TRANSFORMING SPATIAL POINT PROCESSES INTO POISSON PROCESSES USING RANDOM SUPERPOSITION

By Jesper Møller* and Kasper K. Berthelsen*

Aalborg University

Most finite spatial point process models specified by a density are locally stable, implying that the Papangelou intensity is bounded by some integrable function β defined on the space for the points of the process. It is possible to superpose a locally stable spatial point process X with a complementary spatial point process Y to obtain a Poisson process $X \cup Y$ with intensity function β. Underlying this is a bivariate spatial birth-death process (X_t, Y_t) which converges towards the distribution of (X, Y). We study the joint distribution of X and Y, and their marginal and conditional distributions. In particular, we introduce a fast and easy simulation procedure for Y conditional on X. This may be used for model checking: given a model for the Papangelou intensity of the original spatial point process, this model is used to generate the complementary process, and the resulting superposition is a Poisson process with intensity function β if and only if the true Papangelou intensity is used. Whether the superposition is actually such a Poisson process can easily be examined using well known results and fast simulation procedures for Poisson processes. We illustrate this approach to model checking in the case of a Strauss process.

1. Introduction. A spatial birth-death process is a continuous time jump process where each jump consists in either adding or removing a point from a finite spatial point pattern. Preston (1977) provided a detailed mathematical study of such processes, and showed among other things that under suitable conditions, (approximate) realisations of a finite spatial point process can be obtained by running a spatial birth-death process for a long enough time; this point was also taken by Kelly and Ripley (1976) and Ripley (1977), and in connection to perfect simulation algorithms by Kendall (1998), Kendall and Møller (2000), and Fernández, Ferrari and Gar-

*Supported by the Danish Natural Science Research Council, grants 272-06-0442 and 09-072331, 'Point process modelling and statistical inference', and by the Centre for Stochastic Geometry and Advanced Bioimaging, funded by a grant from the Villum Foundation.

AMS 2000 subject classifications: Primary 60G55, 62M99; secondary 62M30, 66C60.

Keywords and phrases: complimentary point process, coupling, local stability, model checking, Papangelou conditional intensity, spatial birth-death process, Strauss process

cia (2002). Spatial birth-death processes have also been used as statistical models for geological data (Fiksel, 1984; Stoyan, Kendall and Mecke, 1995) and sand dunes (Møller and Sørensen, 1994), and for Bayesian analysis of mixture models with an unknown number of components (Stephens, 2000).

Preston (1977) established the existence of a spatial birth-death process through a coupling to a non-explosive birth-death process on the non-negative integers, which can be extended to a 'dominating' spatial birth-death process. This coupling is particular useful in connection with locally stable point processes, which is a property satisfied by most spatial point process models specified by a density; this condition and other background material are presented in Section 2. Briefly, local stability implies that the Papangelou conditional intensity $\lambda(x,u)$ is bounded from above by an integrable function $\beta(u)$ defined on S, where S denotes the state space of the points, x is any finite point pattern (i.e. a finite subset of S), and $u \in S \setminus x$ is any point. To describe the coupling construction consider a (dominating) birth-death process D_t with birth rate $\beta(u)$ and death rate one so that its equilibrium distribution is a Poisson process on S with intensity function β. It is possible by a dependent thinning of D_t to obtain a (target) birth-death process X_t with birth rate $\lambda(x,u)$ and death rate one such that its distribution converges towards that of X as time t tends to infinity: the dependent thinning is such that if $X_0 \subseteq D_0$ then $X_t \subseteq D_t$ for all $t > 0$ (explaining what is meant by 'dominating'). Further details are given in Section 2.2. This coupling construction also plays a key role in connection with the perfect simulation algorithms of locally stable point processes given by the dominating coupling from the past algorithm (Kendall, 1998; Kendall and Møller, 2000) and the method of clans of ancestors (Fernández, Ferrari and Garcia, 2002).

In this paper we study the birth-death process $Y_t = D_t \backslash X_t$, i.e. the points in the dominating process D_t that are *not* included in the target birth-death process X_t. We refer to Y_t as the *complementary* birth-death process (to the target birth-death process X_t). Section 2.2 defines the bivariate birth-death process (X_t, Y_t), and Section 3 establishes that (X_t, Y_t) converges towards a bivariate point process (X, Y), where we call Y the *complementary* point process (to the target point process X). In general it seems difficult to say anything detailed about this equilibrium distribution except in special cases considered in Section 3 and in Appendix A.

Although the distribution of Y conditional on $X = x$ seems complicated in general, it turns out to be simple to simulate from this conditional distribution. Section 4.1 presents an algorithm which is both fast and easily implemented. Section 4.2 studies the speed of the algorithm which, unlike

dominating CFTP, depends only on β, i.e. it does not depend on any (interaction or monotonicity) properties of λ expect on its upper bound β.

The algorithm may be used for model checking: given data x (a finite point pattern in S) and a model for the Papangelou intensity of the underlying spatial point process X, this model is used for generating a realisation y from the complementary process conditional on $X = x$. Section 5.1 establishes that the resulting superposition $x \cup y$ is a Poisson process with intensity function β if and only if the true Papangelou intensity is used. Whether the superposition is actually such a Poisson process can easily be examined using theoretical results for (functional) summary statistics of Poisson processes, where quantiles of the summary statistics can be quickly simulated. Section 5.2 illustrates this approach to model checking in the case of a Strauss process (Strauss, 1975; Kelly and Ripley, 1976).

The above model checking procedure of superimposing the complementary point pattern on the data pattern and checking if the resulting point pattern is Poisson has some similarities to the approach considered by Møller and Schoenberg (2010). Their procedure is based on dependent thinning of the data pattern x obtaining a realisation of a Poisson process if the assumed model for X is correct. This construction relies on an assumption of a positive lower bound on the Papangelou intensity on S which is typically not available for most point process of interest.

2. Preliminaries.

2.1. *Assumptions.* For simplicity and specificity we consider a spatial point process X defined on a Borel set $S \subset \mathbb{R}^k$ ($k \in \{1, 2, \ldots\}$) of finite and positive Lebesgue measure $|S|$, where with probability one, X is finite and simple (i.e. has no multiple points). This means that X can be considered as a random finite subset of S, so realizations of X are finite point configurations $x = \{x_1, \ldots, x_n\} \subset S$, with $0 \leq n < \infty$ (for $n = 0$, $x = \emptyset$ is the empty point configuration). For measure theoretical details, see e.g. Appendix B in Møller and Waagepetersen (2004). The setting covers most cases of practical interest, but our methods can easily be extended to non-simple point processes defined on a general state space and using an exponential state space setting (Carter and Prenter, 1972; Preston, 1977; Ripley and Kelly, 1977). We refer to X as our *target point process*.

Let β be a non-negative Lebesgue integrable function β defined on S, denote Poisson(S, β) the distribution of the Poisson process on S with intensity function β, and set $\nu =$ Poisson$(S, 1)$ (the distribution of the homogeneous

Poisson process on S with intensity one). Note that

(1)
$$\int h(x)\,\mathrm{d}\nu(x) = \mathrm{e}^{-|S|}h(\emptyset) + \sum_{n=1}^{\infty} \frac{\mathrm{e}^{-|S|}}{n!} \int_S \cdots \int_S h(\{x_1,\ldots,x_n\})\,\mathrm{d}x_1 \cdots \mathrm{d}x_n$$

for any non-negative measurable function h defined on the space of all finite subsets of S.

We assume that X is absolutely continuous with respect to ν and denotes its density f. We also assume that f is *locally stable* with respect to β, i.e. for any finite point configuration $x \subset S$ and any point $u \in S \setminus x$,

$$f(x \cup \{u\}) \le \beta(u) f(x). \tag{2}$$

This condition is satisfied for most point process models specified by a density (where of course the choice of β depends on the density), see Geyer (1999) and Møller and Waagepetersen (2004). Clearly (2) implies that the *Papangelou conditional intensity* defined for any finite $x \subset S$ and any $u \in S \setminus x$ by

$$\lambda(x,u) = f(x \cup \{u\})/f(x) \quad (\text{taking } 0/0 = 0) \tag{3}$$

is bounded by $\beta(u)$. In fact local stability also implies that many desirable properties for simulation algorithms are satisfied, cf. Møller and Waagepetersen (2004) and the references therein. Note that

$$b := \int_S \beta(u)\,\mathrm{d}u$$

is finite and equal to the mean number of points under Poisson(S,β). Henceforth, to avoid the trivial case where $f(x) = 0$ whenever $x \neq \emptyset$, we assume that $b > 0$.

2.2. *Coupled spatial birth-death processes.* We shall exploit that (3) ensures a coupling of a continuous time spatial birth-death process $\{X_t : t \ge 0\}$ with a dominating spatial birth-death process $\{D_t : t \ge 0\}$ such that $X_t \subseteq D_t$ for all times $t \ge 0$, where the X_t process has birth rate $\lambda(x,u)$ and death rate one, and the D_t process has birth rate $\beta(u)$ and death rate one. Both processes are time reversible, X_t has equilibrium density f, and the equilibrium distribution of D_t is Poisson(S,β). See Preston (1977), Kendall (1998), Kendall and Møller (2000), and Appendix G in Møller and Waagepetersen (2004). However, in general, as shown in Section 3, the coupled process (X_t, D_t) is not time reversible.

For later use, we now recall the details of the coupling construction, where we let the initial states be arbitrary except that it is assumed that $X_0 \subseteq D_0$.

First, we generate the dominating birth-death process $\{D_t : t \geq 0\}$ as follows. We start by generating a pure birth process on S with birth rate $\beta(u)$. Viewed as a space-time point process, this is simply a Poisson process on $S \times [0, \infty)$ with intensity function $\rho(u, t) = \beta(u)$. To each space-time point (u, t) in this Poisson process, we generate a lifetime $\tau(u)$ which is exponentially distributed with mean one; these lifetimes are independent of the birth process, and the lifetimes are mutually independent. In the dominating birth-death process, the point u is then included for the time period starting at the birth time t and ending at the death time $t + \tau(u)$ (where u is excluded).

Second, the X_t process is obtained from the D_t process by a dependent thinning. Write D_{t-} and X_{t-} for the states of the processes just before time t. If a birth happens in the dominating spatial birth-death process at time $t > 0$ so that $D_t = D_{t-} \cup \{u\}$, then conditional on this knowledge and what previously has happened in the two processes, with probability $\lambda(X_{t-}, u)/\beta(u)$ set $X_t = X_{t-} \cup \{u\}$, and otherwise do nothing, i.e. $X_t = X_{t-}$ is unchanged. Moreover, if a death happens in the dominating spatial birth-death process at time $t > 0$ so that $D_t = D_{t-} \setminus \{u\}$ where $u \in D_{t-}$, then $X_t = X_{t-} \setminus \{u\}$ (of course $X_t = X_{t-}$ is unchanged if u is not in X_{t-}). Finally, as a transition in the X_t process can only happens if a similar transition happens in the D_t process, it follows that $X_t \subseteq D_t$ for all $t \geq 0$.

We call $Y_t = D_t \setminus X_t$, $t \geq 0$, the complementary spatial birth-death process. Note that $\{(X_t, Y_t) : t \geq 0\}$ is a bivariate jump process such that a transition from a given state $(x, y) = (\{x_1, \ldots, x_m\}, \{y_1, \ldots, y_n\}) \subset S \times S$ of the process (with $x = \emptyset$ if $m = 0$, and $y = \emptyset$ if $n = 0$) can be of one of four types (i)-(iv), where the rate of the transition is

(i) $\lambda(x, u)$ if $(x \cup \{u\}, y)$ is the new state, i.e. when a birth of a point $u \in S$ happens in the X_t process;
(ii) $\beta(u) - \lambda(x, u)$ if $(x, y \cup \{u\})$ is the new state, i.e. when a birth of a point $u \in S$ happens in the Y_t process;
(iii) one if $(x \setminus \{x_i\}, y)$ is the new state, i.e. when the ith point in the X_t process dies ($i \in \{1, \ldots, m\}$ and provided $m > 0$),
(iv) one if $(x, y \setminus \{y_j\})$ is the new state, i.e. when the jth point in the X_t process dies ($j \in \{1, \ldots, n\}$ and provided $n > 0$).

3. The equilibrium distribution of the bivariate jump process. The (X_t, Y_t) process converges in distribution towards a unique equilibrium distribution Π; in fact the process converges geometrically fast towards Π as

seen by combining results in Møller (1989) with Appendix G in Møller and Waagepetersen (2004). Henceforth, assume that (X, Y) follows Π. We call Y the *complementary point process* (to the target point process X). Note that $D = X \cup Y$ follows Poisson(S, β), but what else can we say about Π? The following Propositions 1-3 are verified in Appendix A.

Proposition 1. The equilibrium distribution Π is absolutely continuous with respect to the product measure $\nu \times \nu$.

We need some further notation. Let $\pi(x, y)$ denote the density of Π with respect to $\nu \times \nu$. For any finite $x \subset S$, let $n(x)$ denote the cardinality of x. If $n(x) = 0$, set $\sum_{u \in x} q(x, u) = 0$ for any real function $q(x, u)$.

Proposition 2. Apart from a $\nu \times \nu$-nullset, the equilibrium density π is the unique density satisfying the equation

$$\begin{aligned}
&[b+n(x)+n(y)]\pi(x,y) \\
&\quad = \sum_{u\in x} \lambda(x \setminus \{u\}, u)\pi(x \setminus \{u\}, y) + \sum_{u \in y} [\beta(u) - \lambda(x,u)]\pi(x, y \setminus \{u\}) \\
&\quad + \int_S \pi(x \cup \{u\}, y)\, \mathrm{d}u + \int_S \pi(x, y \cup \{u\})\, \mathrm{d}u \qquad (4)
\end{aligned}$$

for all finite $x, y \subset S$.

We have not been able to solve (4) without imposing rather restrictive conditions, such as in the following Proposition 3 or as in the examples discussed in Appendix A.

One attempt at solving (4) is given by solving the detailed balance condition

$$\pi(x,y)\lambda(x,u) = \pi(x \cup \{u\}, y), \quad \pi(x,y)(\beta(u) - \lambda(x,u))\pi(x, y \cup \{u\}), \qquad (5)$$

which is equivalent to time reversibility of the (X_t, Y_t) process. This is, however, only satisfied in the following simple case.

Proposition 3. The equilibrium density $\pi(\cdot, \cdot)$ satisfies the detailed balance condition (5) if and only if $\lambda(x, u) = \lambda(u)$ does not depend on x, in which case X and Y are independent Poisson processes on S with intensity functions $\lambda(u)$ and $\beta(u) - \lambda(u)$, respectively.

As noticed in Remark 1 (Section 5.1) and in Appendix A, apart from the case where the detailed balance condition (5) holds, the conditional distribution of Y given $X = x$ is in general a complicated distribution—nevertheless we can easily simulate from this conditional distribution as shown in Section 4.1. So in general it seems difficult to explicitly evaluate

the joint density of X and Y. Also the density of Y seems in general to be very complicated as discussed in Appendix A.

4. Conditional simulation of the complimentary point process.

4.1. *Simulation procedure.* The following algorithm provides an easy way to make a conditional simulation $Y(x)$ of the complimentary point process given that x is a realization from the target point process X.

Algorithm 1.

(a) Set $Y(x) = \emptyset$ and generate Z from Poisson(S, β). If $Z = \emptyset$, then set $T = 0$ and go to (e).
(b) For each point $u \in Z$, generate an exponentially distributed lifetime T_u with mean one, and a uniformly distributed "mark" M_u on $[0, 1]$, assuming that all these times and marks are mutually independent. Set $T = \max\{T_u : u \in Z\}$.
(c) Set $X_0 = x$ and generate the spatial birth-death process X_t with birth rate λ and death rate one, stopping the generation at time T, assuming this generation conditional on T is independently of everything else in (a)-(b).
(d) For each $u \in Z$, if $M_u > \lambda(X_{T_u}, u)/\beta(u)$, add u to $Y(x)$, i.e. $Y(x) \leftarrow Y(x) \cup \{u\}$.
(e) Return $Y(x)$.

Theorem 1. The output $Y(x)$ in Algorithm 1 is a realization from the conditional distribution of the complimentary point process Y given that $X = x$ is a realization from the target point process.

Proof. Intuitively, this follows by

- imaging that we have extended the (X_t, Y_t) process to all times $t \in \mathbb{R}$ such that it is in equilibrium; this is easily done, since the (X_t, D_t) process regenerates each time $D_t = \emptyset$, see e.g. Appendix G in Møller and Waagepetersen (2004);
- observing that by (i)-(iv) in Section 2.2, conditional on $\{X_t : t \in \mathbb{R}\}$, the births in $\{Y_t : t \in \mathbb{R}\}$ form a space-time Poisson process B on $S \times \mathbb{R}$ with intensity function $\mu(u, t) = \beta(u) - \lambda(X_t, u)$, the corresponding lifetimes are mutually independent and independent of B, and each lifetime is exponentially distributed with mean one;
- noting that B can be obtained by an independent thinning from a space-time Poisson process on $S \times \mathbb{R}$ with intensity function $\rho(u, t) = \beta(u)$, where the retention probability for a space-time point (u, t) is $\mu(u, t)/\rho(u, t) = 1 - \lambda(X_t, u)/\beta(u)$.

For a formal proof, it is convenient to reverse time, and to imagine that we have generated more than is actually needed, as described in the following.

Let $\{D_t^* : t \leq 0\}$ be an independent copy of the dominating spatial birth-death process $\{D_t : t \leq 0\}$ considered backwards in time, where $D_0^* = Z$. We use this notation, since (as discussed at the beginning of Section 4.2) the D_t process may instead be used when generating the X_t process. In step (b), T is the largest lifetime of the points in Z; correspondingly, let $\tilde{T}$ be the first time before time 0 where a point in D_0^* was born when the D_t^* process is considered forwards in time, setting $\tilde{T} = 0$ if $D_0^* = \emptyset$; so $-\tilde{T}$ is distributed as T. Moreover, suppose that we have generated the X_t process backwards in time $t \leq 0$, independently of the D_t^* process and anything else associated to this process as considered below, and with $X_0 = X$.

By time reversibility, the generation of these processes is just like running them forwards it time. To each birth time t in D_t^* (considered forwards in time) we attach a mark given by a uniformly distributed random variable M_t on $[0, 1]$. All these marks are assumed to be mutually independent and independent of $\{(X_t, D_t^*) : t \leq 0\}$.

Moreover, suppose that for any time $s < 0$, we have generated a complimentary spatial birth-death process Y_t^s forwards in time $t \in [s, 0]$ in the following way. Initially, $Y_s^s = \emptyset$. Further, a birth in the Y_t^s process can only happen if it also happens in the D_t^* process: if $D_t^* = D_{t-}^* \cup \{u\}$, then $Y_t^s = Y_{t-}^s \cup \{u\}$ if $M_t > \lambda(X_t, u)/\beta(u)$, and $Y_t^s = Y_{t-}^s$ otherwise. Furthermore, a death in the Y_t^s process can only happen if it also happens in the D_t^* process: if a death happens so that $D_t^* = D_{t-}^* \setminus \{u\}$ (where $u \in D_{t-}^*$), then $Y_t^s = Y_{t-}^s \setminus \{u\}$. Hence $\{(X_t, Y_t^s) : s \leq t \leq 0\}$ is seen to be a jump process with transition rates as given in (i)-(iv) in Section 2. Consequently, $\{(X_t, Y_t^s) : s \leq t \leq 0\}$ is distributed as $\{(X_t, Y_t) : 0 \leq t \leq -s\}$ with X_0 in equilibrium and $Y_0 = \emptyset$.

Note that $Y_0^s \subseteq D_0^*$ and whether or not a birth happens in the complimentary spatial birth-death process does not depend on the history in this process. Hence to generate Y_0^s, if $s \leq \tilde{T}$, we need only to consider the death times of the points in D_0^* (when D_t^* is viewed backwards in time) and to use the states of the X_t process at these death times. So in our simulation procedure we need only the steps (a)-(e), and $(X, Y(X))$ is distributed as the limiting distribution of (X_{-s}, Y_{-s}) as $-s \to \infty$. Therefore $(X, Y(X))$ follows Π, and so $Y(X)$ is distributed as Y conditional on X. Thereby Theorem 1 is verified.

Remark 1. It follows from the proof of Theorem 1 that the conditional distribution of Y given $\{X_t : t \geq 0\}$ is a Poisson process on S with intensity

function

$$\beta(u) - \sum_{i=0}^{\infty} \left(\mathrm{e}^{-\tau_i} - \mathrm{e}^{-\tau_{i+1}}\right) \lambda(X_{\tau_i}, u), \quad u \in S, \tag{6}$$

where $\tau_0 = 0$ and $\tau_1 < \tau_2 < \ldots$ denote the transition (or jump) times of $\{X_t : t \geq 0\}$, and where $\mathrm{e}^{-\tau_i} - \mathrm{e}^{-\tau_{i+1}}$ is the probability that an exponentially distributed lifetime with mean one is falling in the interval from τ_i to τ_{i+1} (within this interval X_t is constant). Now, Y conditional on $X = x$ is distributed as Y conditional on $X_0 = x$, and the latter distribution may in principle be obtained by considering the Poisson process on S with intensity function (6) and integrating over all possible paths of $\{X_t : t > 0\}$ when $X_0 = x$. However, apart from the special case where $\lambda(x, u) = \lambda(u)$ does not depend on x, this calculation appears to be very complicated, indicating that the conditional distribution of Y given X is in general a complicated distribution.

Remark 2. A comparison of Algorithm 1 with perfect simulation algorithms seems in order, since the proof of Theorem 1 has some similarity to arguments used when establishing the correctness of the CFTP (coupling from the past) algorithm in Propp and Wilson (1996), the dominating CFTP algorithm in Kendall (1998) and Kendall and Møller (2000), and the method of clans of ancestors in Fernández, Ferrari and Garcia (2002). The latter two algorithms are used for perfect simulation of a locally stable point process, using spatial birth-death processes in different ways. As argued below, Algorithm 1 is much simpler to implement and much faster than these perfect simulation algorithms.

The speed of the dominating CFTP algorithm depends much on monotonicity properties of $\lambda(x, u)$ (used for generating a sequence of so-called lower and upper processes) and how strong the interaction is between u and neighbouring points in x (a point $v \in x$ is said to be a neighbour to u if $\lambda(x, u)$ depends on v). In fact, in cases where $\lambda(x, u)$ can be much smaller than $\beta(u)$, the dominating CFTP algorithm can be very slow (Berthelsen and Møller, 2002). Moreover, a doubling scheme is used in the dominating CFTP algorithm (this doubling scheme is for the abovementioned sequence of lower and upper processes), where Berthelsen and Møller (2002) recommends that the first time in the doubling scheme should be random and distributed as T (more precisely, this is the case if the waiting times for transitions are included; in fact one needs only to consider the jump chain in the dominating CFTP algorithm, so this partly reduces the computations). Furthermore, the method of clans of ancestors depends on how $\lambda(x, u)$ specifies which points are neighbours (but not on how strong the interaction

is), and this method can be very slow, in fact even slower than dominating CFTP (Kendall and Møller, 2000; Berthelsen and Møller, 2002).

In contrast, Algorithm 1 depends neither on any monotonicity property of $\lambda(x,u)$, or on how strong the interaction is, or on how $\lambda(x,u)$ specifies which points are neighbours. The speed of our simulation procedure (a)-(e) depends only on b, as further discussed in Section 4.2.

Remark 3. Algorithm 1 is useful when verifying Theorem 1. In practice it is easier to use the following algorithm, where we avoid simulating the lifetimes but the output is still a conditional simulation of Y given $X = x$.

Algorithm 2.

(a) Set $Y(x) = \emptyset$ and $w = x$. Generate M from a Poisson distribution with mean b.
(b) While $M > 0$ repeat steps (c) to (f):
  (c) Set $n = n(w)$ and $m = M$. Generate a uniformly distributed variable v on $[0,1]$.
  (d) If $v < \frac{m}{m+n+b}$ perform steps (d.1) to (d.3):
    (d.1) reduce M by one, i.e. $M \leftarrow M - 1$;
    (d.2) generate a point u on S with density $\beta(\cdot)/b$;
    (d.3) with probability $1-\lambda(w,u)/\beta(u)$, add u to $Y(x)$, i.e. $Y(x) \leftarrow Y(x) \cup \{u\}$.
  (e) If $v \in [\frac{m}{m+n+b}, \frac{m+n}{m+n+b}]$ then
    (e.1) remove a point u from w chosen uniformly at random, i.e. $w \leftarrow w \setminus \{u\}$.
  (f) If $v > \frac{m+n}{m+n+b}$ then perform steps (f.1) and (f.2):
    (f.1) generate a point u on S with density $\beta(\cdot)/b$;
    (f.2) with probability $\lambda(w,u)/\beta(u)$, add u to w, i.e. $w \leftarrow w \cup \{u\}$.
(g) Return $Y(x)$.

4.2. *The speed of the algorithm.* In this section we consider the computational load of using Algorithm 2. Inspecting Algorithm 2 we notice that essentially it only involves two computational aspects, one is generating a point on S with density $\beta(\cdot)/b$ and the other is evaluating $\lambda(w,u)/\beta(u)$. Typically the complexity of the latter is (much) higher than the former. This leads us to quantify the computational load of Algorithm 2 in terms of the number of times $\lambda(w,u)/\beta(u)$ is evaluated. Algorithm 2 evaluates $\lambda(w,u)/\beta(u)$ only in step (d.3), which happens M times, and in step (f.2). Letting N denote the number of times (f.2) is evaluated, the computational

load is $C = M+N$. Step (f) corresponds to a birth in the dominating process D_t. Hence, N corresponds to the number of births in the dominating process in a time interval of random length T, where T is defined as in Algorithm 1.

Proposition 4. The mean computational load is

$$\mathrm{E}(C) = b\left(1 - \mathrm{e}^{-b}\right) + b\mathrm{E}(T), \tag{7}$$

where

$$\mathrm{E}(T) = \left(1 - \mathrm{e}^{-b}\right)\left[\left(1 - \mathrm{e}^{-b}\right)\ln(b) - \int_0^b \ln(s)\mathrm{e}^{-s}\,\mathrm{d}s\right] \tag{8}$$

$$\leq \left(1 - \mathrm{e}^{-b}\right)\left[\left(1 - \mathrm{e}^{-b}\right)\ln(b) + a\right], \tag{9}$$

where $a = -\int_0^1 \ln(s)\mathrm{e}^{-s}\,\mathrm{d}s \approx 0.7966$.

Proof. Conditional on T, M and N are independent and Poisson distributed, with mean b and bT, respectively. Hence, as $\exp(-b)$ is the probability that $T = 0$, we obtain (7). Furthermore, conditional on $M > 0$, T has density

$$g(t) = \sum_{n=1}^{\infty} \frac{b^n}{n!}\mathrm{e}^{-b} n\mathrm{e}^{-t}\left(1 - \mathrm{e}^{-t}\right)^{n-1} = b\mathrm{e}^{-t}\exp\left(-b\mathrm{e}^{-t}\right), \quad t > 0,$$

and so

$$\mathrm{E}(T) = \left(1 - \mathrm{e}^{-b}\right)\int_0^\infty t g(t)\,\mathrm{d}t = \left(1 - \mathrm{e}^{-b}\right)\int_0^b \ln(b/s)\mathrm{e}^{-s}\,\mathrm{d}s,$$

where we have made a change of variable to $s = b\mathrm{e}^{-t}$. This reduces to (8), and (9) is easily obtained.

The integral in (8) can easily be evaluated by numerical integration, and in most applications, the term $\exp(-b)$ appearing in both (7), (8), and (9) will be effectively zero. It follows that

$$\mathrm{E}(C) \leq b + b(\ln(b) + a),$$

where in most applications, $\leq$ can be replaced by $\approx$.

5. Model checking.

5.1. *The random superposition procedure.* Suppose that a realization x from a spatial point process X^* with "true" density f^* is observed, and we want to check the goodness of fit for a fitted model with density f, where both f^* and f are locally stable. By Theorem 1, the *random superposition procedure* $X \cup Y(X)$ is a realization of Poisson(S, β) if $f = f^*$. Conversely, the following theorem establishes that if f and f^* are not specifying the same model, then the superposition

$$D^* = X^* \cup Y(X^*)$$

is not following Poisson(S, β).

Theorem 2. D^* follows Poisson(S, β) if and only if f and f^* agree except on a ν-nullset.

Proof. We already noticed that the "if"-part holds. The following equation (10) becomes useful when verifying the "only if"-part. Using (1) it is straightforwardly verified that for any non-negative measurable function $h(x, y)$,

$$\int \sum_{x \subseteq z} h(x, z \setminus x) \, \mathrm{d}\nu(z) = \mathrm{e}^{|S|} \int \int h(x, y) \, \mathrm{d}\nu(x) \, \mathrm{d}\nu(y).$$

This immediately implies that if X_1 and X_2 are spatial point processes on S such that X_1 has density π_1 and X_2 conditional on $X_1 = x$ has density $\pi_2(\cdot|x)$, then $X_1 \cup X_2$ has density

$$\pi(z) = \mathrm{e}^{-|S|} \sum_{x \subseteq z} \pi_1(x) \pi_2(z \setminus x | x). \tag{10}$$

Let

$$q(z) = \mathrm{e}^{|S|-b} \prod_{u \in z} \beta(u)$$

denote the density of Poisson(S, β). By (a)-(c) in Algorithm 1, $Y(x)$ has a density $g(\cdot|x)$; specifically

$$g(y|x) = \int q(y \cup w) h(x, y, w) \, \mathrm{d}\nu(w),$$

where

$$h(x, y, w) = \mathrm{E}\left[\left\{ \prod_{u \in y} \left(1 - \frac{\lambda(X_{T_u}, u)}{\beta(u)} \right) \right\} \left\{ \prod_{u \in w} \frac{\lambda(X_{T_u}, u)}{\beta(u)} \right\} \right],$$

with the expectation calculated conditional on that $X_0 = x$. In particular,

$$g(\emptyset|x) \geq \mathrm{P}(Z = \emptyset) = \mathrm{e}^{-b} > 0.$$

Moreover, by Theorem 1 and (10), for ν almost all z,

$$q(z) = \mathrm{e}^{-|S|} \sum_{x \subseteq z} f(x) g(z \setminus x | x). \tag{11}$$

This means that for ν almost all z, $f(z)$ is determined by q and g, since first

$$f(\emptyset) = \mathrm{e}^{|S|} q(\emptyset) / g(\emptyset | \emptyset), \tag{12}$$

and second by induction $f(z)$ is given in terms of those $f(x)$ with x strictly contained in z, using that

$$f(z) = \mathrm{e}^{|S|} \left(q(z) - \sum_{x \subset z} f(x) g(z \setminus x | x) \right) / g(\emptyset | z). \tag{13}$$

Similarly, if D^* follows Poisson(S, β), then for ν almost all z, we also obtain (11) but with f replaced by f^*, and hence (12)-(13) but with f replaced by f^*. Hence the "only if"-part is verified.

Remark 4. In Møller and Schoenberg (2010) model checking is based on a random thinning method—where it is assumed that $\lambda(x, u) \geq \beta(u)$, a condition which is rarely satisfied for point process models as discussed in Møller and Schoenberg (2010)— and where the method of clans of ancestors (discussed in Remark 2) is playing a key role. This seems a less appealing procedure than our random superposition procedure, since the latter is faster, simpler, and general applicable.

5.2. *Example: the Strauss process.* In this section we consider an example of how to utilise Theorem 2 for model checking. The basic idea is as follows. Given data x we generate a realisation y of $Y(x)$ where f has been estimated in some way based on data. According to Theorem 2 the union $x \cup y$ is a realisation of a Poisson process on S with intensity function β if and only if f and f^* specify the same model, provided f and f^* are locally stable. The model check consists in testing the hypothesis that $x \cup y$ is in fact a realisation of a Poisson process on S with intensity function β.

There exist numerous ways of testing if a given point pattern is a realisation of a Poisson point process. In the sequel we assume that β is constant and restrict attention to methods based on Besag's L function which is a useful transformation of Ripley's K function (Besag, 1977). Informally, $L(r)$ is a non-negative functional point process summary which indicates to what extend a given stationary point process exhibits clustering or repulsion at inter-point distance $r > 0$. In the present context the most important property of the L function is that for a Poisson process $L(r) = r$. Further, for

a point process with $L(r) < r$, the expected number of points within a distance r from a 'typical point' is lower than what is expected under a Poisson process, indicating repulsion between the points if r is small. Similarly, $L(r) > r$ implies that more points are expected within distance r from a typical point when compared to a Poisson process, indication clustering between the points if r is small. For more details, including extensions to the case where β is not constant, see Møller and Waagepetersen (2004) and the references therein.

We let $\hat{L}(r) = \hat{L}(r; z)$ denote an estimate of L based on a point pattern z obtained by observing a stationary point process within S, see e.g. Møller and Waagepetersen (2004) and Illian *et al.* (2008). Usually this estimate involves first estimating the intensity of the process; below, unless otherwise stated, we make use of the fact that the intensity is given by β which is assumed known. If z is a realisation of a Poisson process, we expect that $\hat{L}(r; z) - r \approx 0$ for all $r > 0$. Accordingly, if $\hat{L}(x \cup y; r) - r$ deviates too much from zero we have an indication that the model specified by f is not (close to) the true model. To get an handle on the 'too much' part we need to take into account the variation in $\hat{L}$. Let $\underline{L}(r) - r$ and $\overline{L}(r) - r$ denote estimated 2.5% and 97.5% quantiles for $\hat{L}(r; W) - r$ when W follows Poisson(S, β). These estimates are based on independent simulations from Poisson(S, β); unless otherwise stated, we use $k = 239$ such simulations $W^{(1)}, \ldots, W^{(k)}$ so that $\underline{L}(r)$ is the 5th smallest and $\overline{L}(r)$ the 5th largest among $\hat{L}(W^{(1)}; r), \ldots, \hat{L}(W^{(k)}; r)$. We refer to the pair of functions $\underline{L}(r) - r$ and $\overline{L}(r) - r$ as the (pointwise) estimated 95% Poisson envelopes. If $\hat{L}(r; x \cup y) - r$ deviates too much outside these envelopes, we reject the assumed model. For more details on this (informal) test procedure, see e.g. Illian *et al.* (2008).

We illustrate the test procedure in the case of a planar Strauss process, with density

$$f(x) \propto \beta^{n(x)} \gamma^{s_R(x)},$$

where $\beta > 0$, $\gamma \in [0, 1]$, and $R > 0$ are parameters, and where $S_R(x)$ is the number of point pairs $\{u, v\} \subseteq x$ (with $u \neq v$) separated by a distance less than R (Strauss, 1975; Kelly and Ripley, 1976). The Strauss process is locally stable, since

$$\lambda(x, u) = \beta \gamma^{S_R(x,u)} \leq \beta,$$

where $S_R(x, u)$ is the number of points in x within a distance R from u. Thus β is the intensity of the dominating Poisson process, while γ is an interaction parameters and R determines the range of interaction in the Strauss process.

The top left panel in Figure 1 shows a realisation x of a Strauss process on the unit square $S = [0, 1]^2$ with $\beta = 250$, $\gamma = 0.1$, and $R = 0.05$. In

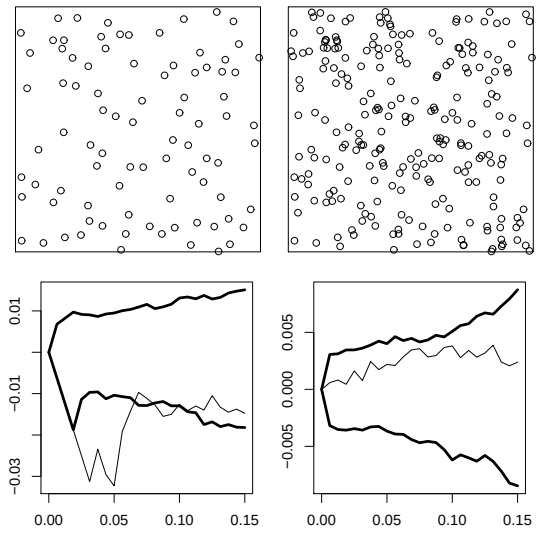

FIG 1. *Top left: realisation x of a Strauss process on the unit square and with $(\beta, \gamma, R) = (250, 0.1, 0.05)$. Top right: union of x and a realisation y of the complementary process $Y(x)$. Bottom left: $L(r; x) - r$ corresponding to top left panel (thin line). Bottom right: $L(r; x \cup y) - r$ corresponding to top right panel (thin line). The thick lines in the two lower panels are estimated 95% Poisson envelopes.*

the following we will refer to this model as the true model. Here $n(x) = 87$ and x is a perfect simulation obtained by the dominating CFTP algorithm described in Berthelsen and Møller (2002). The bottom left panel shows a plot of $\hat{L}(r; x) - r$ compared to estimated 95% Poisson envelopes where $\underline{L}$ and $\overline{L}$ are estimated assuming $\beta = n(x)/|S|$ (the maximum likelihood estimate under the stationary Poisson process. As $\hat{L}(r; x) - r$ is well outside the 95% envelopes, this correctly indicates that x is not a realisation of a Poisson process. In fact, the dip in $\hat{L}(r; x) - r$ around $r = R$ indicates a repulsive point process.

As outlined above, our model checking consists in checking if $x \cup y$ is a realisation of a Poisson process, where y is a realisation of $Y(x)$. The top right panel in Figure 1 shows $x \cup y$ where y is a realisation of $Y(x)$ under the true model with x as in the top left panel of Figure 1; here $n(x \cup y) = 235$. The bottom right panel in Figure 1 shows $\hat{L}(r; x \cup y) - r$ together with estimated 95% Poisson envelopes. As $\hat{L}(r; x) - r$ is well within these envelopes we cannot dismiss that $x \cup y$ is a realisation of a Poisson process. In turn this implies, correctly, that we cannot reject that the assumed model is the true model.

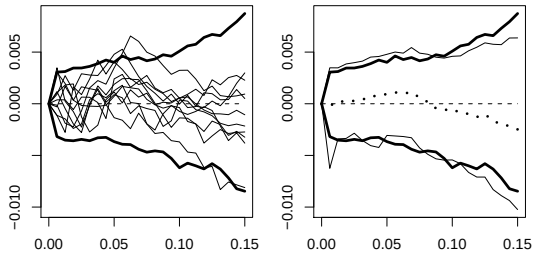

FIG 2. *Left panel shows* $\hat{L}(r; x \cup Y(x)) - r$ *for ten independent realisations of* $Y(x)$ *assuming the true model. Right panel shows estimated 95% envelopes for* $\hat{L}(r; x \cup Y(x)) - r$ *(thin lines) together with an estimate of* $\mathrm{E}[\hat{L}(r; x \cup Y(x)) - r]$ *(dotted line). The thick lines in the two panels are estimated 95% Poisson envelopes for* $\hat{L}(r; W) - r$ *when* $W \sim$ *Poisson*(S, β) *(they are identical to those in the right lower panel of Figure 1).*

Before we investigate the power of out test procedure, we notice that even with x fixed there is some variation in $\hat{L}(r; x \cup y) - r$ due to the variation in y coming from $Y(x)$. The left panel in Figure 2 shows $\hat{L}(r; x \cup y) - r$ for ten realisations of y from $Y(x)$, and estimated 95% Poisson envelopes. The right panel in Figure 2 shows estimated 95% envelopes for $\hat{L}(r; x \cup Y(x)) - r$ together with an estimate of $\mathrm{E}[\hat{L}(r; x \cup Y(x)) - r]$ and 95% Poisson envelopes. Notice that the the two sets of envelopes are a close match. This is in general not to be expected even if the estimated model equals the true model since, for a given x, $x \cup Y(x)$ is in general not distributed as Poisson(S, β).

To investigate the power of our test procedure we consider two misspecifications of the model. In model A we assume that x is a realisation of a Strauss process with $\beta = 150$, $\gamma = 0.5$, and $R = 0.05$ (i.e. incorrect β and γ but correct R), and in model B we assume that x is a realisation of a Strauss process with $\beta = 125$, $\gamma = 0.1$, and $R = 0.025$ (i.e. incorrect β and R but correct γ). Under both the true model and the two misspecified models the expected number of points in the point processes are roughly the same (confirmed by simulations).

The results obtained under model A are summarised in Figure 3. Compared to Figure 2 the realisations of $\hat{L}(r; x \cup Y(x)) - r$ are no longer centred around zero and the two pairs of envelopes in the right panel do not match. Since most of the $\hat{L}(r; x \cup Y(x)) - r$ curves in the left panel of Figure 3 are within the 95% Poisson envelopes and keeping in mind that $x \cup Y(x)$ is not distributed as Poison(S, β), Figure 3 is not a strong indication that the model specified in model A is wrong.

For model B the conclusions based on Figure 4 are much clearer. The fact that both the estimated mean and estimated 2.5% envelope for $L(r; x \cup$

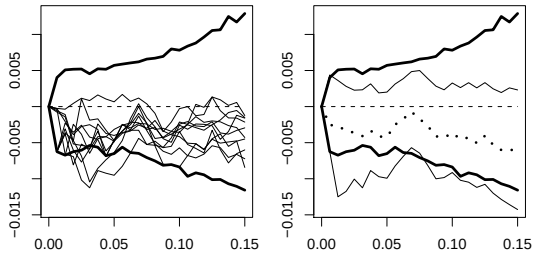

FIG 3. *As in Figure 2 but assuming* $\beta = 150$, $\gamma = 0.5$, *and* $R = 0.05$ *when generating* $Y(x)$ *and* $W \sim Poisson(S, \beta)$.

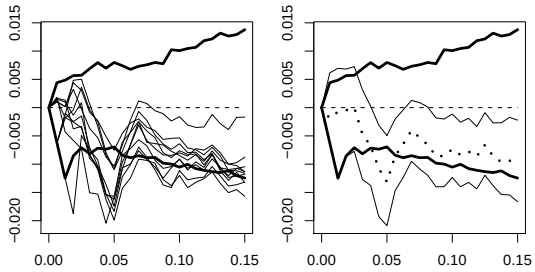

FIG 4. *As in Figure 2 but assuming* $\beta = 125$, $\gamma = 0.1$, *and* $R = 0.025$ *when generating* $Y(x)$ *and* $W \sim Poisson(S, \beta)$.

$Y(x)) - r$ are well outside the Poisson envelopes for $r \approx 0.05$ gives clear indications of a misspecified model. Further, the distinct V-shaped deviation from zero in the curves in the left panel of Figure 4 are unlikely under Poisson.

To further assess the power of our test procedure we consider two test-statistics for testing the hypothesis that a point pattern z is a realisation of a Poisson process with intensity β on S. The two test statistics are $T_1(z) = \int_0^{\tilde{r}_1} (L(r; z) - r)^2 \mathrm{d}r$ and $T_2(z) = \max_{r \in (0, \tilde{r}_2]} d(r; z) - \min_{r \in (0, \tilde{r}]} d(r; z)$, where $d(r; z) = (\hat{L}(r; z) - r)/(\overline{L}(r) - \underline{L}(r))$ and $\tilde{r}_i$ is a user-specified parameter,$i = 1, 2$. Note that T_1 captures the overall deviation from zero expected if the model is correct, and T_2 should capture large V-shaped deviations as seen in Figure 4 while taking into account that the variance of $\hat{L}(r; W) - r$ varies with r, where $W \sim \text{Poisson}(S, \beta)$. Based on 1000 realisation from Poisson(S, β) we obtain for each test statistic T_i an estimate $\hat{T}_{i,c}$ of the critical value at the 5% significance level. For each of the misspecified models A and B we have generated 1000 independent realisations $x^{(1)}, \ldots, x^{(1000)}$ of X, and for

each of realisation $x^{(j)}$ we have generated a realisation $y^{(j)}$ from $Y(x^{(j)})$. For each test statistic T_i, the estimated power is given by the fraction of the 1000 realisations where the value of $T_i(x^{(j)} \cup y^{(j)})$ is larger than $\hat{T}_{i,c}$.

For both models A and B and for both T_1 and T_2 we have used $\tilde{r}_1 = \tilde{r}_2 = 0.15$. For model A the estimated power is very low: 3.7% and 7.9%, respectively, for T_1 and T_2. This is in agreement with the conclusion above based on a visual inspection of Figure 3. For model B the estimated power is 11.4% and 47.7%, respectively, for the two test statistics. Here T_2 has a reasonable level of power, which is in accordance with the conclusion based on a visual inspection of Figure 4. We expect that it is in general difficult to find a good test statistic which captures the type of deviations in Figure 4 which we noticed above.

Remark 5. One inherent limitation of our proposed testing procedure stems from the fact that we compare $x \cup y$ to a homogeneous Poisson process with intensity β, where $b = \beta|S|$ can be far from the maximum likelihood estimate $n(x)$ under this Poisson process. Consider the case where $b \gg n(x)$, which is the case for point processes with strong interaction and dense packing, e.g. a Strauss process with $\gamma \approx 0$ and $R > 0$ combined with a high value of β. Assuming $b \gg n(x)$ implies that $\mathrm{E}[n(Y(x))] \approx b \gg n(x)$, i.e. in the union $x \cup y$ we expect the points of the complimentary point process to vastly outnumber the data points—essentially the data "drowns" in the complementary point process. Consequently the distribution of $Y(x)$ is very similar to Poisson(S, β) no matter what the true model is. Hence, the probability of rejecting a wrong model effectively equals the significance level.

Remark 6. A by-product of Algorithm 2 is that when generating i.i.d. realisations of $Y(x)$, we obtain i.i.d. realisations of X_T conditional on $X_0 = x$. If the model is correct any summary of x is expected not to be extreme compared to the same summary for realisations of X_T. A simple summary would be the number of points.

Remark 7. The conventional way to use the L function for model checking is to compare $\hat{L}(r; x) - r$ to estimated 95% envelopes for $\hat{L}(r; X) - r$ when X is distributed according to the assumed model. Obtaining the envelopes typically involves either generating i.i.d. realisations of X or subsampling from a long Markov chain converging towards the assumed model. In general both methods are computationally more expensive than generating both $Y(x)$ and Poisson(S, β) used in our test. Hence, our approach has a computational advantage compared to the more conventional approach for model checking.

Acknowledgements. We thank Ege Rubak and Frederic R. Paik Schoenberg for helpful discussions.

Appendix A. For $n = 0, 1, \ldots$, define $\Omega_n = \{x \subset S : n(x) = n\}$, recalling that $n(x)$ is the cardinality of x. So for any event $F_n \subseteq \Omega_n$, if $n \geq 1$,

$$\nu(F_n) = \frac{\mathrm{e}^{-|S|}}{n!} \int_S \cdots \int_S 1[\{x_1, \ldots, x_n\} \in F_n] \, \mathrm{d}x_1 \cdots \mathrm{d}x_n \tag{14}$$

where $1[\cdot]$ is the indicator function. Moreover, $\nu(F_0) = 1[\emptyset \in F_0]$ if $F_0 \subseteq \Omega_0$, i.e. when either F_0 is empty or it is the set consisting of the empty point configuration $\emptyset$.

Proof of Proposition 1. By assumption X is absolutely continuous with respect to ν, so it suffices to verify that for any point configuration $x \in \Omega_m$, event $F_n \subseteq \Omega_n$, and non-negative integers m and n, $\mathrm{P}(Y \in F_n | X = x) = 0$ if $\nu(F_n) = 0$. This follows immediately from (14) and Theorem 1.

Proof of Proposition 2. For any events $F_m \subseteq \Omega_m$ and $F_n \subseteq \Omega_n$, with $m, n = 0, 1, \ldots$, the total rate of moving away from any state in $F_m \times F_n$ is $b+m+n$; the mean of the total rate of moving into $F_m \times F_n$ by a birth in the X_t process is

$$\begin{aligned}
&G_1(F_m \times F_n) \\
&= \int_S \cdots \int_S \sum_{i=1}^{m} 1[\{x_1, \ldots, x_m\} \in F_m, \, \{y_1, \ldots, y_n\} \in F_n] \\
&\quad \lambda(\{x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_m\}, x_i) \pi(\{x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_m\}, \{y_1, \ldots, y_n\}) \\
&\quad \frac{\mathrm{e}^{-2|S|}}{m!n!} \, \mathrm{d}x_1 \cdots \mathrm{d}x_m \, \mathrm{d}y_1 \cdots \mathrm{d}y_n
\end{aligned}$$

(setting $G_1(F_m \times F_n) = 0$ if $m = 0$); the mean of the total rate of moving into $F_m \times F_n$ by a birth in the Y_t process is

$$\begin{aligned}
&G_2(F_m \times F_n) \\
&= \int_S \cdots \int_S \sum_{i=1}^{n} 1[\{x_1, \ldots, x_m\} \in F_m, \, \{y_1, \ldots, y_n\} \in F_n] \\
&\quad [\beta(y_i) - \lambda(\{x_1, \ldots, x_m\}, x_i)] \\
&\quad \pi(\{x_1, \ldots, x_m\}, \{y_1, \ldots, y_{i-1}, y_{i+1}, \ldots, y_n\}) \frac{\mathrm{e}^{-2|S|}}{m!n!} \, \mathrm{d}x_1 \cdots \mathrm{d}x_m \, \mathrm{d}y_1 \cdots \mathrm{d}y_n
\end{aligned}$$

(setting $G_2(F_m \times F_n) = 0$ if $n = 0$); the mean of the total rate of moving into $F_m \times F_n$ by a death in the X_t process is

$$\begin{aligned} &G_3(F_m \times F_n) \\ &= \int_S \cdots \int_S \sum_{i=1}^{m+1} 1[\{x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_{m+1}\} \in F_m, \{y_1, \ldots, y_n\} \in F_n] \\ &\quad \pi(\{x_1, \ldots, x_{m+1}\}, \{y_1, \ldots, y_n\}) \frac{\mathrm{e}^{-2|S|}}{(m+1)!n!} \, \mathrm{d}x_1 \cdots \mathrm{d}x_{m+1} \, \mathrm{d}y_1 \cdots \mathrm{d}y_n; \end{aligned}$$

and the mean of the total rate of moving into $F_m \times F_n$ by a death in the Y_t process is

$$\begin{aligned} &G_4(F_m \times F_n) \\ &= \int_S \cdots \int_S \sum_{i=1}^{n+1} 1[\{x_1, \cdots, x_m\} \in F_m, \{y_1, \ldots, y_{i-1}, y_{i+1}, \ldots, y_{n+1}\} \in F_n] \\ &\quad \pi(\{x_1, \ldots, x_m\}, \{y_1, \ldots, y_{n+1}\}) \frac{\mathrm{e}^{-2|S|}}{m!(n+1)!} \, \mathrm{d}x_1 \cdots \mathrm{d}x_m \, \mathrm{d}y_1 \cdots \mathrm{d}y_{n+1}. \end{aligned}$$

Consequently, by Proposition 8.1 in Preston (1977), the equilibrium distribution Π is the unique distribution satisfying

$$(b+m+n)\Pi(F_m \times F_n) = G_1(F_m \times F_n) + G_2(F_m \times F_n) + G_3(F_m \times F_n) + G_4(F_m \times F_n)$$

for all events $F_m \subseteq \Omega_m$ and $F_n \subseteq \Omega_n$, with $m, n = 0, 1, \ldots$. This is seen to be equivalent to the statement in Proposition 2, since

$$\begin{aligned} \Pi(F_m \times F_n) = &\int_S \cdots \int_S 1[\{x_1, \ldots, x_m\} \in F_m, \{y_1, \ldots, y_n\} \in F_n] \\ &\pi(\{x_1, \ldots, x_m\}, \{y_1, \ldots, y_n\}) \frac{\mathrm{e}^{-2|S|}}{m!n!} \, \mathrm{d}x_1 \cdots \mathrm{d}x_m \, \mathrm{d}y_1 \cdots \mathrm{d}y_n. \end{aligned}$$

Proof of Proposition 3. The "if"-part is easily verified, since then

$$f(x) = \exp\left(-\int_S \lambda(u) \, \mathrm{d}u\right) \prod_{u \in x} \lambda(u)$$

is the density of X, and

$$g(y) = \exp\left(b - \int_S \lambda(u) \, \mathrm{d}u\right) \prod_{u \in y} (\beta(u) - \lambda(u))$$

is the density of Y. So suppose that (5) holds. The first equation in (5) implies that $f(x)\lambda(x, u) = f(x \cup \{u\})$, which is clearly satisfied and just

means that X has density f. Thus $g(y|x) = \pi(x,y)/f(x)$ is the conditional density of Y given $X = x$ (when $f(x) > 0$), and the first equation in (5) gives that

$$g(y|x) = g(y|x \cup \{u\}) \quad \text{whenever } f(x \cup \{u\}) > 0,$$

meaning that X and Y are independent and $g(y|x) = g(y)$ does not depend on x. The second equation in (5) is then equivalent to

$$g(y)(\beta(u) - \lambda(x,u)) = g(y \cup \{u\}),$$

so $\lambda(x,u) = \lambda(u)$ does not depend on x. Consequently, by induction,

$$f(x) \propto \prod_{u \in x} \lambda(u), \quad g(y) \propto \prod_{u \in y} (\beta(u) - \lambda(u)),$$

whereby also the "only-if"-part is verified.

Example 1. Consider the very simple case with $\lambda(\emptyset, u) = \beta > 0$ and $\lambda(x,u) = 0$ whenever $n(x) > 0$, that is,

$$f(\emptyset) = \mathrm{e}^{|S|}/(1+b), \quad f(\{u\}) = \beta \mathrm{e}^{|S|}/(1+b), \quad f(x) = 0 \quad \text{whenever } n(x) > 1,$$

meaning that $n(X) \leq 1$ and with probability $b/(1+b)$, $n(X) = 1$, in which case X consists of a uniformly distributed point in S. Note that $b = \beta|S|$, and by Algorithm 1 and Theorem 1, conditional on $X = x$, the points in $Y(x)$ are independent and uniformly distributed in S. Thus conditional on $(n(X), n(Y)) = (m,n)$, the $m+n$ points in X and Y are independent and uniformly distributed in S. So the joint distribution of X and Y is effectively given by the distribution of $(n(X), n(Y))$. Defining

$$\pi_{m,n} = \mathrm{P}(n(X) = m, n(Y) = n), \quad m,n \in \{0,1,\ldots\},$$

we have for any (x,y) with $(n(x), n(y)) = (m,n)$,

$$\pi_{m,n} = \frac{|S|^{m+n}\mathrm{e}^{-2|S|}}{m!n!}\pi(x,y)$$

and $\pi_{m,n} = 0$ if $m \geq 2$. Hence (4) is seen to be equivalent to

$$\pi_{0,0} = \mathrm{e}^{-b}, \quad \pi_{1,n} = \frac{b^{n+1}}{(n+1)!}\mathrm{e}^{-b} - \pi_{0,n+1}, \quad (b+n)\pi_{0,n} = \pi_{1,n} + (n+1)\pi_{0,n+1},$$

where the two first equations follow from the fact that $n(X)+n(Y)$ is Poisson distributed with parameter b, and the last equation follows since (4) gives

$$(b+n)\frac{0!n!}{|S|^n \mathrm{e}^{2|S|}}\pi_{0,n} = 0 + 0 + |S|\frac{1!n!}{|S|^{n+1}\mathrm{e}^{2|S|}}\pi_{1,n} + |S|\frac{0!(n+1)!}{|S|^{n+1}\mathrm{e}^{2|S|}}\pi_{0,n+1}.$$

Consequently, the $\pi_{m,n}$ are determined by $\pi_{0,1}$, since $\pi_{1,n}$ is determined by $\pi_{0,n+1}$ for $n = 0, 1, \ldots$, and $\pi_{0,n+1}$ is determined by $\pi_{0,1}$ for $n = 1, 2, \ldots$, since

$$\pi_{0,n+1} = \frac{1}{n}\left[(b+n)\pi_{0,n} - \frac{b^{n+1}}{(n+1)!}\mathrm{e}^{-b}\right], \quad n = 1, 2, \ldots. \tag{15}$$

Using induction, it follows easily from (15) that

$$\pi_{0,n+1} = \left[\prod_{i=1}^{n}\frac{b+i}{i}\right]\left[\pi_{0,1} - \mathrm{e}^{-b}\sum_{i=2}^{n+1}\frac{b}{i(i-1)}\prod_{j=1}^{i-1}\frac{b}{b+j}\right], \quad n = 1, 2, \ldots. \tag{16}$$

which can be rewritten as

$$\begin{aligned}\pi_{0,n+1} = \frac{\Gamma(b+n+1)}{\Gamma(n+1)\Gamma(b+1)}\Bigg[\pi_{0,1} - b^{-b}\left(\Gamma(b+1,b) - \frac{\Gamma(b+n+1,b)\Gamma(b+1)}{\Gamma(b+n+1)}\right)\\ - \mathrm{e}^{-b}b\left(1 - \frac{b^n\Gamma(b+1)}{(n+1)\Gamma(b+n+1)}\right)\Bigg], \quad n = 1, 2, \ldots,\end{aligned} \tag{17}$$

where $\Gamma(a,x) = \int_x^\infty t^{a-1}\mathrm{e}^{-t}\,\mathrm{d}t$ is the incomplete gamma function which has the property that $\Gamma(a,x) = (a-1)\Gamma(a-1,x) + x^{a-1}\mathrm{e}^{-x}$.

As $n \to \infty$ we have that $\pi_{0,n+1} \to 0$. Since $\prod_{i=1}^{n}(b+i)/i \to \infty$ as $n \to \infty$ equation (16) implies that

$$\pi_{0,1} = \mathrm{e}^{-b}\sum_{i=2}^{\infty}\frac{b^i}{i(i-1)}\frac{\Gamma(b+1)}{\Gamma(b+i)}.$$

Using similar argument but taking (17) as the starting point we obtain

$$\pi_{0,1} = b^{-b}(\Gamma(b+1,b) - \Gamma(b+1)) + \mathrm{e}^{-b}b.$$

Inserting this in equation (17) and noting that $\mathrm{P}(n(Y) = n) = \pi_{0,n} + \pi_{1,n} = \pi_{0,n} + \frac{b^{n+1}}{(n+1)!}\mathrm{e}^{-b} - \pi_{0,n+1}$ we find that the marginal distribution of $n(Y)$ is given by

$$\mathrm{P}(n(Y) = n) = \frac{b^{1-b}\left(\Gamma(b+n) - \Gamma(b+n,b)\right)}{\Gamma(n+1)}, \quad n = 1, 2, \ldots.$$

Example 2. The case above extends to when $\lambda(x,u) = \mu_{n(x)}$ depends only on the number of points in x, where for $n = 0, 1, \ldots$, $0 \le \mu_n \le \beta$ and if $\mu_n = 0$ then $\mu_{n+1} = 0$. Again, conditional on $(n(X), n(Y)) = (m,n)$, the

$m+n$ points in X and Y are independent and uniformly distributed in S, and solving (4) becomes equivalent to solve

$$(b+m+n)\pi_{m,n} = \lambda_{m-1}\pi_{m-1,n}+(b-\lambda_m)\pi_{m,n-1}+(m+1)\pi_{m+1,n}+(n+1)\pi_{m,n+1}$$

for $m,n = 1,2,\ldots$, where $\lambda_m = |S|\mu_m$, $\pi_{-1,n} = 0$, and $\pi_{m,-1} = 0$. It follows by induction that the $\pi_{0,n}$, $n = 0,1,\ldots$, determine all the $\pi_{m,n}$, $m,n = 1,2,\ldots$. We know that $\pi_{0,0} = \mathrm{e}^{-b}$ but in general, for $n \geq 1$, we do not have a simple recursion for the $\pi_{0,n}$; this is in contrast to (15). So finding an expression for the $\pi_{0,n}$ seems now to be a much harder problem.

In conclusion, apart from the rather trivial case where X and Y are independent Poisson processes (see Proposition 3), the joint distribution of X and Y seem to be complicated. Furthermore, apart from simple cases (such as Example 1) the marginal distribution of Y seem also to be very complicated.

References.

Berthelsen, K. K. and Møller, J. (2002). A primer on perfect simulation for spatial point processes. *Bulletin of the Brazilian Mathematical Society* **33** 351–367.

Besag, J. (1977). Some methods of statistical analysis for spatial data. *Bulletin of the International Statistical Institute* **47** 77–92.

Carter, D. S. and Prenter, P. M. (1972). Exponential spaces and counting processes. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* **21** 1-19.

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

Fiksel, T. (1984). Simple spatial-temporal models for sequences of geological events. *Elektronische Informationsverarbeitung und Kypernetik* **20** 480–487.

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

Illian, J., Penttinen, A., Stoyan, H. and Stoyan, D. (2008). *Statistical Analysis and Modelling of Spatial Point Patterns*. John Wiley and Sons, Chichester.

Kelly, F. P. and Ripley, B. D. (1976). A note on Strauss' model for clustering. *Biometrika* **63** 357-360.

Kendall, W. S. (1998). Perfect simulation for the area-interaction point process. In *Probability Towards 2000* (L. Accardi and C. C. Heyde, eds.) 218–234. Springer Lecture Notes in Statistics 128, Springer Verlag, New York.

Kendall, W. S. and Møller, J. (2000). Perfect simulation using dominating processes on ordered spaces, with application to locally stable point processes. *Advances in Applied Probability* **32** 844–865.

Møller, J. (1989). On the rate of convergence of spatial birth-and-death processes. *Annals of the Institute of Statistical Mathematics* **3** 565–581.

Møller, J. and Schoenberg, R. P. (2010). Thinning spatial point processes into Poisson processes. *Advances in Applied Probability* **42** 347–358.

Møller, J. and Sørensen, M. (1994). Parametric models of spatial birth-and-death processes with a view to modelling linear dune fields. *Scandinavian Journal of Statistics* **21** 1–19.

Møller, J. and Waagepetersen, R. P. (2004). *Statistical Inference and Simulation for Spatial Point Processes*. Chapman and Hall/CRC, Boca Raton.

Preston, C. J. (1977). Spatial birth-and-death processes. *Bulletin of the International Statistical Institute* **46** 371–391.

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

Ripley, B. D. (1977). Modelling spatial patterns (with discussion). *Journal of the Royal Statistical Society Series B* **39** 172–212.

Ripley, B. D. and Kelly, F. P. (1977). Markov point processes. *Journal of the London Mathematical Society* **15** 188-192.

Stephens, M. (2000). Bayesian analysis of mixture models with an unknown number of components–an alternative to reversible jump methods. *Annals of Statistics* **28** 40–74.

Stoyan, D., Kendall, W. S. and Mecke, J. (1995). *Stochastic Geometry and Its Applications*, Second ed. Wiley, Chichester.

Strauss, D. J. (1975). A model for clustering. *Biometrika* **63** 467–475.

Department of Mathematical Sciences
Aalborg University
Fredrik Bajers Vej 7G
9220 Aalborg Øst,
Denmark
E-mail: jm@math.aau.dk; kkb@math.aau.dk