, 0)$	0
$0.006, \ldots, 0.018$	$(9, 1)$	1
$0.019, \ldots, 0.026$	$(8, 2)$	1
$0.027, \ldots, 0.044$	$(7, 3)$	1
$0.045, \ldots, 0.087$	$(6, 4)$	1
0.088	$(5, 5)$	1
$0.089, \ldots, 0.094$	$(5, 3, 2)$	2
0.095	$(4, 4, 2)$	2
$0.096, \ldots, 0.1$	$(4, 4, 1, 1)$	3

the maximum at the boundary of the subset we took a larger one and repeated the computations). For example, if a value R_k was such that $\widehat{N} = 2$, we computed $l_{\psi,m}(\theta, \mathbf{x})$ at points $\theta = (\theta_0, \theta_1, \theta_2)$ corresponding to the following discrete set described in terms of intensities β: $\beta_0 = 0.5k$, $k = 1, \ldots, 30$, $\beta_1 = 0.5k, k = 1, \ldots, 5000$, and $\beta_2 = k$, $k = 10, \ldots, 500$. The finer grid approximation can be taken to get better approximation. We decide on these sizes of the mesh, compensating it by taking rather fine discrete approximation for the values of the interaction radius.

The sample X_i, $i = 1, \ldots, m$ in the formula (26) was taken from a reference distribution P_ψ corresponding to the hard core model with radius R and non-zero intensities $\psi_0, \ldots, \psi_{\widehat{N}(R)}$. We took different reference parameters for different pairs (N, R). For example, for $R = 0.49$ we had $\widehat{N} = 1$ and put $\beta_0 = 5$ and $\beta_1 = 180$ as reference

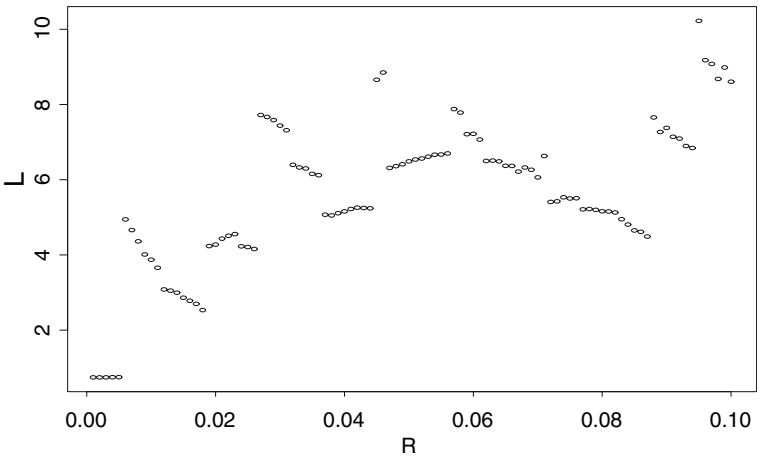

Fig. 9 Plot of $L(R_k) = l_{\psi,m}(\theta(R_k), R_k)$ for $k = 1, \ldots, 100$

intensities. It is easy to see that the model with the minimal radius and the maximal number of non-zero intensities can be taken as a common reference model for all situations considered, since its density support has larger support. Also any Poisson density can be taken as a reference one for any density h_θ. But these choices have the following disadvantage. Assume that the support of the reference density h_ψ is larger than the support of the density h_θ. It yields that when computing the function (26) we have to discard all reference samples X_i such that $h_\theta(X_i) = 0,\ h_\psi(X_i) \neq 0$. By the ergodic theorem the more samples we use the better approximation we have. Hence we need to be sure that sufficiently many reference samples are left. Practically this is not convenient to control. It is much more convenient to use the reference model with the same radius and the same number of non-zero intensities. In this case we "do not loose" any reference samples and the approximation routine is robust. This is why we take the hard core model with $R = R_k$ and $N = \widehat{N}(R_k)$ while approximating the maximizers $\theta(\mathbf{x}, R_k)$ of the cross-section $R = R_k$ of the likelihood surface.

In all cases we run the Metropolis–Hastings algorithm for 25×10^6 steps sampling at every 10, 000th step, so 2, 500 reference samples were simulated. Thus, for a fixed value R we computed the values $\theta_m(\mathbf{x}, R)$ that maximize the function $l_{\psi,m}(\theta, \mathbf{x})$. The points $\theta_m(\mathbf{x}, R_k),\ k = 1, \ldots, 100$ are themselves approximated by some points from some discrete sets.

Determine now $\widehat{R}$ (see (27)). To find the point of the global maximum we need to compare the values of the function (23) at all points of local (profile) maximums. We do it by computing approximations (26) to the function (25) and at this stage the common reference model has to be taken. The Poisson model with intensity 20 is taken as the reference one. The parameter 20 has been taken since we want to have the reference samples $\mathbf{x}$ such that $t(\mathbf{x})$ is comparable 10 and $\widehat{N}(\mathbf{x}) \leq 2$ as often as possible. We computed $m = 3,000$ samples from the reference model drawing a sample every 10, 000 steps in a long run of Metropolis–Hastings algorithm. The values of $l_{\psi,m}(\theta(R_k), R_k),\ k = 1, \ldots, 100$ are plotted in Fig. 9. One can see that the

Table 3 Values of R and $\beta(R)$

R	Statistics $t_k,\ k = 0, \ldots$	$\widehat{N}$	β	$l_m(\theta, R)$
0.005	$t_0 = 10$	0	$\beta_0 = 14$	0.750
*0.006	$(t_0, t_1) = (9, 1)$	1	$(\beta_0,\ \beta_1) = (11, 2010)$	4.944
*0.019	$(t_0, t_1) = (8, 2)$	1	$(\beta_0,\ \beta_1) = (8.5, 441)$	4.235
*0.022	$(t_0, t_1) = (8, 2)$	1	$(\beta_0,\ \beta_1) = (9, 323)$	4.511
*0.027	$(t_0, t_1) = (7, 3)$	1	$(\beta_0,\ \beta_1) = (6.5, 367)$	7.720
0.046	$(t_0, t_1) = (6, 4)$	1	$(\beta_0,\ \beta_1) = (5, 187)$	8.851
*0.088	$(t_0, t_1) = (5, 5)$	1	$(\beta_0,\ \beta_1) = (3.5, 83.5)$	7.655
*0.09	$(t_0, t_1, t_2) = (5, 3, 2)$	2	$(\beta_0,\ \beta_1\ \beta_2) = (4, 35, 130)$	7.378
0.095	$(t_0, t_1, t_2) = (4, 4, 2)$	2	$(\beta_0,\ \beta_1\ \beta_2) = (2.5, 50.5, 105)$	10.224
*0.096	$(t_0, t_1, t_2, t_3) = (4, 4, 1, 1)$	3	$(\beta_0,\ \beta_1\ \beta_2,\ \beta_3) = (2.5, 51, 40, 158)$	9.177

log-likelihood function has clear jumps at those values of the radius where the statistics $\widehat{N}$ and t_k's change their values. We put in Table 3 the values of R and $\beta(R)$ such that $\theta(R) = \log(\beta(R))$ maximizes the function $l(\theta, R)$. For convenience we included in the table the values of sufficient statistics. Also Table 3 contains the same information for those values of R where the statistics values change, these rows are labeled by a star. One can see from this table that the point of the global maximum and therefore the MLE estimate is $(\widehat{R} = 0.095, \widehat{N} = 2, \widehat{\beta}_0 = 2.5, \widehat{\beta}_1 = 50.5, \widehat{\beta}_2 = 105)$.

Acknowledgments This research was supported by the Technology Foundation STW, applied science division of NWO, and the technology program of the Ministry of Economic Affairs, The Netherlands (project CWI.6155 'Markov sequential point processes for image analysis and statistical physics'). The author would like to thank M.N.M. van Lieshout for helpful discussions and valuable comments on the present paper, A.G. Steenbeek for programming assistance and R. Huele for the real life data. The author gratefully thanks both anonymous referees for suggestions and comments leading to an improved exposition.

References

Abramowitz, M., Stegun, I. (1972). *Handbook of mathematical functions with formulas, graphs and mathematical tables*. New York: Dover.

Baddeley, A. J., van Lieshout, M. N. M., Møller, J. (1996). Markov properties of cluster processes. *Advances in Applied Probability, 28*(N2), 346–355.

Comtet, L. (1974). *Advanced combinatorics. The art of finite and infinite expansions.* Dordrecht, Netherlands: Reidel.

Evans, J. W. (1993). Random and cooperative sequential adsorption. *Reviews of Modern Physics, 65*(N4), 1281–1329.

Geyer, C., Thompson, E. (1992). Constrained Monte Carlo likelihood for dependent data. *Journal of Royal Statistical Society B, 54*(N3), 657–699.

Geyer, C. (1994). On the convergence of Monte Carlo maximum likelihood calculations. *Journal of Royal Statistical Society B, 56*(N1), 261–274.

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

van Lieshout, M. N. M. (2006a). Markovianity in space and time. In Dynamics & stochastics: Festschrift in honour of M. S. Keane, D. Denteneer, F. den Hollander, E. Verbitskiy (Eds.), *Institute of Mathematical Statistics, Lecture Notes-Monograph Series, 48*, 154–168.

van Lieshout, M. N. M. (2006b). Campbell and moment measures for finite sequential spatial processes. Research Report PNA-R0601, CWI, Amsterdam.

van Lieshout, M. N. M. (2006c). Maximum likelihood estimation for random sequential adsorption. *Advances in Applied Probability*, *38*(N4), 889–898.

Privman, V. (Ed.) (2000). A special issue of *Colloids and Surfaces A, 165*.