Journal of Statistical Software

November 2013, Volume 55, Issue 11. *http://www.jstatsoft.org/*

Hybrids of Gibbs Point Process Models and Their Implementation

Adrian Baddeley
University of Western Australia

Rolf Turner
University of Auckland

Jorge Mateu
Universitat Jaume I

Andrew Bevan
University College London

Abstract

We describe a simple way to construct new statistical models for spatial point pattern data. Taking two or more existing models (finite Gibbs spatial point processes) we multiply the probability densities together and renormalise to obtain a new probability density. We call the resulting model a hybrid. We discuss stochastic properties of hybrids, their statistical implications, statistical inference, computational strategies and software implementation in the R package **spatstat**. Hybrids are particularly useful for constructing models which exhibit interaction at different spatial scales. The methods are demonstrated on a real data set on human social interaction. Software and data are provided.

Keywords: spatial point processes, **spatstat**, R, C, maximum pseudolikelihood, Berman-Turner device, Markov chain Monte Carlo, birth-death-shift Metropolis-Hastings algorithm, Gordon Square data, social interaction.

1. Introduction

In the statistical analysis of spatial point pattern data, a parametric modeling approach requires a supply of stochastic models for point patterns. One important source of models is the class of (finite) Gibbs point processes. Many models of this class can be fitted rapidly to real data sets containing large numbers of points (Baddeley and Turner 2000, 2006; Diggle 2003; Geyer 1999; Møller and Waagepetersen 2004). Model-fitting, prediction and simulation of Gibbs models are supported in the **spatstat** package (Baddeley and Turner 2005) for R (R Core Team 2013).

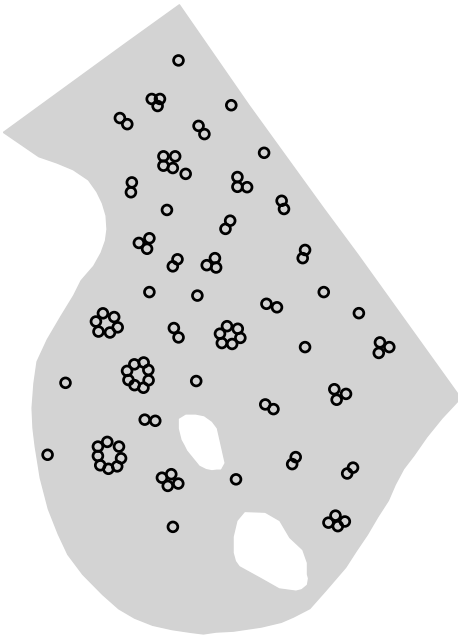

Figure 1: Gordon Square data. Locations of 99 people (circles) sitting on grass (gray shading) in Gordon Square, London, UK on a sunny afternoon. Data collected by the authors.

However, in practice, Gibbs models currently available for use in data analysis are few in number and limited in scope. The available models for interpoint interaction tend to have simple mathematical structure, which simplifies theoretical study and software coding, but makes them unrealistic in applications. The most commonly used Gibbs models exhibit spatial interaction at only a single spatial scale, whereas most natural processes exhibit dependence at multiple scales. This has motivated statisticians to construct 'multi-scale' generalizations of the classical Gibbs models (Ambler 2002; Picard, Bar-hen, Mortier, and Chadoeuf 2009). The most commonly known example is the multi-scale pairwise interaction model (Diggle 1983, Section 4.9).

One application of multi-scale analysis, which we shall consider, arises in the study of human social interaction. Figure 1 shows the spatial locations of people sitting on the grass in a park on a sunny afternoon. The pattern appears to show spatial organization at several different scales.

It would be possible to model the point pattern illustrated in Figure 1 as a realization of a cluster process, but this would not be a Poisson cluster process, so there is an inescapable need to model the interactions between points at some level, and the natural models for interactions are Gibbs processes.

The key problem is that the construction of new Gibbs models *ab initio* is not trivial. Gibbs models are usually defined using the unnormalized probability density h. While it is easy to write down a new functional form for h, it is not always obvious whether h is integrable (so that the normalizing constant is finite and the process is well-defined). A famous example is the Strauss process (Strauss 1975; Kelly and Ripley 1976). In addition to the prerequisite of integrability, h would often be required to satisfy further conditions such as local stability, to

ensure geometric convergence of the Metropolis-Hastings algorithm (Geyer and Møller 1994; Møller and Waagepetersen 2004). Despite Geyer's assertion that it is "easy to invent new processes and do statistical inference for them" (Geyer 1999, p. 110), it was not straightforward to prove that his two examples of new processes were locally stable (Geyer 1999, p. 111). Application of newly-invented point process models to data requires, at a very minimum, software implementation of algorithms for parameter estimation and simulation, and usually also requires research to ensure the validity and efficiency of these algorithms. Meaningful interpretation of the model is a further challenge.

This paper explores a simple, practical technique for constructing new Gibbs models from existing ones. The unnormalized densities h_1, h_2 of two existing models are multiplied together, to form a new 'hybrid' unnormalized density $h(\mathbf{x}) = h_1(\mathbf{x})h_2(\mathbf{x})$. Equivalently the energy potentials $U = \log h$ are combined to form a hybrid potential $\log h(\mathbf{x}) = \log h_1(\mathbf{x}) + \log h_2(\mathbf{x})$. We explore the practical use of hybridization as a way to construct new point process models to be fitted to data.

In principle $h_1 h_2$ might not define a point process at all, since it might not be integrable. We state some conditions which guarantee that the hybrid point process exists.

In this paper we show that the hybrid $h_1 h_2$ of two unnormalized Gibbs densities h_1, h_2 preserves many desirable stochastic properties, including the Gibbs and Markov properties, local and Ruelle stability, and monotonicity of the point process density. From a statistical viewpoint, hybridization is also very natural, particularly for exponential family models, where it preserves properties such as the log-concavity of the likelihood and pseudolikelihood.

The main advantage of our approach is that we can re-use *existing* software, for statistical inference and simulation of basic point process models, to provide the same functionality for hybrids of these basic models. Existing algorithms for fitting and simulating Gibbs point process models require minor modification in order to fit and simulate hybrids of the same models. With an appropriate software design, the components of the hybrid can be specified dynamically. This means the hybrid model can be built interactively by the user. Thus, hybridization provides a quick method for constructing new Gibbs models, that is flexible enough for modeling real data sets.

The R package **spatstat** (Baddeley and Turner 2005), available from the Comprehensive R Archive Network at `http://CRAN.R-project.org/package=spatstat`, supports a wide range of statistical methods for spatial point pattern data. We have recently extended the syntax of models in **spatstat** so that hybrid models can be specified interactively, and extended the computational infrastructure to handle hybrid models. The software implementation is described in Sections 7 and 9 of this paper.

It has become widely accepted in spatial statistics that Gibbs models are suitable primarily for modeling regularity or inhibition between points, and are not able to model strong clustering (Illian, Penttinen, Stoyan, and Stoyan 2008, pp. 138, 157, 171; Møller and Waagepetersen 2004) although moderate clustering can be achieved. However, this conclusion is based on experience with the simple Gibbs models that have predominated in the literature, in particular with *pairwise interaction* processes. We would argue that the much wider class of hybrid Gibbs models should be able to produce a wide variety of spatial patterns, including the combination of quite strong clustering and inhibition seen in Figure 1.

Hybrids are particularly useful for modeling interactions at multiple scales. In fact the literature already contains several examples. The multi-scale generalization of the Strauss process

– namely, a pairwise interaction process with a step function potential (Diggle 1983, Section 4.9; Diggle 2003, p. 109; Heikkinen and Penttinen 1999) – is a hybrid of several Strauss processes. Ambler and Silverman's (Ambler 2002; Ambler and Silverman 2004; Picard *et al.* 2009) multi-scale generalization of the area-interaction model (Widom and Rowlinson 1970; Baddeley and van Lieshout 1995) is a hybrid of area-interaction potentials. However, generalizing from a single-scale to a multi-scale process is not a trivial procedure. In this paper we show that one plausible multi-scale generalization of Geyer's saturation process (Geyer 1999) fails to have the desired properties. However, the hybrid of several Geyer saturation processes is guaranteed to be well-behaved, by our theoretical results (Section 4).

Section 2 gives some basic background and definitions. Section 3 defines hybrids of point process densities, and offers several statistical interpretations. Section 4 gives basic technical results about hybrids. Section 5 discusses multi-scale processes. Section 6 discusses statistical inference for parametric models, and Section 7 describes our software implementation of model-fitting. Section 8 discusses simulation of hybrid models, and Section 9 describes our software implementation of simulation algorithms. Section 10 reports a simulation experiment on the performance of our model-fitting techniques. Section 11 demonstrates an analysis of the data in Figure 1. We conclude with a discussion in Section 12.

2. Background

This section gives some basic definitions for spatial point processes.

2.1. Point processes and densities

A *spatial point pattern* data set $\mathbf{x}$ is an unordered set

$$\mathbf{x} = \{x_1, \ldots, x_n\}, \quad n \geq 0,\ x_i \in W$$

of points x_i in a spatial 'window' W in d-dimensional space $\mathbb{R}^d$, where $d \geq 1$. The window W is assumed to have finite positive volume $|W|$. Write $n(\mathbf{x}) = n$ for the number of points in the pattern $\mathbf{x}$; note that $n(\mathbf{x})$ is not fixed, and may be zero.

A (finite) *spatial point process* $\mathbf{X}$ in W is, briefly speaking, a random element of the space $\mathcal{X}$ of all point patterns in W, subject to measurability requirements (Møller and Waagepetersen 2004). Intuitively the number of points $N = n(\mathbf{X})$ is a random variable, and conditional on $N = n$, the points of the pattern $\mathbf{X}$ are random d-dimensional points in W with a certain joint distribution. The distribution of $\mathbf{X}$ is completely determined by knowledge of the distributions, or even just the expected values, of the random variables $g(\mathbf{X})$ for any measurable functional $g : \mathcal{X} \to [0, \infty)$.

In formulating stochastic models, the reference model will be the Poisson point process with intensity 1 on W, which will be denoted by $\mathbf{Z}$. The expectation of $g(\mathbf{Z})$ for a measurable functional $g : \mathcal{X} \to [0, \infty)$ is, from first principles,

$$\mathsf{E}[g(\mathbf{Z})] = e^{-|W|} g(\emptyset) + e^{-|W|} \sum_{n=1}^{\infty} \frac{1}{n!} \int_W \cdots \int_W g(\{x_1, \ldots, x_n\})\, \mathrm{d}x_1\, \mathrm{d}x_2 \cdots\, \mathrm{d}x_n. \tag{1}$$

A point process $\mathbf{X}$ will be said to have *probability density* $f(\mathbf{x})$ (with respect to the unit rate Poisson process) if

$$\mathsf{E}[g(\mathbf{X})] = \mathsf{E}[g(\mathbf{Z}) f(\mathbf{Z})] \tag{2}$$

for any measurable integrand g, where the density f is a measurable function $f : \mathcal{X} \to [0, \infty)$, which is normalized in the sense that

$$\mathsf{E}[f(\mathbf{Z})] = 1. \tag{3}$$

Any such density determines the probability distribution of a spatial point process on W. For further details, see Møller and Waagepetersen (2004).

For example, the uniform Poisson process of intensity $\beta > 0$ on W has probability density

$$f(\mathbf{x}) = \alpha\beta^{n(\mathbf{x})},$$

where the normalizing constant is $\alpha = \exp((1 - \beta)|W|)$. It can be verified directly that this function satisfies (3).

2.2. Unnormalized densities

In formulating new models, it is common to write down an *unnormalized* density $h(\mathbf{x})$, a nonnegative measurable function which must be normalized in order to serve as a probability density (Geyer 1999; Ripley and Kelly 1977). Desirable properties of the unnormalized density include the following.

Definition 1 *An unnormalized density h is said to be*

integrable *if $\mathsf{E}[h(\mathbf{Z})] < \infty$;*

Ruelle stable *if there are finite constants A and M such that $h(\mathbf{x}) \le AM^{n(\mathbf{x})}$;*

locally stable *if there is a finite constant B such that $h(\mathbf{x} \cup \{u\}) \le Bh(\mathbf{x})$ for all $\mathbf{x} \in \mathcal{X}$ and $u \in W$;*

hereditary *(or has hereditary positivity) if, for any configuration $\mathbf{x} \in \mathcal{X}$, $h(\mathbf{x}) > 0$ implies $h(\mathbf{y}) > 0$ for all sub-configurations $\mathbf{y} \subset \mathbf{x}$.*

See Ruelle (1969); Geyer (1999), Møller and Waagepetersen (2004, p. 83–84).

Integrability is an essential condition for constructing a point process. If h is integrable and not identical to zero, then h defines a finite point process, since $f(\mathbf{x}) = \alpha h(\mathbf{x})$ is a point process density, where $\alpha = 1/\mathsf{E}[h(\mathbf{Z})]$. If h is not integrable then it cannot be normalized and does not define a point process.

Ruelle stability or local stability are necessary conditions for the validity of popular Monte Carlo simulation algorithms (Geyer 1999; Møller and Waagepetersen 2004).

If h is Ruelle stable, then it is integrable, since it is dominated by a Poisson density:

$$\mathsf{E}[h(\mathbf{Z})] \le \mathsf{E}[AM^{n(\mathbf{Z})}] = A\exp((M - 1)|W|) < \infty.$$

If h is locally stable, then it is hereditary, Ruelle stable (since $h(\mathbf{x}) \le h(\emptyset)B^{n(\mathbf{x})}$ by induction), and integrable.

Definition 2 *A finite point process whose probability density f is hereditary, is called a (finite)* Gibbs *point process. For a Gibbs process, the (Papangelou)* conditional intensity *at a location $u \in W$ given a configuration $\mathbf{x} \in \mathcal{X}$ may be defined as*

$$\lambda(u, \mathbf{x}) = \frac{f(\mathbf{x} \cup \{u\})}{f(\mathbf{x} \setminus \{u\})} \tag{4}$$

provided the numerator and denominator are nonzero, and $\lambda(u, \mathbf{x}) = 0$ otherwise.

The conditional intensity has important stochastic interpretations and is useful in model-building, model-fitting and simulation.

Notice that, if we start from an unnormalized density h that is hereditary and integrable, then normalize it to obtain the probability density $f(\mathbf{x}) = \alpha h(\mathbf{x})$, the conditional intensity (4) does not involve the normalizing constant α:

$$\lambda(u, \mathbf{x}) = \frac{f(\mathbf{x} \cup \{u\})}{f(\mathbf{x} \setminus \{u\})} = \frac{h(\mathbf{x} \cup \{u\})}{h(\mathbf{x} \setminus \{u\})}.$$

Thus, local stability of h is equivalent to local stability of f, which in turn is equivalent to saying that the conditional intensity is uniformly bounded by a constant.

2.3. Examples

Apart from the Poisson point process, the most common models considered in the literature are *pairwise interaction processes* of the general form

$$h(\mathbf{x}) = \left[\prod_{i=1}^{n} b(x_i)\right] \left[\prod_{i<j} c(x_i, x_j)\right], \tag{5}$$

where $b : W \to [0, \infty)$ and $c : W \times W \to [0, \infty)$ are given functions. The second bracketed term is a product over all unordered pairs of points in $\mathbf{x}$. If b is constant and c is translation-equivariant, i.e., if $c(u, v) = c(u - v)$, then we shall call the point process *stationary* (although this is a slight abuse of terminology because the point process is restricted to a bounded region.)

A simple special case of (5) is the *Strauss process* (Strauss 1975), where $b(u) \equiv \beta > 0$ and

$$c(u, v) = \begin{cases} \gamma & \text{if } \|u - v\| < r \\ 1 & \text{otherwise} \end{cases} \tag{6}$$

where β, γ and r are parameters. It follows that

$$h(\mathbf{x}) = \beta^{n(\mathbf{x})} \gamma^{s(\mathbf{x})}, \tag{7}$$

where $s(\mathbf{x}) = s(\mathbf{x}; r)$ is the number of unordered pairs of distinct points in $\mathbf{x}$ which lie closer than r units apart. Kelly and Ripley (1976) showed that h is integrable iff $\gamma \leq 1$. In this case, h is Ruelle stable since $h(\mathbf{x}) \leq \beta^{n(\mathbf{x})}$. The conditional intensity is

$$\lambda(u, \mathbf{x}) = \beta \, \gamma^{t(u, \mathbf{x})}, \tag{8}$$

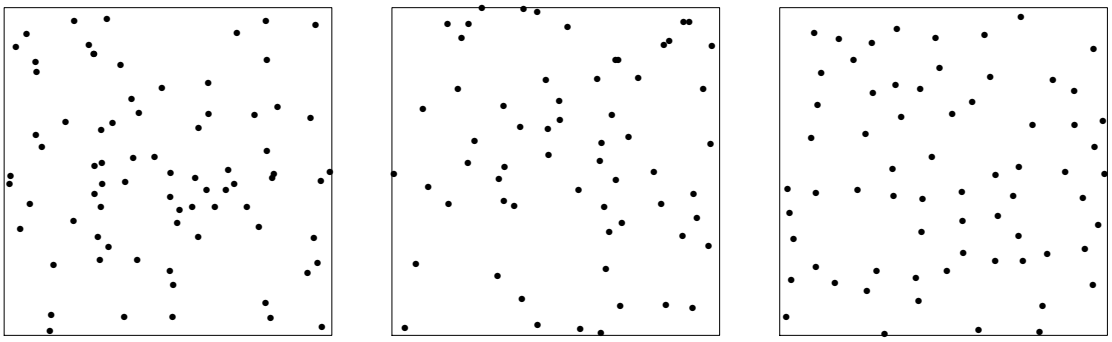

Figure 2: Simulated realizations of the Strauss process with interaction parameter γ equal to 1 (*Left*, equivalent to Poisson process), 0.5 (*Middle*) and 0 (*Right*, equivalent to hard core process) and interaction range $r = 0.07$ in the unit square. Expected number of points equal to 70 in each panel.

where $t(u, \mathbf{x}) = s(\mathbf{x} \cup \{u\}) - s(\mathbf{x} \setminus \{u\})$ is the number of points of $\mathbf{x}$ that lie within a distance r of the location u. If $\gamma \leq 1$ then $\lambda(u, \mathbf{x}) \leq \beta$ so that h is also locally stable.

In the special case of a Strauss process with $\gamma = 0$, if we interpret $0^0 = 1$, then (7) defines a point process $\mathbf{X}$ in which $s(\mathbf{X}) = 0$ with probability 1, known as the *'hard core process'*. In a hard core process, no points ever come closer to each other than the threshold of r units. This is equivalent to conditioning a Poisson point process in W of intensity β on the event that $s(\mathbf{X}) = 0$.

Figure 2 shows simulated realizations of the Strauss process with different values of γ with approximately the same numbers of points in each case. The left panel, with $\gamma = 1$, is a Poisson process, where the point locations are completely random. The right panel, with $\gamma = 0$, is a hard core process, where the points are constrained never to come closer than the threshold distance. The middle panel, with $\gamma = 0.5$, is intermediate between the two other cases. The parameter β was adjusted (using the Poisson-saddlepoint approximation of Baddeley and Nair 2012a,b) to achieve roughly equal numbers of points in each panel.

Returning for a moment to the general pairwise interaction process (5), observe that the conditional intensity is

$$\lambda(u, \mathbf{x}) = b(u) \prod_i c(u, x_i).$$

A sufficient condition for local stability of (5) is that the function c is bounded above by 1. However this condition is quite restrictive, since it implies that the process must exhibit regularity (inhibition) at all spatial scales. The Jones (1924) pairwise interaction process is Ruelle stable but not locally stable (Møller and Waagepetersen 2004, p. 88).

Gibbs processes are commonly used to model inhibitory (regular) patterns, as it is more difficult to construct locally-stable Gibbs models which exhibit attraction (clustering). Notable exceptions that are locally stable even in the attractive case include the Widom-Rowlinson (Widom and Rowlinson 1970; Rowlinson 1980) penetrable spheres model or 'area-interaction' process (Baddeley and van Lieshout 1995), the Geyer (1999) saturation process, continuum random cluster process (Klein 1982), quermass-interaction processes (Kendall, van Lieshout, and Baddeley 1999) and shot-noise weighted processes (van Lieshout and Molchanov 1998).

The stationary Widom-Rowlinson penetrable spheres model or 'area-interaction' process with parameters β, κ, r has unnormalized density

$$h(\mathbf{x}) = \beta^{n(\mathbf{x})} \exp(-\kappa U(\mathbf{x}, r)), \tag{9}$$

where $U(\mathbf{x}, r) = |B(\mathbf{x}, r)|$ is the d-dimensional volume of the set

$$B(\mathbf{x}, r) = W \cap \bigcup_{i=1}^{n(\mathbf{x})} b(x_i, r) \tag{10}$$

formed by taking the union of spheres $b(x_i, r)$ of radius $r > 0$ centered at the points x_i of the configuration $\mathbf{x}$, and restricting the union to the window W. Since $0 \le U(\mathbf{x}) \le |W|$, the density (9) is integrable for all values of $\beta > 0$ and $\kappa \in \mathbb{R}$. If $\kappa < 0$, the density favors configurations where $U(\mathbf{x})$ is large, so that the process is regular or inhibitory; while if $\kappa > 0$, the density favors configurations where $U(\mathbf{x})$ is small, so that the process is clustered or attractive.

The stationary Geyer (1999) saturation process with parameters β, γ, r, s has unnormalized density

$$h(\mathbf{x}) = \beta^{n(\mathbf{x})} \prod_{i=1}^{n(\mathbf{x})} \gamma^{\min(s, t(x_i, \mathbf{x} \setminus x_i))}, \tag{11}$$

where $t(x_i, \mathbf{x} \setminus x_i)$ is the number of neighbors of x_i in $\mathbf{x}$, that is, the number of points x_j with $j \neq i$ such that $\|x_i - x_j\| \le r$. The parameter $s > 0$ is a saturation threshold which ensures that each term in the product is bounded by γ^s, so that the density is integrable and Ruelle stable for all values of $\gamma > 0$. A geometrical argument (Geyer 1999) establishes that the density is locally stable. The process is clustered if $\gamma > 1$.

The stationary triplet interaction process (Geyer 1999) with parameters β, γ, r has unnormalized density

$$h(\mathbf{x}) = \beta^{n(\mathbf{x})} \gamma^{v(\mathbf{x})}, \tag{12}$$

where $v(\mathbf{x}) = v(\mathbf{x}, r)$ is the number of unordered triplets (x_i, x_j, x_k) of distinct points in $\mathbf{x}$ in which each pair of points is closer than r units: $\|x_i - x_j\| \le r$, $\|x_i - x_k\| \le r$ and $\|x_j - x_k\| \le r$.

The following results describe the general form of Gibbs models and give a classification of them.

Theorem 1 (Hammersley-Clifford-Ripley-Kelly) *(Ripley and Kelly 1977) If the unnormalized density h is hereditary, then it has a* Gibbs representation

$$h(\mathbf{x}) = \exp\{V_0 + \sum_{x \in \mathbf{x}} V_1(x) + \sum_{\{x,y\} \subset \mathbf{x}} V_2(x, y) + \ldots\}, \tag{13}$$

where V_0 is a constant, and $V_k : \mathcal{X}^k \to \mathbb{R} \cup \{-\infty\}$ is called the potential of order k. *The representation is unique.*

Definition 3 *An unnormalized density h has* interaction order m *if, in the Gibbs representation (13), the potentials V_k for $k > m$ are identically zero. It has* interaction range R *if all the potentials satisfy $V_k(\mathbf{y}) = 0$ whenever $\mathbf{y}$ contains two points $y_1, y_2 \in \mathbf{y}$ such that $\|y_1 - y_2\| > R$.*

For example, the Poisson processes have interaction order 1 and interaction range 0. The pairwise interaction processes (5) have interaction order 2, ignoring trivial cases which reduce to a Poisson process. The Strauss process (7) with parameters β, γ, r has interaction order 2 and interaction range $R = r$. The triplet interaction process (12) has interaction order 3 and interaction range $R = r$. The Widom-Rowlinson model (9) and the Geyer saturation process (11) have infinite interaction order, and interaction range $R = 2r$.

3. Hybrids and their interpretation

Definition 4 *Let $h_1, h_2, \ldots, h_m$ be unnormalized densities. Their* simple hybrid *is the unnormalized density*

$$h(\mathbf{x}) = h_1(\mathbf{x}) \cdots h_m(\mathbf{x}), \quad \mathbf{x} \in \mathcal{X}. \tag{14}$$

The weighted hybrid *with exponents $\kappa_1, \ldots, \kappa_m$ (where $0 < \kappa_j < \infty$) is*

$$h(\mathbf{x}) = h_1(\mathbf{x})^{\kappa_1} \cdots h_m(\mathbf{x})^{\kappa_m}, \quad \mathbf{x} \in \mathcal{X}. \tag{15}$$

Note that the hybrid of several point process densities does not necessarily define a point process density. Additional assumptions are required to ensure that the hybrid is integrable. This is discussed further in Section 4.

Hybrids have a straightforward interpretation from the viewpoint of statistical physics. In that context, $\log h(\mathbf{x})$ is interpreted as the potential energy of a state $\mathbf{x}$, so a hybrid is created simply by adding the potential energy functionals of several models.

The statistical interpretation of a hybrid is somewhat more complicated. The likelihood of a hybrid is a product of component likelihoods, and the Papangelou conditional intensity of a hybrid is the product of the conditional intensities of the components. However, these product forms do not imply any kind of stochastic independence, since the usual factorization lemma does not apply to $h_1(\mathbf{x})h_2(\mathbf{x})$.

It is more appropriate to draw a connection with *relative distributions* (Handcock and Morris 1999). If X and Y are random variables with probability densities f_X and f_Y, then the *relative density* of Y with respect to X is $\rho(x) = f_Y(x)/f_X(x)$. The denominator f_X is sometimes called the "reference density". In constructing probability distributions it is sometimes convenient to specify f_Y by giving the reference density f_X and the relative density ρ, since $f_Y(x) = f_X(x)\rho(x)$. Similarly for finite point processes, it may be convenient to specify an unnormalized point process density h by specifying its relative density h_1 with respect to another point process density h_2, so that $h(\mathbf{x}) = h_1(\mathbf{x})h_2(\mathbf{x})$ is a hybrid.

A hybrid of several point process densities, each having a different characteristic 'scale' of interaction, will typically exhibit 'multi-scale' interaction.

Hybridization includes *conditioning* as a special case. Let $h_1(\mathbf{x}) = \mathbf{1}\{\mathbf{x} \in \mathcal{A}\}$ be the indicator of an event $\mathcal{A} \subset \mathcal{X}$. If h_2 is the unnormalized density of a point process $\mathbf{X}$, then the hybrid $h(\mathbf{x}) = h_1(\mathbf{x})h_2(\mathbf{x})$ is the unnormalized density of the point process $\mathbf{X}$ *conditioned on the event* $\mathcal{A}$. Examples of useful events $\mathcal{A}$ include $\mathcal{A} = \{n(\mathbf{X}) = m\}$, where m is a fixed integer, and $\mathcal{A} = \{\mathbf{X} \subset D\}$, where $D \subset W$ is a spatial domain. The corresponding hybrid models are equivalent to conditioning on the event that the process contains exactly m points, and that the process lies inside the domain D, respectively.

Hybridization also includes the concept of an *offset* in a model. If $g(\mathbf{x})$ is an unnormalized density and $\{v_\theta(\mathbf{x}) : \theta \in \Theta\}$ is a parametric family of unnormalized densities, then the family of hybrids

$$h_\theta(\mathbf{x}) = g(\mathbf{x})v_\theta(\mathbf{x}) \tag{16}$$

can be regarded as a modification of the original model $\{v_\theta\}$ by including an *'offset'* term $g(\mathbf{x})$. The stochastic interpretation of the parameter θ in the model h_θ is of course slightly different from its interpretation in the model v_θ.

For example, suppose it is known in advance that the data points must satisfy a hard core condition with distance c. Then we might fit a model of the form (16), where $g(\mathbf{x})$ is the density of the classical hard core process with hard core diameter c. In fitting the model h_θ to the data, we effectively fit the model v_θ while allowing for the presence of a hard core.

A special case of a model offset is *spatial inhomogeneity*. Suppose $g(\mathbf{x})$ is the density of an inhomogeneous Poisson process corresponding to a known source of spatial inhomogeneity. Then the model h_θ in (16) effectively 'adjusts' for this inhomogeneity. This is the standard approach to modeling spatially-inhomogeneous point processes. A Poisson point process with intensity function $\lambda(u)$, $u \in W$, has unnormalized density

$$g(\mathbf{x}) = \prod_{i=1}^{n(\mathbf{x})} \lambda(x_i). \tag{17}$$

Consequently we can recognize (5) as the hybrid of the unnormalized densities of an inhomogeneous Poisson process with intensity function $b(u)$, and a stationary pairwise interaction process with interaction function $c(u, v)$. In a similar way we can introduce a factor similar to (17) in combination with any point process density, to introduce spatial inhomogeneity.

Alternatively a model of the form (16) can be regarded as a *perturbation* of $g(\mathbf{x})$ which may be used to assess goodness-of-fit of the model $g(\mathbf{x})$. For this purpose, the family of models $\{h_\theta\}$ should include g itself, so we require that for $\theta = \mathbf{0}$ (say) we have $v_{\mathbf{0}} \equiv 1$. A formal goodness-of-fit test could be conducted by testing $H_0 : \theta = \mathbf{0}$. Suitable test statistics may include the maximum likelihood or maximum pseudolikelihood estimate of θ, and the score or pseudoscore (derivative of the log-likelihood or log pseudolikelihood) evaluated at $\theta = \mathbf{0}$. This approach is explored in Baddeley, Rubak, and Møller (2011).

4. Properties of hybrids

4.1. General results

The hybrid of several point process densities does not necessarily define a point process density. That is, if all factors h_j are integrable, it does not necessarily follow that the hybrid $h_1 \dots h_m$ is integrable. For a counterexample, take any unnormalized density h_1 which is integrable but not square-integrable (i.e., $h_1(\mathbf{x})^2$ is not integrable) and form the hybrid $h = h_1 h_1 = h_1^2$.

Additional assumptions on the component densities h_i are needed to ensure that the hybrid $h_1 \dots h_m$ is integrable and therefore defines a point process.

Lemma 1 *Hybridization preserves the hereditary property. That is, if $h_1, \dots, h_m$ are hereditary, then the hybrid $h = h_1 \dots h_m$ is hereditary. Furthermore, the interaction order of h is*

the maximum of the interaction orders of the components h_i. The interaction range of h is the maximum of the interaction ranges of h_i.

The proof is straightforward.

Thus a hybrid of Gibbs point processes is a Gibbs process, provided it is integrable. A hybrid of pairwise interaction Gibbs models is again a pairwise interaction Gibbs model, provided it is integrable.

Lemma 2 *Hybridization preserves Ruelle stability. That is, if all factors h_j are Ruelle stable, then the hybrid $h_1 \dots h_m$ is Ruelle stable (hence integrable).*

Proof: Suppose $h_j(\mathbf{x}) \le M_j^{n(\mathbf{x})}$. Then the weighted hybrid (15) satisfies

$$h(\mathbf{x}) = \prod_{j=1}^{m} h_j(\mathbf{x})^{\kappa_j} \le \prod_{j=1}^{m} \left(M_j^{n(\mathbf{x})}\right)^{\kappa_j} = \left(\prod_{j=1}^{m} M_j^{\kappa_j}\right)^{n(\mathbf{x})} = M^{n(\mathbf{x})},$$

where $M = \prod_{j=1}^{m} M_j^{\kappa_j} < \infty$. □

Lemma 3 *Hybridization preserves local stability. That is, if every factor h_j is locally stable, then the hybrid $h_1 \dots h_m$ is locally stable.*

Lemma 4 *The conditional intensity of a hybrid is the product of the conditional intensities of the components. That is, suppose $\mathbf{X}_1, \dots, \mathbf{X}_m$ are finite point processes with unnormalized densities $h_1, \dots, h_m$ and Papangelou conditional intensities $\lambda_i(u, \mathbf{x}) = h_i(\mathbf{x} \cup \{u\})/h_i(\mathbf{x} \setminus \{u\})$. If the hybrid $h(\mathbf{x}) = h_1(\mathbf{x}) \dots h_m(\mathbf{x})$ is integrable, then the corresponding point process has conditional intensity*

$$\lambda(u, \mathbf{x}) = \prod_{i=1}^{m} \lambda_i(u, \mathbf{x}). \tag{18}$$

The proof is trivial.

4.2. Hybridizing with a non-integrable density

It is not necessary that both component factors h_1 and h_2 be integrable in order to obtain an integrable hybrid $h = h_1 h_2$. This is very useful in practice, because it allows us to construct models involving clustering at some scales.

An instructive example occurs when h_1 is the Strauss density (7) with $\gamma > 1$, which is not integrable, and h_2 is the unnormalized density of a classical hard core process:

$$h_2(\mathbf{x}) = \mathbf{1}\{\|x_i - x_j\| > c \text{ for all } i \ne j\}, \tag{19}$$

where $c > 0$ is fixed. Since the window W has finite volume $|W|$, there is a finite upper bound $M \le |W|/(\pi c^2/4)$ on the number of discs of diameter c that can be placed in W without overlapping. Hence, realizations of the hard core process have at most M points

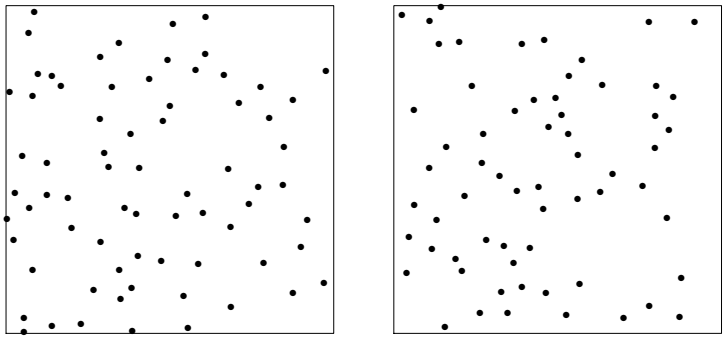

Figure 3: Simulated realizations of the hybrid Strauss-Hard core process with hard core distance $c = 0.03$, Strauss interaction distance $r = 0.07$, and Strauss interaction parameter $\gamma = 0.5$ (*Left*) and $\gamma = 1.5$ (*Right*) in the unit square, with intensity approximately equal to 70.

almost surely, i.e., $h_2(\mathbf{x}) > 0$ implies $n(\mathbf{x}) \leq M$. Now the term $s(\mathbf{x})$ in the Strauss density (7) satisfies $0 \leq s(\mathbf{x}) \leq n(\mathbf{x})(n(\mathbf{x}) - 1)$, so the Strauss density satisfies $h_1(\mathbf{x}) \leq K(n(\mathbf{x}))$, where $K(n) = \max\{1, \beta\}^n \max\{1, \gamma\}^{n(n-1)}$. The sequence $K(n)$ is non-decreasing. The hybrid $h(\mathbf{x}) = h_1(\mathbf{x})h_2(\mathbf{x})$ is zero if $n(\mathbf{x}) > M$, so we find that $h(\mathbf{x}) \leq K(M) < \infty$, that is, the hybrid Strauss-Hard core density is uniformly bounded, hence Ruelle stable and integrable, for all values of γ. By a similar argument, it is also locally stable for all γ, justifying the use of standard simulation algorithms.

Figure 3 shows simulated realizations of the hybrid Strauss-Hard core process with $\gamma = 0.5$ and $\gamma = 1.5$, generated by a Metropolis-Hastings algorithm. At short range, both patterns have a hard core, while at slightly longer range, the left panel (with $\gamma = 0.5$) exhibits inhibition while the right panel (with $\gamma = 1.5$) exhibits attraction. The discrepancy is already quite subtle in this simple example of a hybrid.

Experts in analysis will recognize that this phenomenon is connected with compact support of the integrand. We say that an unnormalized density h has *compact support* if there is an upper bound N such that $h(\mathbf{x}) > 0$ implies $n(\mathbf{x}) \leq N$. We say that h is *conditionally bounded* if

$$h(\mathbf{x}) \leq K(n(\mathbf{x})) \tag{20}$$

and *conditionally locally bounded* if

$$h(\mathbf{x} \cup \{u\}) \leq K(n(\mathbf{x}))h(\mathbf{x}) \tag{21}$$

for some constants $K(n) < \infty$ for each $n = 0, 1, 2, \ldots$. Then we have the following result.

Lemma 5 *Consider a hybrid $h(\mathbf{x}) = h_1(\mathbf{x}) \ldots h_m(\mathbf{x})$ where $h_1(\mathbf{x})$ has compact support.*

(a) If $h_1(\mathbf{x}), \ldots, h_m(\mathbf{x})$ are conditionally bounded, then $h(\mathbf{x})$ is uniformly bounded, has compact support, is Ruelle stable and is integrable.

(b) If $h_1(\mathbf{x}), \ldots, h_m(\mathbf{x})$ are conditionally locally bounded, then $h(\mathbf{x})$ is uniformly bounded, has compact support, is locally stable, Ruelle stable and integrable.

This lemma gives a sufficient condition for the validity of popular simulation algorithms, such as the birth-death Metropolis-Hastings algorithm, which require either local stability or Ruelle stability.

Examples of functions h with compact support include (19) and $h(\mathbf{x}) = \mathbf{1}\{n(\mathbf{x}) \le m\}$ where m is fixed. These are both hereditary. A non-hereditary example is $h_2(x) = \mathbf{1}\{n(\mathbf{x}) = m\}$. If $h_1(\mathbf{x})$ is the unnormalized Strauss density with $\gamma > 1$, then $h_1 h_2$ is the unnormalized density of the Strauss process with fixed number of points m and interaction parameter $\gamma > 1$ (which *is* integrable).

5. Multi-scale processes

Multi-scale processes can be constructed easily using hybridization. Indeed this is effectively how multi-scale models have been constructed in the literature. Here we survey the existing multi-scale models, representing them as hybrids, and also construct new multi-scale models. We also note the relation between these multi-scale models and the popular summary statistics F, G and K (Illian *et al.* 2008; Møller and Waagepetersen 2004).

5.1. Multi-scale pairwise interaction processes

The natural multi-scale generalization of the Strauss process (7) is a pairwise interaction process (5) in which the pair interaction term is a step function of the interpoint distance.

Consider a pairwise interaction process (5) with $b(u) \equiv \beta > 0$ and $c(u, v) = C(\|u - v\|)$, where C is a step function, with $C(t) = \epsilon_\ell$ when $r_{\ell-1} < t \le r_\ell$, where $0 = r_0 < r_1 < \ldots < r_L$, and $C(t) = 1$ for $t > r_L$. Then (5) reduces to

$$h(\mathbf{x}) = \beta^{n(\mathbf{x})} \prod_{\ell=1}^{L} \epsilon_\ell^{A_\ell(\mathbf{x})}, \tag{22}$$

where

$$A_\ell(\mathbf{x}) = \sum_{i<j} \mathbf{1}\{r_{\ell-1} < \|x_i - x_j\| \le r_\ell\}$$

is the number of unordered pairs of points in $\mathbf{x}$ whose interpoint distance lies in the interval $I_\ell = (r_{\ell-1}, r_\ell]$. The corresponding conditional intensity is

$$\lambda(u, \mathbf{x}) = \beta \prod_{\ell} \epsilon_\ell^{B_\ell(u,\mathbf{x})}$$

for $u \notin \mathbf{x}$, where $B_\ell(u, \mathbf{x}) = A_\ell(\mathbf{x} \cup \{u\}) - A_\ell(\mathbf{x} \setminus \{u\})$ is the number of points of $\mathbf{x}$ which are separated from the point u by distances in the interval I_ℓ.

Lemma 6 *The piecewise constant pairwise interaction density (22) is equivalent to a hybrid of Strauss densities (7) with the same interaction ranges $r_1, \ldots, r_L$. The interaction parameters ϵ_ℓ of (22) are related to the interaction parameters γ_ℓ of the Strauss densities by $\gamma_L = \epsilon_L$ and*

$$\begin{aligned} \gamma_\ell &= \epsilon_\ell / \epsilon_{\ell+1} \\ \epsilon_\ell &= \gamma_\ell \gamma_{\ell+1} \ldots \gamma_L \end{aligned}$$

for $\ell = 1, \ldots, L-1$.

Proof: Write

$$S_\ell(\mathbf{x}) = \sum_{i<j} \mathbf{1}\{\|x_i - x_j\| \le r_\ell\}, \quad \ell = 1, \ldots, L$$

for the number of unordered pairs of points of $\mathbf{x}$ that lie within a distance r_ℓ of one another. Then $A_1(\mathbf{x}) = S_1(\mathbf{x})$ and $A_\ell(\mathbf{x}) = S_\ell(\mathbf{x}) - S_{\ell-1}(\mathbf{x})$ for $\ell = 2, \ldots, L$. Substituting in (22) yields

$$h(\mathbf{x}) = \beta^{n(\mathbf{x})} \gamma_L^{S_L(\mathbf{x})} \prod_{\ell=1}^{L-1} \left(\frac{\epsilon_\ell}{\epsilon_{\ell+1}} \right)^{S_\ell(\mathbf{x})} \tag{23}$$

which is the hybrid of L unnormalized Strauss densities (7) with interaction radii $r_1, \ldots, r_L$ and interaction parameters ϵ_1/ϵ_2, ϵ_2/ϵ_3, $\ldots$, $\epsilon_{L-1}/\epsilon_L$ and ϵ_L respectively. □

Lemma 6 is instructive for developing skill in interpretation of hybrid models. It says that the combined interaction between two points separated by a distance s is equal to the product of all the component interactions that apply to distance s. In the case of a hybrid of Strauss processes, this means we must consider all the components with interaction radii r greater than s. An illustrative example is given in the online supplements.

Notice that the model (22) is integrable at least when all $\epsilon_\ell \le 1$, whereas the individual Strauss models in the hybrid representation are integrable iff $\gamma_L = \epsilon_L \le 1$ and $\gamma_\ell = \epsilon_\ell/\epsilon_{\ell+1} \le 1$ for $\ell < L$, which is a more restrictive condition. This is an example of the principle enunciated in Section 4.2 that the hybrid may be well-defined for a larger set of parameter values than are permissible for the components of the hybrid.

The canonical sufficient statistic for the piecewise constant pairwise interaction density (22) is the vector of values $(S_1(\mathbf{x}), \ldots, S_L(\mathbf{x}))$. Extending an observation in Assunção (2003) and Baddeley *et al.* (2011), this statistic is essentially equivalent (ignoring details about edge correction) to the empirical K-function (Ripley 1976) evaluated at the distances $r_1, \ldots, r_L$.

Inhomogeneous point processes with multi-scale pairwise interaction can be constructed similarly, by taking $b(u)$ in (5) to be a non-constant function.

5.2. Multi-scale area-interaction processes

The Widom-Rowlinson 'area-interaction' process (Widom and Rowlinson 1970; Rowlinson 1980; Baddeley and van Lieshout 1995) was defined in (9) above. Ambler (2002); Ambler and Silverman (2004) and Picard *et al.* (2009) proposed a multi-scale generalization of this model, with unnormalized density

$$h(\mathbf{x}) = \beta^{n(\mathbf{x})} \exp\left(-\sum_{\ell=1}^{L} \kappa_\ell U(\mathbf{x}, r_\ell) \right), \tag{24}$$

where $U(\mathbf{x}, r)$ is the area of the dilation by radius r, defined in (10).

This is clearly the hybrid of the individual area-interaction models (9). Extending an observation made in Baddeley and van Lieshout (1995) and Baddeley *et al.* (2011), the canonical sufficient statistic for (24) is essentially the empirical empty space function F.

5.3. Multi-scale saturation processes

The Geyer saturation process (11) is one case where the apparently natural generalization to

a multi-scale process does not have the desired properties. We describe this first, and then study the hybrid of Geyer models.

Saturated step function interaction

Consider the piecewise constant pairwise interaction process (22). Recall that $A_\ell(\mathbf{x})$ counts all unordered pairs of distinct data points (x_i, x_j) whose pairwise distance lies in the interval I_ℓ, while $B_\ell(x_i, \mathbf{x})$ counts those pairs involving x_i. Summing $B_\ell(x_i, \mathbf{x})$ over i will count each pair twice:

$$\sum_{i=1}^{n(\mathbf{x})} B_\ell(x_i, \mathbf{x}) = 2A_\ell(\mathbf{x}).$$

Hence the density (22) can be rewritten as a product over points of $\mathbf{x}$,

$$h(\mathbf{x}) = \prod_{i=1}^{n(\mathbf{x})} H(x_i, \mathbf{x}),$$

where

$$H(x_i, \mathbf{x}) = \beta \prod_{\ell=1}^{L} \epsilon_\ell^{\frac{1}{2}B_\ell(x_i,\mathbf{x})}.$$

It may therefore seem reasonable to define a multi-scale generalization of the Geyer saturation process as follows. Introduce saturation parameters $s_1, \ldots, s_L$ satisfying $0 \le s_\ell \le \infty$. Define the 'saturated annular neighbor count'

$$B^*_\ell(u, \mathbf{x}) = \min\{s_\ell, B_\ell(u, \mathbf{x})\} \tag{25}$$

the 'saturated' version of H

$$H^*(x_i, \mathbf{x}) = \beta \prod_{\ell=1}^{L} \epsilon_\ell^{\frac{1}{2}B^*_\ell(x_i,\mathbf{x})}$$

and the saturation density

$$h^*(\mathbf{x}) = \prod_{i=1}^{n(\mathbf{x})} H^*(x_i, \mathbf{x}). \tag{26}$$

By construction, this unnormalized density is Ruelle stable, and therefore integrable. However, we give a counterexample below to show that h^* is not locally stable.

The conditional intensity associated with h^* is

$$\begin{aligned} \lambda(u, \mathbf{x}) &= \beta H^*(u, \mathbf{x} \cup \{u\}) \prod_{i=1}^{n(\mathbf{x})} \frac{H^*(x_i, \mathbf{x} \cup \{u\})}{H^*(x_i, \mathbf{x})} \\ &= \beta H^*(u, \mathbf{x}) \prod_{i=1}^{n(\mathbf{x})} \prod_{\ell=1}^{L} \epsilon_\ell^{\frac{1}{2}D_\ell(x_i,u,\mathbf{x})}, \end{aligned} \tag{27}$$

where

$$D_\ell(x_i, u, \mathbf{x}) = B^*_\ell(x_i, \mathbf{x} \cup \{u\}) - B^*_\ell(x_i, \mathbf{x}).$$

Note that $D_\ell(x_i, u, \mathbf{x})$ equals 1 if $r_\ell < \|x_i - u\| \le r_{\ell+1}$ *and* $A_\ell(x_i, \mathbf{x}) \le s_\ell - 1$, and equals 0 otherwise.

The following counterexample shows that the unnormalized density h^* defined in (26) is not locally stable. Take $L = 2$, $r_1 = 15$, $r_2 = 17$, $\beta = 1$, $\epsilon_1 = 1$, $\epsilon_2 = 2$ and $s_1 = s_2 = 3$. Take u to be the origin in $\mathbb{R}^2$, and let $\mathbf{x}$ be the point pattern in $\mathbb{R}^2$ consisting of a sequence of n equispaced points on the circular arc of radius 16, centered at the origin, between the angles $-\pi/10$ and $\pi/10$.

All the points of $\mathbf{x}$ are within the annulus with radii r_1 and r_2 centered at the origin. However, the annuli with radii r_1 and r_2 centered at any of the points $x_i \in \mathbf{x}$ do not intersect the pattern $\mathbf{x}$. Thus $A_2(x_i, \mathbf{x}) = 0$ for all i, so $D_2(x_i, u, \mathbf{x}) = 1$ for all i, and hence $\lambda(u, \mathbf{x}) = 2^{(n+1)/2}$. Thus h^* is not locally stable.

Hybrid Geyer process

Consider instead the hybrid of several Geyer densities (11) with interaction ranges $r_1, \ldots, r_m$, saturation parameters $s_1, \ldots, s_m$ and interaction parameters $\gamma_1, \ldots, \gamma_m$ respectively,

$$h(\mathbf{x}) = \beta^{n(\mathbf{x})} \prod_{i=1}^{n(\mathbf{x})} \prod_{j=1}^{m} \gamma_j^{\min(s_j, t(x_i, \mathbf{x}\setminus x_i; r_j))}, \tag{28}$$

where $t(u, \mathbf{x}; r) = \sum_i \mathbf{1}\{\|u - x_i\| \le r\}$. Since the Geyer density (11) is hereditary and locally stable (Geyer 1999), the same is true for the hybrid (28) by Lemmas 1 and 3. Hence it is also Ruelle stable. The interaction range is $R = 2 \max_j r_j$.

Extending an observation made in Baddeley *et al.* (2011), if all the saturation parameters s_j are equal to 1, then the canonical sufficient statistic for (28) is (ignoring edge corrections) the empirical nearest-neighbor distance distribution G, evaluated at the distances $r_1, \ldots, r_m$. This is a further indication that the hybrid (28) is the best choice for a multi-scale generalization of the saturation process.

6. Parametric inference

This section gives some statistical theory for hybrid point process models. We consider statistical inference for hybrids of parametric point process models of regular linear exponential family form (Wei 1998; Küchler and Sørensen 1997; Barndorff-Nielsen 1978).

6.1. Linear exponential family

First we recall some facts about exponential families of point process densities. Consider an unnormalized density of the form

$$h(\mathbf{x}; \theta) = \exp(\theta^\top S(\mathbf{x}) + B(\mathbf{x})) \tag{29}$$

with p-dimensional canonical parameter $\theta \in \Theta \subseteq \mathbb{R}^p$, and p-dimensional canonical sufficient statistic $S : \mathcal{X} \to \mathbb{R}^p$. The function $B : \mathcal{X} \to \mathbb{R} \cup \{-\infty\}$ is an optional offset term. The parameter space Θ is assumed to be a convex subset of $\mathbb{R}^p$.

If $h(\cdot; \theta)$ is integrable, then it defines a point process model, with likelihood

$$L(\theta) = f(\mathbf{x}; \theta) = M(\theta)^{-1} h(\mathbf{x}; \theta)$$

called a *linear exponential family* likelihood, where $M(\theta) = \mathsf{E}[h(\mathbf{Z};\theta)]$ is the normalizing constant and also the moment generating function for $S(\mathbf{Z})$. By standard results the log-likelihood is a concave function. Under regularity conditions, the likelihood is maximized at a root of the likelihood score

$$U(\mathbf{x};\theta) = \frac{\partial}{\partial\theta}\log L(\theta) = S(\mathbf{x}) - \mathsf{E}_\theta S(\mathbf{X}), \tag{30}$$

where E_θ denotes the expectation with respect to $f(\cdot;\theta)$. The score is an unbiased estimating function.

The conditional intensity is

$$\lambda(u,\mathbf{x};\theta) = \exp(\theta^\top T(u,\mathbf{x}) + C(u,\mathbf{x})) \tag{31}$$

if $h(\mathbf{x}) > 0$, and 0 otherwise, where $T(u,\mathbf{x}) = S(\mathbf{x}\cup\{u\}) - S(\mathbf{x}\setminus\{u\})$ and $C(u,\mathbf{x}) = B(\mathbf{x}\cup\{u\}) - B(\mathbf{x}\setminus\{u\})$. The *pseudolikelihood* (Besag 1977; Jensen and Møller 1991; Jensen and Künsch 1994) is

$$\mathsf{PL}(\theta;\mathbf{x}) = \left[\prod_{i=1}^{n(\mathbf{x})}\lambda(x_i,\mathbf{x};\theta)\right]\exp\left(-\int_W \lambda(u,\mathbf{x};\theta)\,\mathrm{d}u\right) \tag{32}$$

and the pseudoscore (derivative of the log pseudolikelihood) for θ is

$$\tilde{U}(\mathbf{x};\theta) = \sum_i T(x_i,\mathbf{x}) - \int_W T(u,\mathbf{x})\lambda(u,\mathbf{x};\theta)\,\mathrm{d}u. \tag{33}$$

The pseudoscore is also an unbiased estimating function.

6.2. Hybrid

Now suppose we have m parametric point process models whose unnormalized densities are each of regular linear exponential family form

$$h_j(\mathbf{x};\theta^{(j)}) = \exp({\theta^{(j)}}^\top S^{(j)}(\mathbf{x}) + B^{(j)}(\mathbf{x})), \quad j = 1,\ldots,m \tag{34}$$

with p_j-dimensional canonical parameters $\theta^{(j)} \in \Theta^{(j)} \subseteq \mathbb{R}^{p_j}$, p_j-dimensional canonical sufficient statistics $S^{(j)} : \mathcal{X} \to \mathbb{R}^{p_j}$, and optional offset terms $B^{(j)} : \mathcal{X} \to \mathbb{R}\cup\{-\infty\}$. Each parameter space $\Theta^{(j)}$ is a convex subset of $\mathbb{R}^{p_j}$. We do not necessarily assume that all the components h_j are integrable.

Consider the hybrid

$$h(\mathbf{x};\theta) = \prod_{j=1}^m h_j(\mathbf{x};\theta^{(j)}), \tag{35}$$

where $\theta = (\theta^{(1)},\ldots,\theta^{(m)})$ is the combined parameter vector, belonging to the convex set $\Theta = \Theta^{(1)}\times\ldots\times\Theta^{(m)}$ in $\mathbb{R}^p$ where $p = \sum_j p_j$. Then (35) is again of linear exponential family form, and is supported by the existing theory of estimation and inference for exponential families (Barndorff-Nielsen 1978; Jensen and Møller 1991; Geyer 2009).

Note that we only need to consider the case of simple hybrids, because in a weighted hybrid of the densities (34) the weights κ_j would be absorbed into the parameters $\theta^{(j)}$.

Assume the hybrid $h(\cdot;\theta)$ is integrable for all $\theta \in \Theta$, with normalized density $f(\cdot;\theta)$. Then its likelihood score $U(\mathbf{x},\theta)$ has jth component

$$U(\mathbf{x};\theta)^{(j)} = S^{(j)}(\mathbf{x}) - \mathsf{E}_\theta S^{(j)}(\mathbf{X}) \tag{36}$$

where E_θ denotes the expectation with respect to $f(\cdot;\theta)$.

Further assume that each component h_j is hereditary (which implies that the hybrid h is hereditary). The conditional intensity of the hybrid (35) is

$$\lambda(u,\mathbf{x};\theta) = \prod_{j=1}^{m} \lambda^{(j)}(u,\mathbf{x};\theta^{(j)}). \tag{37}$$

Substituting in (32) gives the pseudolikelihood of the hybrid. The pseudoscore $\tilde{U}(\mathbf{x};\theta)$ of the hybrid model has jth component

$$\tilde{U}(\mathbf{x};\theta)^{(j)} = \sum_i T^{(j)}(x_i,\mathbf{x}) - \int_W T^{(j)}(u,\mathbf{x})\lambda(u,\mathbf{x};\theta)\,\mathrm{d}u, \tag{38}$$

where again $T(u,\mathbf{x}) = S(\mathbf{x}\cup\{u\}) - S(\mathbf{x}\setminus\{u\})$. Existing algorithms for maximizing the likelihood or pseudolikelihood of a Gibbs model (Baddeley and Turner 2000; Mateu and Montes 2001a,b) apply equally to hybrids.

In a hybrid, some of the parameters of the component models may be *unidentifiable*. For example, the hybrid of two homogeneous Poisson processes with intensities β_1 and β_2, respectively, is another homogeneous Poisson process with intensity $\lambda = \beta_1\beta_2$. The parameters β_1, β_2 cannot be determined from λ, and therefore certainly cannot be estimated from a realization of the hybrid process. Similarly a hybrid of the Strauss processes with parameters (γ_1, r_1) and (γ_2, r_2) with $r_1 = r_2 = r$ is simply a Strauss process with parameters (γ, r) where $\gamma = \gamma_1\gamma_2$. The separate parameters γ_1, γ_2 are not identifiable. Many such examples arise.

6.3. Irregular parameters

Many popular Gibbs models are expressible in the linear exponential form (29) only if the values of certain "irregular" or "nuisance" parameters η are held fixed. The unnormalized density can be written in the form

$$h(\mathbf{x};\theta,\eta) = \exp(\theta^\top S(\mathbf{x},\eta) + B(\mathbf{x},\eta)) \tag{39}$$

instead of (29), where θ are the "regular" parameters and η the "irregular" parameters. For example the interaction range r of the Strauss density (7) is an irregular parameter.

Typically the likelihood $L(\theta,\eta)$ and pseudolikelihood $\mathsf{PL}(\theta,\eta)$ are not well-behaved as functions of the irregular parameters η. For example the Strauss likelihood (7) is not differentiable as a function of the interaction range r. Thus, analytic methods for maximizing the likelihood or pseudolikelihood are typically not applicable to the irregular parameters, and very little statistical theory is available for estimates of the irregular parameters.

A general statistical technique for estimating nuisance parameters is through the *profile loglikelihood*

$$\ell(\eta) = \max_\theta \log L(\theta,\eta) \tag{40}$$

because the value of η which maximizes $\ell(\eta)$ is also the maximum likelihood estimate of η (assuming these maxima are attained). Values of the function $\ell(\eta)$ can be computed efficiently for any given η using computational methods for linear exponential families. Maximization over η can then be performed by brute-force algorithms which require evaluating $\ell(\eta)$ at many values of η.

In the context of point process models, the analogous *profile pseudolikelihood* (Baddeley and Turner 2000, 2006)

$$p(\eta) = \max_{\theta} \log \mathsf{PL}(\theta, \eta) \tag{41}$$

can be used to estimate irregular parameters. The value of $p(\eta)$ for any given η can be computed rapidly using the Berman-Turner device (Berman and Turner 1992; Baddeley and Turner 2000). Maximization over η must typically be performed by brute force.

One tractable case is where the irregular parameter is a *hard core distance*.

Lemma 7 *Consider a point process* $\mathbf{X}$ *with unnormalized density* $h(\mathbf{x}; \theta, c) = h_1(\mathbf{x}; \theta) h_2(\mathbf{x}; c)$ *where* h_1 *is always positive* ($h_1(\mathbf{x}; \theta) > 0$ *for all* $\mathbf{x}, \theta$) *and* h_2 *is the unnormalized hard core density (19) with hard core distance* $c > 0$. *Then the maximum likelihood estimate and maximum pseudolikelihood estimate of* c *are both equal to*

$$\hat{c} = m(\mathbf{x}) = \min_{i \neq j} \|x_i - x_j\|$$

the minimum interpoint distance (with $m(\mathbf{x}) = \infty$ *if* $n(\mathbf{x}) < 2$*).*

Proof: The hard core density (19) can equivalently be written

$$h_2(\mathbf{x}; c) = \mathbf{1}\{m(\mathbf{x}) \geq c\}.$$

The likelihood is

$$L(\theta, c) = A(\theta, c)^{-1} h_1(\mathbf{x}; \theta) \mathbf{1}\{m(\mathbf{x}) \geq c\}$$

with normalizing constant

$$A(\theta, c) = \mathsf{E}[h_1(\mathbf{Z}; \theta) \mathbf{1}\{m(\mathbf{Z}) \geq c\}],$$

where again $\mathbf{Z}$ denotes the Poisson process with unit intensity. Note that h_2 is decreasing as a function of c. Hence $A(\theta, c)$ is decreasing as a function of c, for fixed θ. Consequently, for fixed θ, the likelihood $L(\theta, c)$ is increasing as a function of c for $c \leq m(\mathbf{x})$, and $L(\theta, c) = 0$ for $c > m(\mathbf{x})$. Hence $L(\theta, c)$ is maximized at $\hat{c} = m(\mathbf{x})$.

The conditional intensity is

$$\lambda_{\theta,c}(u, \mathbf{x}) = I_c(u, \mathbf{x}) \kappa_\theta(u, \mathbf{x})$$

where $\kappa_\theta(u, \mathbf{x}) = h_1(\mathbf{x} \cup \{u\}, \theta) / h_1(\mathbf{x}, \theta)$ and $I_c(u, \mathbf{x}) = \mathbf{1}\{\min_i \|u - x_i\| \geq c\}$. The pseudolikelihood is

$$\mathsf{PL}(\theta, c) = \prod_i \kappa_\theta(x_i, \mathbf{x}) \exp\left(- \int I_c(u, \mathbf{x}) \kappa_\theta(u, \mathbf{x}) \, \mathrm{d}u \right)$$

provided $c \leq m(\mathbf{x})$, and $\mathsf{PL}(\theta, c) = 0$ otherwise. The pseudolikelihood is an increasing function of c for $c \leq m(\mathbf{x})$. Hence the maximum pseudolikelihood estimate of c is $\hat{c} = m(\mathbf{x})$. □

Lemma 7 implies that the maximum likelihood and maximum pseudolikelihood estimate of a hard core distance c is always positively biased, since $m(\mathbf{x}) \geq c$ with probability 1.

6.4. Stepwise model selection

Hybrid models can be built up incrementally, by adding one component at a time. To decide whether a particular component should be added, one can use the closely-related residual summary statistics of Baddeley *et al.* (2011).

Suppose that a hybrid with m components has been fitted, with fitted parameter vector $\hat{\theta}_0$, and we now consider the addition of an additional $(m+1)$th hybrid component of the form (34). Treating the fitted m-component hybrid as a submodel of the $(m+1)$-component model, we can calculate the $(m+1)$th component of the score residual,

$$U(\mathbf{x};\theta)^{(m+1)} = S^{(m+1)}(\mathbf{x}) - \mathsf{E}_{\hat{\theta}} S^{(m+1)}(\mathbf{X}) \tag{42}$$

or the $(m+1)$th component of the pseudoscore residual

$$\tilde{U}(\mathbf{x};\theta)^{(m+1)} = \sum_i T^{(m+1)}(x_i, \mathbf{x}) - \int_W T^{(m+1)}(u, \mathbf{x}) \lambda(u, \mathbf{x}; \hat{\theta}) \, \mathrm{d}u. \tag{43}$$

In these equations $\mathsf{E}_{\hat{\theta}}$ is the expectation with respect to the fitted m-component hybrid, and $\lambda(u, \mathbf{x}; \hat{\theta})$ is its conditional intensity. The score residual (42) and pseudoscore residual (43) should have mean zero if the fitted m-component hybrid is the correct model.

This approach is used to analyze the Gordon Square data in Section 11. The performance of maximum profile pseudolikelihood estimation is studied in a simulation experiment in Section 10.

7. Software implementation of parametric inference

Maximum pseudolikelihood estimation is implemented in the R package **spatstat** (Baddeley and Turner 2005, 2006). The implementation has recently been extended to hybrid models (in **spatstat** version 1.29-0 and later).

7.1. User interface

A typical command line for fitting a point process model in **spatstat** is

```
R> fit <- ppm(X, ~ 1, Strauss(0.1))
```

indicating that the stationary Strauss process with interaction range $r = 0.1$ units is to be fitted to the point pattern data set `X` and the resulting fitted model stored in the object `fit`. Note that we have to specify the Strauss interaction range r, but not the interaction strength γ, which is estimated in the fitting process. Similarly

```
R> fit <- ppm(X, ~ polynom(x, y, 3), Geyer(0.2, 1))
```

indicates an inhomogeneous Geyer saturation process with interaction distance $r = 0.2$ and saturation parameter $s = 1$, together with a first-order potential V_1 that is a cubic polynomial

in the Cartesian coordinates. Models can be fitted by maximum pseudolikelihood (Baddeley and Turner 2000) (the default) or by the Huang and Ogata (1999) method.

The fitted object `fit` belongs to the class 'ppm' of fitted point process models. This is an S3 class for which we have implemented methods for the generic R commands `print`, `plot`, `summary`, `anova`, `coef`, `dfbetas`, `extractAIC`, `fitted`, `formula`, `influence`, `labels`, `leverage`, `logLik`, `model.frame`, `nobs`, `residuals`, `simulate`, `terms`, `update` and `vcov`, as well as additional generics introduced in **spatstat**. This provides extensive capabilities for parametric inference. Original research was needed for the development of the `residuals` (Baddeley, Turner, Møller, and Hazelton 2005), `vcov` (Baddeley, Møller, and Pakes 2008; Coeurjolly and Rubak 2012) and `influence` and `dfbetas` (Baddeley, Chang, and Song 2012) methods for 'ppm' objects.

We have extended this interface to allow hybrid models. A hybrid of the two models mentioned above would be

```
R> fit <- ppm(X, ~ polynom(x, y, 3), Hybrid(Strauss(0.1), Geyer(0.2, 1))
```

Implementation details are explained below.

For the estimation of irregular parameters (such as the Strauss interaction range r) by maximum profile pseudolikelihood, there is a separate function `profilepl` which performs brute force maximization. In a future version of **spatstat**, this capability will be included in `ppm` itself.

7.2. Software design

The model-fitting function `ppm` fits a point process model using maximum pseudolikelihood (Baddeley and Turner 2000) and optionally computes an improved fit using the Huang and Ogata (1999) updating method. To maximize the pseudolikelihood of linear exponential family models we use the Berman and Turner (1992) device. This approximates the log of the pseudolikelihood (32) by a finite sum over sample points $u_j \in W$, in such a way that the approximate log pseudolikelihood is equivalent to the weighted log-likelihood of a Poisson loglinear regression model, which can then be fitted reliably using existing software for generalized linear models. Our code only needs to compute the values $t_j = T(u_j, \mathbf{x})$ of the statistic T appearing in the conditional intensity (31), evaluated at the sample points u_j.

Separating trend and interaction

For implementation purposes, we rearrange the Gibbs representation of the conditional intensity into the form

$$\lambda(u, \mathbf{x}) = \exp\{B(u) + \theta^\top V_1(u) + \theta^\top G(u, \mathbf{x})\}, \tag{44}$$

where $B(u)$ is an offset, $\theta^\top V_1(u)$ is the first order potential, and $\theta^\top G(u, \mathbf{x})$ is the interaction potential, the sum of all terms of order 2 and higher in the Gibbs representation.

The "trend" terms $B(u)$ and $V_1(u)$ are determined by the second argument to `ppm`. This is a formula in the R language, with no left hand side, which specifies the functions $B(u)$ and $V_1(u)$. Variable names appearing in the formula may be the reserved symbols `x` and `y` representing the Cartesian coordinates, or the names of spatial covariate functions or pixel

images. If a term in the formula is enclosed by `offset()` then this determines the offset function $B(u)$.

The "interaction" term $G(u, \mathbf{x})$ is controlled by the third argument to `ppm`, discussed below. Note that the mathematical form of $G(u, \mathbf{x})$ is the same for all models of the same interaction order k. This fact is used in the software design. For example, for all pairwise interaction models (5) the interaction term is of the form

$$G(u, \mathbf{x}) = \sum_i q(u, x_i), \tag{45}$$

where q is the *canonical pair potential*. For example in the Strauss model (7),

$$q(u, v) = \begin{cases} 1 & \text{if } \|u - v\| < r \\ 0 & \text{otherwise} \end{cases} \tag{46}$$

so that the pair interaction function c in (6) is recovered by $c(u, v) = \exp(\theta q(u, v))$ with $\theta = \log \gamma$.

Interaction objects

In a command such as

```
R> fit <- ppm(X, ~ 1, Strauss(0.1))
```

evaluation of the expression `Strauss(0.1)` yields an object, of the special class '`interact`', which describes the interaction structure of the Strauss process with interaction range $r = 0.1$. Users do not need to understand or manipulate these objects; users need only create them by calling an appropriate function (such as `Strauss` for Strauss models) and simply pass the result to the model-fitting function `ppm`.

An object of class '`interact`' includes the following components:

`family:` a pointer to the family of point process interactions to which this interaction belongs.

`pot:` an R function to calculate the canonical interaction potential of the process

`par:` a list of the irregular or nuisance parameters of the process.

For example, for the object `Strauss(0.1)`, the `family` is the family of *pairwise interaction processes*; the `pot` is the function $q(u, v)$ defined in Equation (46); and `par` specifies the interaction range $r = 0.1$.

Interaction family

The parent `family` of an interaction is another object (of class '`isf`' for "interaction structure family") that defines the operations that are common to all models of the same type. This includes the following components:

`eval:` an R function to compute the interaction term $G(u, \mathbf{x})$ in the conditional intensity (44).

`suffstat:` an R function to compute the canonical sufficient statistic $S(\mathbf{x})$ in the likelihood (29).

There are separate families for pairwise interactions, saturated pairwise models (such as Geyer's saturation model) and general infinite-order interactions (such as the area-interaction model).

For example, the `eval` function for the family of pairwise interaction processes computes (45) given the point pattern data set `x`, the sample location u, and the canonical potential q for the particular model.

Hybrid interactions

To extend this implementation to hybrid models, we need only implement the function `Hybrid` and the corresponding interaction structure family `hybrid.family`.

The `Hybrid` function accepts arguments which are objects of class '`interact`', for example,

```
R> Hybrid(Strauss(0.1), Geyer(0.2, 1))
```

The result of this call is a new object of class '`interact`' containing as its `pars` component the list of interactions `list(Strauss(0.1), Geyer(0.2, 1))`.

The main work is in coding the interaction structure family for hybrid models, `hybrid.family`. The `eval` function must, for each component interaction, extract the potential `pot` and the interaction structure `family`, then invoke the `eval` function for the relevant family to compute the values $T^{(j)}(x_i, \mathbf{x})$ and $T^{(j)}(u, \mathbf{x})$ in (38) for each data point x_i and each dummy point u. The results are collected (in separate columns of pseudo-data for each component) and returned. The existing code for maximum pseudolikelihood then continues.

Existing code for interactions includes various mechanisms for checking the validity of the fitted model. A fitted model is valid if its fitted parameters are all finite numeric values (rather than `NA`, `NaN` or `Inf`) and if the model is integrable. This code must be modified slightly to allow the conditions for validity of a hybrid model to be different from the conditions for validity of its components. In the current version of **spatstat** a hybrid model is declared to be valid if *either* all the components are valid *or* if Lemma 5 applies.

First-order trend in hybrids

Note that, in this software design, the `Hybrid` operator is applied to *interactions*, not to point process densities. Our approach to modeling is to separate the first-order ('intensity' or 'trend') potential from the higher-order ('interaction') potential as in (44).

Conveniently, this helps to avoid problems with the unidentifiability of parameters. For example, the hybrid of two Strauss processes with parameters β_1, γ_1, r_1 and β_2, γ_2, r_2 respectively, has first order term $\mu = \beta_1\beta_2$, so the parameters β_1 and β_2 cannot be identified separately. In our software design the hybrid of `ppm(X, ~ 1, Strauss(0.1))` and `ppm(X, ~ 1, Geyer(0.2, 1))` is `ppm(X, ~ 1, Hybrid(Strauss(0.1), Geyer(0.2, 1)))` so that the code will not attempt to estimate β_1 and β_2 separately, but will estimate μ.

In general, to form a hybrid of point process models, we combine their interactions using `Hybrid`, and *combine their trend formulas* by adding all the terms on the right-hand sides. For example the hybrid of `ppm(X, ~ A, Strauss(0.1))` and `ppm(X, ~ B, Geyer(0.2, 1))` is `ppm(X, ~ A + B, Hybrid(Strauss(0.1), Geyer(0.2, 1)))` where `A` and `B` are any model terms.

8. Simulation

Markov chain methods for simulating spatial point process models are surveyed in Møller and Waagepetersen (2004). For Gibbs models, simulation algorithms typically require only computation of the Papangelou conditional intensity. Hence they are easily extensible to simulation of hybrids.

8.1. Metropolis-Hastings algorithm

The standard birth-death Metropolis-Hastings algorithm for point processes (Geyer and Møller 1994) is a discrete-time Markov chain whose states are point patterns $\mathbf{x}$. Each proposal is either a "birth" or a "death" with probabilities p and $1-p$ respectively. In a "birth" proposal $\mathbf{x} \mapsto \mathbf{x} \cup \{u\}$ the existing configuration $\mathbf{x}$ is augmented by adding a random point u with proposal density $b(u, \mathbf{x})$ in the simulation window W. A "death" proposal $\mathbf{x} \mapsto \mathbf{x} \setminus \{x_i\}$ is the deletion of one of the existing points of the configuration $\mathbf{x}$, chosen with equal probability $1/n(\mathbf{x})$. For a target point process density $f(\mathbf{x})$ the Metropolis-Hastings acceptance probabilities are

$$\begin{aligned} A(\mathbf{x} \mapsto \mathbf{x} \cup \{u\}) &= \frac{f(\mathbf{x} \cup \{u\})(1-p)/n(\mathbf{x} \cup \{u\})}{f(\mathbf{x}) p b(u, \mathbf{x})} = \frac{\lambda(u, \mathbf{x})}{b(u, \mathbf{x})} \frac{(1-p)}{p(n(\mathbf{x})+1)}, \\ A(\mathbf{x} \mapsto \mathbf{x} \setminus \{x_i\}) &= \frac{f(\mathbf{x} \setminus \{x_i\}) p b(x_i, \mathbf{x} \setminus \{x_i\})}{f(\mathbf{x})(1-p)/n(\mathbf{x})} = \frac{b(x_i, \mathbf{x} \setminus \{x_i\})}{\lambda(x_i, \mathbf{x})} \frac{p n(\mathbf{x})}{(1-p)}. \end{aligned}$$

By Lemma 4 the conditional intensity of a hybrid is the product $\lambda(u, \mathbf{x}) = \prod_{j=1}^{m} \lambda_j(u, \mathbf{x})$ of the conditional intensities of the components. Thus, to implement a Metropolis-Hastings algorithm for a hybrid, we need only take existing algorithms which perform Metropolis-Hastings for the component processes, and modify the calculation of acceptance probabilities as above.

Computation of the acceptance probabilities requires only the ratio $\lambda(u, \mathbf{x})/b(u, \mathbf{x})$ of the conditional intensity to the birth proposal density. However, convergence rates depend on the choice of proposal density $b(u, \mathbf{x})$.

Suppose we have already implemented a birth-death Metropolis-Hastings algorithm for each of the components of the hybrid, with proposal densities $b_j(u, \mathbf{x})$ for $j = 1, \ldots, m$. This includes generation of proposals with density $b_j(u, \mathbf{x})$ and calculation of the ratio $\lambda_j(u, \mathbf{x})/b_j(u, \mathbf{x})$, and presumably implies that the choice of b_j ensures good convergence. To implement a birth-death Metropolis-Hastings algorithm for the hybrid model, the easiest route is to take the proposal density $b(u, \mathbf{x}) = C \prod_{j=1}^{m} b_j(u, \mathbf{x})$ where C is a normalizing constant, assuming it is feasible to generate proposals with this density. Then only minor modifications are required to the code for calculating acceptance probabilities.

The simulation algorithm implemented in **spatstat** has birth, death and *shift* proposals. In a shift proposal, one of the points in the current configuration is selected at random and moved to another location. Again, only minor modifications are required to accommodate hybrid models.

8.2. Exact simulation

Exact simulation techniques such as coupling-from-the-past (CFTP) have been developed (Berthelsen and Møller 2002, 2003; Kendall and Møller 2000) for point process densities f

which are purely inhibitory, i.e., for which $\mathbf{x} \subset \mathbf{y}$ implies $f(\mathbf{x}) \geq f(\mathbf{y})$. In such techniques the spatial birth-death process associated with a Poisson point process is coupled to a sampler for the target process. Simulation of hybrid models is also possible, provided the hybrid is purely inhibitory, and requires calculations very similar to those above.

9. Software implementation of simulation

9.1. User interface

The **spatstat** package includes a flexible implementation of the birth-death-shift Metropolis-Hastings algorithm in the form of the function `rmh`. This is a generic function with methods for '`ppm`' objects and a default method.

The default method accepts a simple specification of the model that can easily be put together by hand. For example

```
R> rmh(model = list(cif = 'strauss',
+    par = list(beta = 100, gamma = 0.5, r = 0.07), w = square(1)),
+    nrep = 1e6)
```

specifies that the Metropolis-Hastings algorithm for the Strauss process with parameters $\beta = 100$, $\gamma = 0.5$ and $r = 0.07$ in the unit square should be run for one million iterations. The argument `model` should be a list including the entries

`cif:` a character string identifying the interpoint interaction;

`par:` a list of model parameters;

`w:` the spatial domain ('window') containing the simulated pattern.

The default method of `rmh` first validates the model parameters by converting the loosely-structured specification of the model above into a structured object of class '`rmhmodel`'. The digested parameters are then unpacked and passed to our low-level `C` routine implementing the Metropolis-Hastings algorithm.

Other arguments to the default method of `rmh` enable the user to control the initial state, algorithm parameters, algorithm behavior and the amount of information returned. There is also an interactive visual debugger, enabling the user to display each successive state.

The `rmh` method for '`ppm`' objects generates simulated realizations from a fitted point process model. For example the following code would fit a stationary Strauss point process model to a point pattern data set `X`, and then simulate a realization of the model:

```
R> fit <- ppm(X, ~ 1, Strauss(0.1))
R> Y <- rmh(fit)
```

The `rmh` method for '`ppm`' objects first converts the fitted model object `fit` to an object of class '`rmhmodel`', and passes this object to the default method of `rmh`.

Hybrid models are accommodated by an extension of the same interface. For the default method of `rmh`, a hybrid model is specified by a vector of character strings giving the names of

the component interactions, and a list containing the parameter vectors of these interactions. For example a hybrid of the Hard Core model with hard core distance 0.03, and the Strauss model with interaction range 0.07 and interaction parameter $\gamma = 0.5$, could be simulated by

```
R> rmh(model = list(cif = c('hardcore', 'strauss'),
+    par = list(list(beta = 10, hc = 0.03),
+      list(beta = 1, gamma = 0.5, r = 0.07)),
+    w   = square(1)))
```

For the `rmh` method for '`ppm`' objects the user sees only the model objects, e.g.,

```
R> fit <- ppm(X, ~ 1, Hybrid(Hardcore(0.03), Strauss(0.07)))
R> Y <- rmh(fit)
```

9.2. Software design

Handling spatial trend

In our implementation (Baddeley and Turner 2005, 2006) of the birth-death-shift Metropolis-Hastings algorithm, the birth proposal density (described in Section 8) is chosen as $b(u, \mathbf{x}) = \beta(u) = \exp(V_1(u))/B$ where V_1 is the first order term in the Gibbs factorization (13) of the target density, and $B = \int_W \exp(V_1(u))\, \mathrm{d}u$ is the normalizing constant.

This choice of proposal density has the advantage that the first order potential terms cancel in the ratio $\lambda(u, \mathbf{x})/b(u, \mathbf{x})$ so that computation of the acceptance probability involves only the higher order interaction potentials $V_k, k \geq 2$.

Accordingly it is sufficient to implement low-level C code for the birth-death-shift Metropolis-Hastings algorithm for a *stationary* point process with conditional intensity

$$\bar{\lambda}(u, \mathbf{x}) = B \exp(\theta^\top G(u, \mathbf{x})),$$

where $B > 0$ is a known constant and $G(u, \mathbf{x})$ is composed of interpoint interaction terms of order 2 and higher. The C code expects input giving a stream of proposal points which are independent and uniformly distributed over the simulation window. However when this C routine is called from the default method of `rmh`, the proposal points are instead generated from the nonuniform density $\beta(u)$. The result is a simulation from the Metropolis-Hastings algorithm for the nonstationary target process with conditional intensity

$$\lambda(u, \mathbf{x}) = B\beta(u) \exp(\theta^\top G(u, \mathbf{x})) = \exp(\theta^\top V_1(u) + \theta^\top G(u, \mathbf{x})).$$

It is straightforward to generate proposals with density $\beta(u)$ by the rejection method.

Interactions

In our original C code for Metropolis-Hastings simulation, the calculation of the conditional intensity for each model was implemented by hand, in a separate C function for each model. To simulate a chosen model, the name of the model would be passed as a character string from R to C, and the code would look up an internal lookup table which contained a map between model

names (as character strings) and conditional intensity functions (as function pointers). The selected conditional intensity function was then invoked in the generic Metropolis-Hastings algorithm. Other tasks (such as the initialization of model parameters) were handled by a similar mechanism.

Hybrids

To adapt the C code to hybrids, we simply allowed the model name to be a *vector* of character strings. Inside the C code, the conditional intensity function pointer was replaced by a vector of function pointers, to be executed successively to evaluate $\lambda_j(u, \mathbf{x})$ for $j = 1, 2, \ldots, m$. Some extra book-keeping was required to pass irregular parameters correctly to each of the conditional intensity functions.

Note that there are cases where the hybrid is integrable although some of the components are not integrable. An example is the hybrid Strauss-hardcore model which is integrable for all values of the Strauss interaction parameter γ, while the Strauss density itself is only integrable when $\gamma \leq 1$. Consequently, some of the C and R code for checking validity of models must be modified to handle hybrids.

Supporting user options efficiently

Our simulation code supports many options, some of which are time-consuming to perform, such as the interactive debugger. To maximize speed, the C code for the simulation algorithm is compiled several times, with different combinations of these options. Fragments of code are included or excluded (using the C preprocessor) depending on the options selected. This yields a suite of C routines accommodating each combination of options as efficiently as possible. There are currently 16 different routines. The R code selects the routine appropriate to the user's choices, and executes it. We have found that this strategy also makes the code easy to maintain, since there is still only a single source file for the simulation loop.

10. Simulation study

Finally we are able to demonstrate some of the capabilities of the new software. Figure 4 and Table 1 show the results of a typical simulation study of the performance of maximum profile pseudolikelihood for a hybrid model.

The model is a hybrid of a Hardcore process with base intensity $\beta = 300$ and hard core radius $c = 0.04$ with a Strauss process with base intensity 1, interaction distance $r = 0.07$ and interaction strength $\gamma = 0.5$ in the unit square. One hundred realizations of the model were generated using a coupling-from-the-past algorithm (see Sections 8.2 and 9). The model was fitted to each simulated realization by profile maximum pseudolikelihood (Section 6.3) with the parameter r incremented in steps of 0.001.

The hardcore distance c was estimated by direct calculation. However, the maximum pseudolikelihood estimate $\hat{c} = m(\mathbf{x})$ (see Lemma 7) was not used: this leads to computational problems associated with testing the equality of floating-point numbers, since there are interpoint distances exactly equal to $\hat{c}$. A simple solution is to take

$$\tilde{c} = \frac{n}{n+1} m(\mathbf{x}), \tag{47}$$

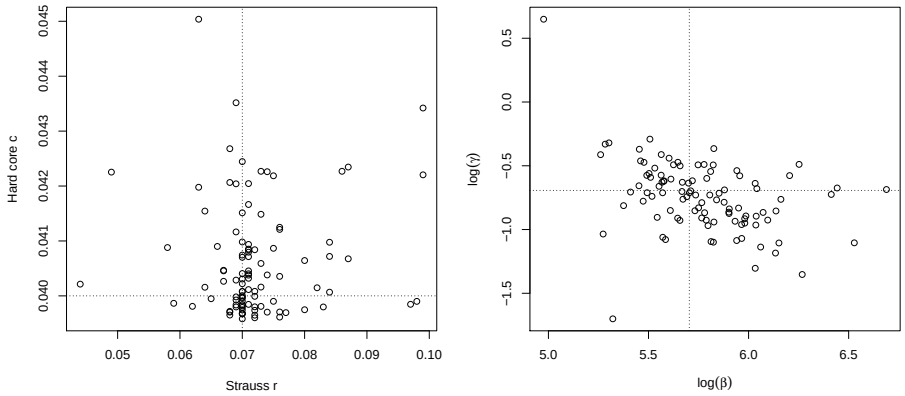

Figure 4: Estimates of the parameters of a hybrid of a Hardcore process and Strauss process using profile maximum pseudolikelihood. *Left:* irregular parameters; *Right:* canonical coefficients.

	$\log\beta$	$\log\gamma$	r	c
Bias	0.067	−0.050	0.002	0.0006
SD	0.289	0.281	0.008	0.0010

Table 1: Estimates of bias and standard deviation for the parameters of the hybrid Strauss-hard core model, with r estimated by profile maximum pseudolikelihood on a grid of step size 0.001, and c estimated by bias-corrected maximum pseudolikelihood.

where $n = n(\mathbf{x})$ is the number of data points. This is analogous to the unbiased estimator for the endpoint of a uniform distribution on the real line, so it may also be expected to reduce bias in the estimation of c.

The left panel of Figure 4 shows the resulting estimates of the irregular parameters r and c (irregular in the sense of Section 6.3). The right panel of Figure 4 shows the estimates of the canonical regular parameters $\log\gamma$ and $\log\beta$ (regular in the sense of Section 6.1). The correct values ($\beta = 300$, $\gamma = 0.5$, $r = 0.07$, $c = 0.04$) are shown as dashed lines. Estimated bias and standard deviation are shown in Table 1. These results indicate good performance for the estimators of each parameter.

11. Analysis of Gordon Square data

11.1. Background

One applied context in which hybrid models may be useful is in the study of human social interaction. The different ways in which humans arrange themselves, when they live together in groups or gather temporarily, have long been of interest to geographers, sociologists and archaeologists amongst others. As long as we are happy to represent an individual person, household or settlement as a simple point event, then point process modeling becomes an attractive approach, both for characterizing a single snapshot of human activity and for comparing many such patterns across different time periods and different contexts. Density

and distance-based summary statistics have been used for this purpose for many decades, particularly for considering patterns of human settlement (Blankholm 1991; Hietala 1984; Hodder and Orton 1976; Robinson 1998), but we would argue that inhomogeneous model-fitting methods, better treatment of anisotropy and hybrid interaction terms now offer new opportunities.

Interpersonal behavioral interactions provide a possible explanation for the way smaller numbers of individuals arrange themselves. The rationale is even more clear-cut than it is for large settlement patterns. In small free-forming groups, for example, there are often (a) spatial patterns of inhibition at very small scales, as people seek to preserve their own personal space (Sommer 1966), and because we ignore the physical size of a human body when we represent it as an ideal point; and (b) patterns of both attraction and inhibition at larger scales, as people reserve certain spaces for particular activities, socialize in groups of friends, cooperate over specific communal tasks and/or avoid other competing groups (Hall 1966; Hendrick, Giesen, and Coy 1974). A comparison of ethnoarchaeological studies (e.g., Binford 1978; Yellen 1977) readily demonstrates that many of these spatial arrangements vary cross-culturally, in different kinds of social context or depending on the mix of genders, ages, perceived statuses and abilities (Baxter 1970). Hence it is of great interest to find ways to fit different models to these different observed spatial patterns in a more explicit and comparative way.

11.2. Description of data

With these applied goals in mind, we turn to Figure 1. These data give a snapshot of people sitting on the ground in Gordon Square (London, U.K.) on a sunny day (3pm, 2011-09-28), ignoring those who were standing, walking or sitting on the benches that line the edge of the square. An approximate record of the locations of all seated individuals was made by hand onto an accurate map of land parcels and built structures in the square. About thirty photographs were also taken from different ground-level positions and used to refine the initial sketched positions. There are 99 seated individuals in a total area of 2164 square meters giving an average intensity of 0.046 people per square meter.

The point pattern in Figure 1 is qualitatively non-random, with a few larger, ring-shaped clusters of friends talking to each other on the grass, and a variety of others sitting in ones, two and threes. There also appears to be an increasing trend in overall intensity from south-west to north-east, as the afternoon shadows were lengthening to the extent that some people were preferring north-eastern spots with greater sunshine.

Without knowledge of the application context, many experts in spatial statistics would probably suggest that Figure 1 contains clearly-defined clusters, and should be modeled using a cluster process model. However, simple cluster models such as a Neyman-Scott process (Neyman and Scott 1958) would be inappropriate, since the spatial organization of people within a group in Figure 1 is not independent, and different groups of people in Figure 1 are not independent. The model would have to be a more general cluster process, including spatial interaction between the members of a cluster, and interaction between clusters. Thus the inescapable key question is to identify interactions between individuals which give rise to the observed spatial pattern. For that reason, a Gibbs model is of primary interest.

11.3. Spatial inhomogeneity

Figure 5 shows two estimates of the spatial trend. The left panel is a kernel estimate of

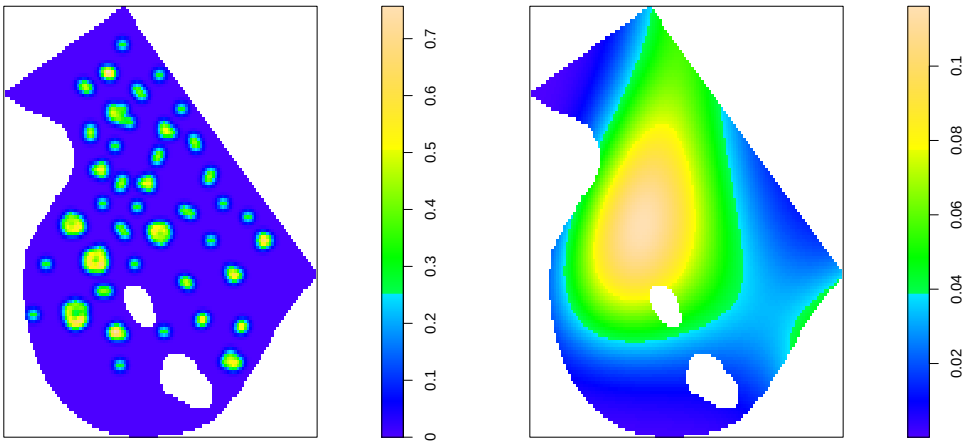

Figure 5: Spatial inhomogeneity of the Gordon Square data. *Left:* Nonparametric kernel estimate of intensity, with bandwidth selected by Berman-Diggle cross-validation. *Right:* Parametric estimate of intensity, log-cubic function of coordinates, fitted by maximum likelihood. Intensity units are people per square meter.

intensity (Diggle 1985), using an isotropic Gaussian kernel, with kernel bandwidth selected by cross-validation, using the method of Berman and Diggle (1989). The right panel is a parametric estimate, a log-cubic function of the Cartesian coordinates, fitted by maximum likelihood using the Berman-Turner device (Berman and Turner 1992; Baddeley and Turner 2000).

The left panel is clearly under-smoothed for our purposes. This is an interesting example where the bandwidth selection method of Berman and Diggle (1989) appears to be "fooled" by the spatial organization at medium scales of about 2 meters. The method of Berman and Diggle (1989) was derived by assuming a Cox point process, and generally performs well for patterns that show positive association between points; apparently the inhibition between groups of sitters in this dataset has caused underestimation of the bandwidth.

We generated simulated realizations (not shown) of a Poisson process with the intensity estimated in the right panel of Figure 5. Comparison of the simulations and the original data suggests that there is a tendency in the Gordon Square data for people to avoid sitting close to the edge of the grassy area. To assess the evidence for this tendency we fitted a model that includes a term for proximity to the boundary of the window: the intensity is of the form

$$\lambda(x, y) = \exp(\zeta I_t(x, y) + p_3(x, y)), \tag{48}$$

where ζ is a parameter, $I_t(x, y) = \mathbf{1}\{b(x, y) \le t\}$ where $b(x, y)$ is the distance from location (x, y) to the boundary of the window, t is a threshold, and $p_3(x, y)$ is a cubic polynomial.

We fitted the model by maximum profile likelihood using the **spatstat** function `profilepl`. For each value of t the algorithm estimates the parameter ζ and the coefficients of p_3 by maximum likelihood using the Berman-Turner device; then the likelihood is maximized over t. The profile log-likelihood shown in Figure 6 has a clear maximum at $t = 3.0$ meters. The fitted coefficient $\hat{\zeta} = -2.26$ is highly significant ($p < 0.01$).

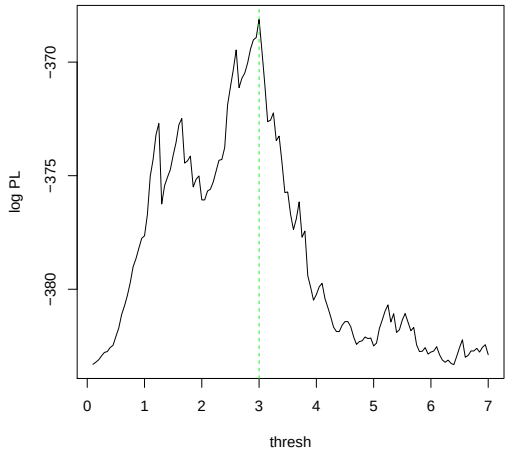

Figure 6: Profile log-likelihood for the distance threshold parameter t in the model (48).

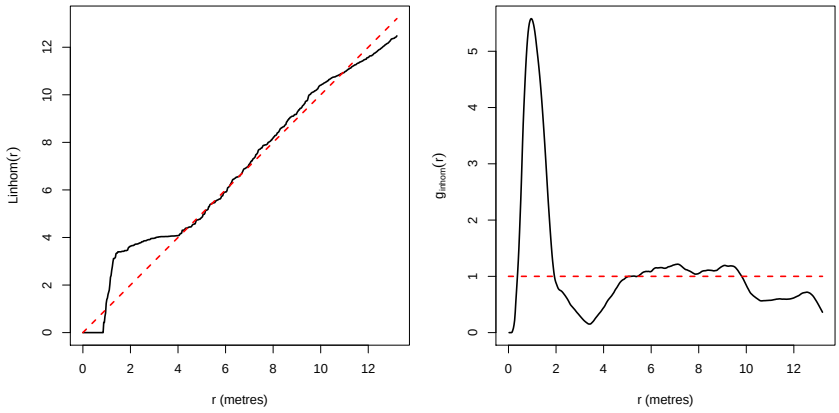

Figure 7: Inhomogeneous L-function (*Left*) and pair correlation function (*Right*) for the Gordon Square data using fitted log-cubic intensity. Solid lines: empirical estimate. Dotted lines: expected value for a Poisson process.

11.4. Evidence for interaction

Figure 7 shows estimates of the inhomogeneous L-function and inhomogeneous pair correlation function (Baddeley, øller, and Waagepetersen 2000) for the Gordon Square data using the fitted log-cubic intensity from the right panel of Figure 5. These summaries show clear evidence of a 'hard core' effect at distances less than about 1 meter, combined with a strong attraction at about 2 meters. A plausible interpretation is that individuals maintain a personal space (Sommer 1966) while forming tightly clustered groups.

11.5. No edge correction

The "standard model" for point process analysis assumes that the point process exists on the whole Euclidean plane, or on a very large region, but is only observed through a bounded window (Baddeley 2010). Edge effects intrude into the analysis because a point near the

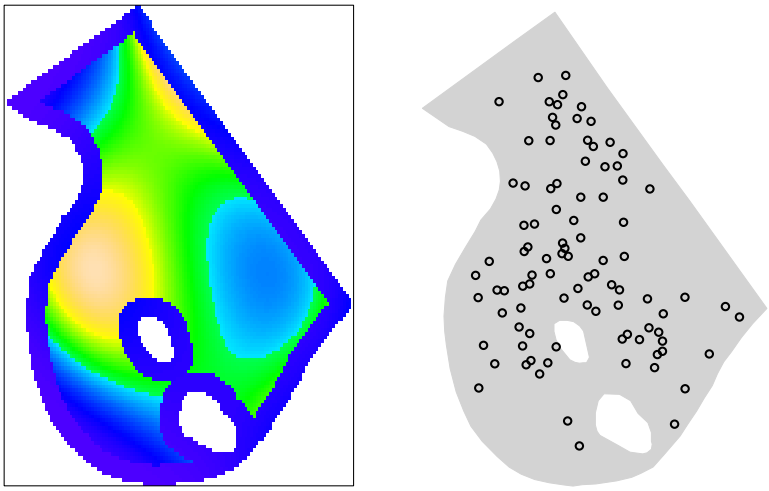

Figure 8: Model 1 (log cubic trend, boundary effect and hard core interaction) for Gordon Square data. *Left:* fitted trend. *Right:* simulated realization of model. Simulated pattern has 99 points.

edge of the observable window might have an unseen neighbor just outside the window. Edge corrections are often advisable in this case (Illian *et al.* 2008).

The "standard model" is inappropriate for the Gordon Square data because people cannot or do not sit outside the boundary of the square. This suggests that we model the data as a finite point process living inside the given boundary, and correspondingly **avoid** edge correction in the exploratory analysis.

11.6. First model

We now fit an inhomogeneous hard core model, the hybrid of our initial Poisson model (48) with a hard core process. That is, the conditional intensity of *model 1* is

$$\lambda(u, \mathbf{x}) = \begin{cases} \exp(\zeta I_t(u) + p_3(u)) & \text{if } d(u, \mathbf{x}) > c \\ 0 & \text{if } d(u, \mathbf{x}) \leq c \end{cases} \tag{49}$$

where c is the hard core threshold distance, and $d(u, \mathbf{x})$ is the shortest distance from location u to the pattern $\mathbf{x}$. As in the initial model (48), $p_3(u)$ denotes a cubic polynomial in the Cartesian coordinates of the point u, and $I_t(u)$ is a term for proximity to the boundary. The maximum likelihood estimate of the hard core threshold distance c is the minimum nearest-neighbor distance, $\hat{c} = 0.86$ meters (Lemma 7 in Section 6.3). We take the approximately unbiased estimate $\tilde{c} = n\hat{c}/(n+1) = 0.85$ meters where $n = 99$ is the number of data points. A simulated realization of the fitted model is shown in Figure 8.

The simulated pattern does not appear to exhibit the same strong clustering as the data. This is corroborated by the diagnostics in Figure 9. The top left panel of Figure 9 shows the residual K-function for the fitted model. As explained in Baddeley *et al.* (2011) and in Section 6.4 above, the residual K-function is the difference between the observed K-function and the *"compensator"* of the K-function based on the fitted model. The compensator effectively 'adjusts' for the behavior predicted by the model (in this case, it adjusts for the fact that

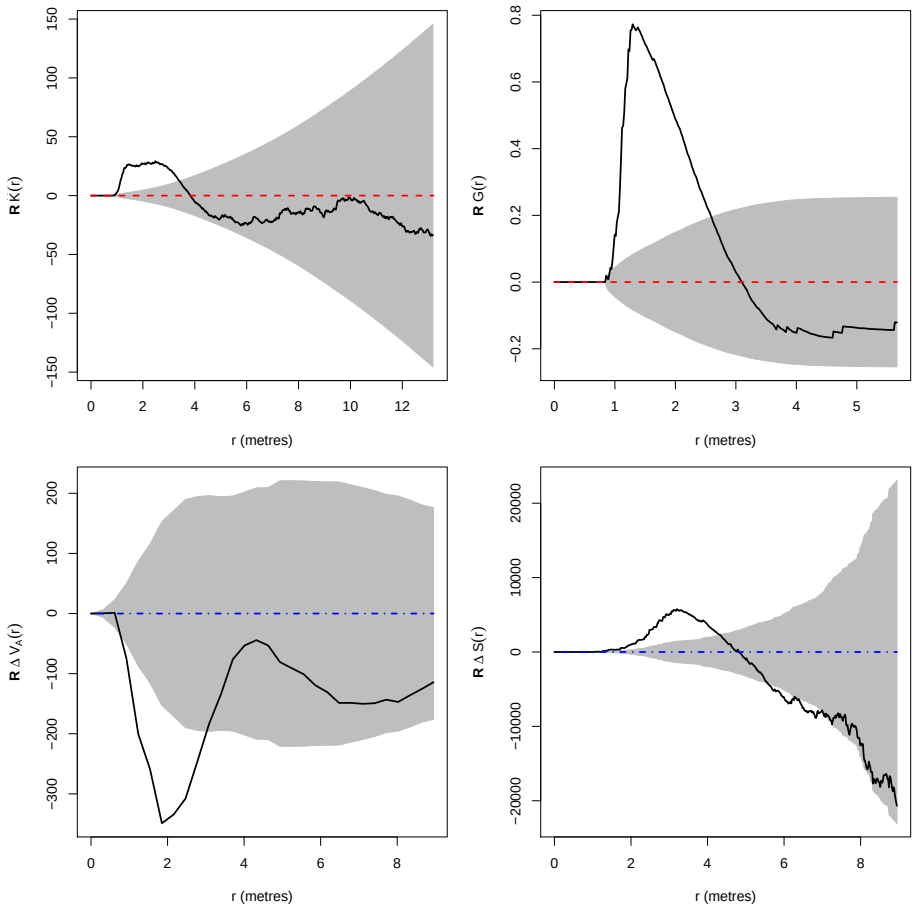

Figure 9: Diagnostics for model 1 of the Gordon Square data. *Top left:* residual K-function; *Top right:* residual G-function; *Bottom left:* pseudoscore residual for area-interaction; *Bottom right:* pseudoscore residual for triplet-interaction.

the model says there is a hard core). If the model is true, the residual should be zero (on average).

More importantly for us, the residual K-function for a fitted model is closely related to a score test, where the null hypothesis is the current fitted model, and the alternative hypothesis is a hybrid of the current fitted model with a Strauss process. Nonzero values of the residual K-function at a particular distance r suggest that it would be appropriate to include a Strauss process component, with interaction range r.

Note that the upper and lower limits of the shaded regions in Figure 9 do not have an exact significance interpretation. They are *approximate* two-sigma significance bands based on a heuristic; they are very unreliable for small values of r. It is possible of course to use simulation to obtain exact significance bands, at a much greater computational expense.

The upper right panel of Figure 9 shows the residual G-function for the same model. This is the difference between the observed nearest neighbor distance function G and the "compensator" of G based on the fitted model. It is related to a score test where the null hypothesis is the current fitted model, and the alternative hypothesis is a hybrid of the current fitted

Figure 10: Simulated realization of model 2 (logcubic trend, boundary effect, hybrid hard core-Geyer interaction) for Gordon Square data. Pattern contains 87 points.

model with a Geyer saturation process with saturation parameter $s = 1$.

The bottom panels show the pseudoscore residuals for area-interaction and triplet-interaction models as defined in Baddeley *et al.* (2011). Again these are closely related to the score test of model 1 where the alternative hypothesis is a hybrid of model 1 with the area-interaction process (Baddeley and van Lieshout 1995) or the triplet-interaction process (Geyer 1999) respectively.

11.7. Second model

The residual plots above suggest that, after allowing for the hard core, there is positive association between points, which peaks at a distance of about 1.35 meters. The sharpness of the peak for the residual G-function encourages us to include a Geyer saturation interaction.

Accordingly, we now fit a hybrid interaction consisting of the previously-fitted hard core at $\tilde{c} = 0.85$ meters and a Geyer saturation model with interaction radius about 1.4 meters. We select the parameters by maximum profile pseudolikelihood (Baddeley and Turner 2000, 2005) giving an interaction radius $r = 1.46$ meters and saturation parameter $s = 2$, with interaction parameter $\gamma = 3.96$ indicating very strong positive association.

The degree of the polynomial appearing in the spatial trend could be selected using the pseudolikelihood counterpart of Akaike's Information Criterion. We have chosen cubic polynomials to reduce computation time.

Figure 10 shows a simulated realization of model 2. It seems to reproduce some of the short-scale clustering visible in the data, but the orderly spacing between clusters in the Gordon Square data is absent in the simulation.

Figure 11 shows the residual diagnostics for model 2. They suggest an improvement over model 1. There is a suggestion of slight inhibition at distance 4 meters.

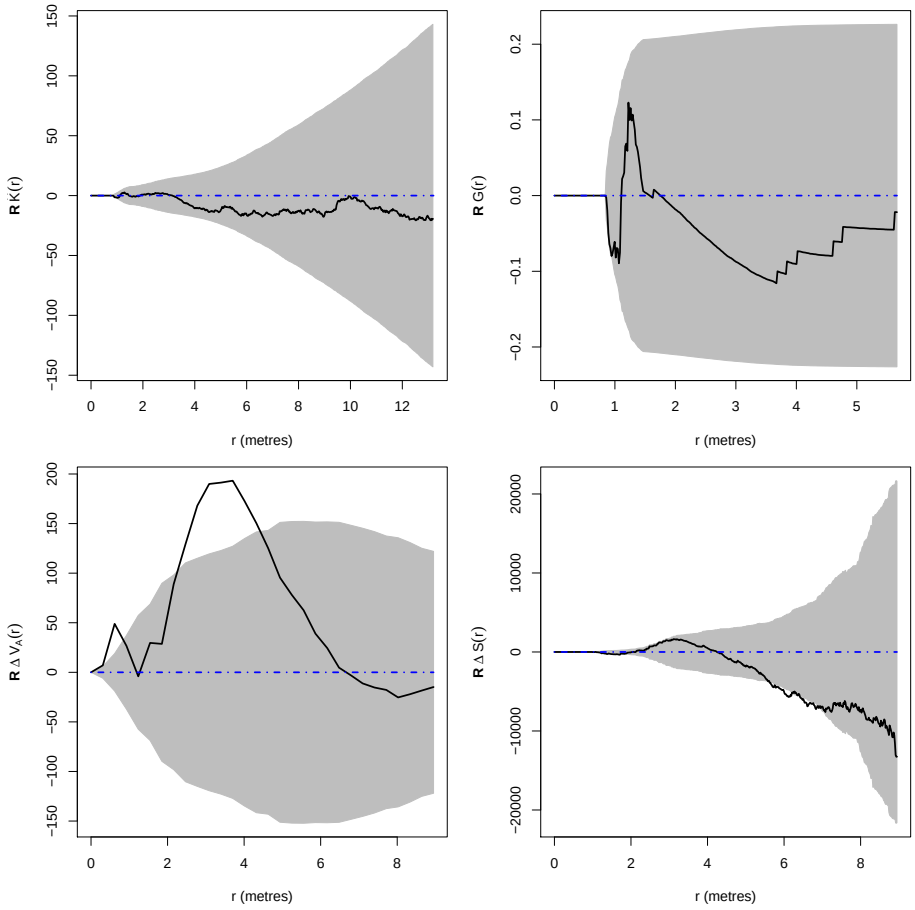

Figure 11: Diagnostics for model 2 of the Gordon Square data. *Top left:* residual K-function; *Top right:* residual G-function; *Bottom left:* pseudoscore residual for area-interaction; *Bottom right:* pseudoscore residual for triplet-interaction.

11.8. Third model

The diagnostics for the second model suggest that a further negative association should be included. The erratic behavior of the G residual for the second model suggests that there is some extra interaction at a distance of about 1.5 meters. The area residual suggests an extra interaction at 2–4 meters.

When the simulated realization of the second model is compared with the original data, focusing on patterns a scale of about 1.5 meters, it is noticeable that the original data contain relatively few triangles of people at this short distance: most of the arrangements of people are circles of about 5 meters in diameter, in which each person is flanked by two close neighbors.

Accordingly an interesting possibility is to add a triplet-interaction component (Geyer 1999) having the *same* interaction distance as the Geyer component (currently estimated as 1.46 meters). The hybrid is fitted by maximum profile pseudolikelihood. The fitted model comprises (a) hard core at $c = 0.85$ meters, (b) attractive Geyer saturation interaction at $r = 1.40$ meters with saturation $s = 2$ and strength $\gamma = 4.82$, and (c) inhibitory triplet-interaction at $r = 1.40$ meters with strength $\gamma = 0.245$. The Geyer component encourages more 'sociable'

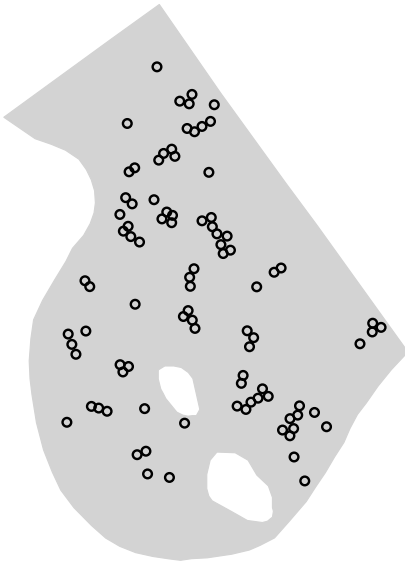

Figure 12: Simulated realization of model 3 (log-cubic offset trend, border effect and hybrid hard core, attractive Geyer interaction, and inhibitive triplet-interaction) for Gordon Square data. Pattern has 92 points.

arrangements where more people sit together, but with no extra benefit when a person has more than 2 neighbors; the triplet interaction component penalizes 'crowding'.

Figure 12 shows a simulated realization of model 3. This looks more realistic than the previous simulations. It appears to capture the short-range interactions quite well, but not the longer-range interaction.

Figure 13 shows the residual diagnostics for model 3. These plots suggest that the model has captured most of the interaction at short scales. At larger scales (a distance of about 4 meters is suggested by the area-interaction residual) there are additional inhibition effects which have not been captured by the model.

The outcome of the analysis is moderately but not completely satisfactory, because it has not completely captured the intricate interactions that are present at all scales. In particular the ring-shaped clusters in Figure 1 have not been reproduced in the simulations. Nearest-neighbor interactions such as the connected component model (Baddeley and Møller 1989) could well be used for this purpose, but these are beyond the scope of the present paper.

Nonetheless, this example serves its intended purpose. It demonstrates how easy it is to formulate and fit hybrid models to data. It also demonstrates that Gibbs models can exhibit spatial clustering.

A more searching approach to this particular application would be to determine which way each person is facing, and to consider this direction as a mark (or third coordinate). Then a Gibbs model could include interactions that depend on direction: whether people are more-or-less facing one another, or have their backs turned to one another. The original photographs suggest that it may also be important to take account of each individual's gender, and of the locations of people sitting on the benches that border the park, who are not represented in Figure 1.

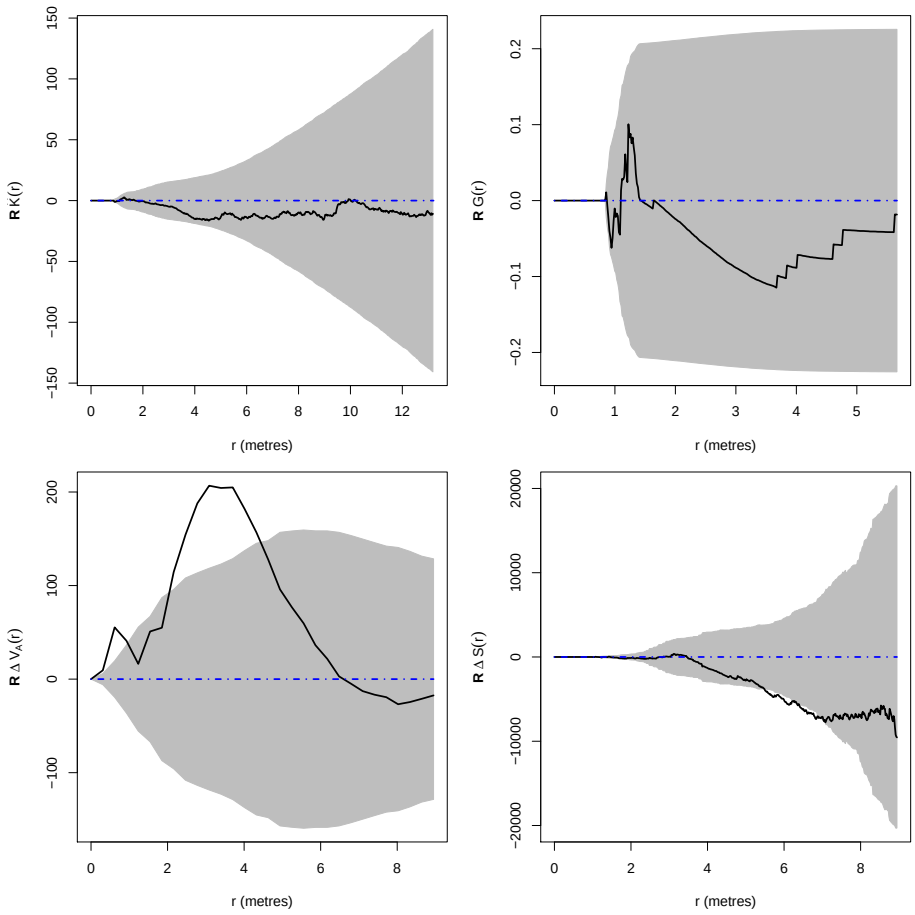

Figure 13: Diagnostics for model 3 of the Gordon Square data. *Top left:* residual K-function; *Top right:* residual G-function; *Bottom left:* pseudoscore residual for area-interaction; *Bottom right:* pseudoscore residual for triplet-interaction.

12. Further commentary

Hybrids, mixtures and superpositions

Note that there is a distinction between hybrids, mixtures, and superpositions of point processes. In a *mixture* of two point processes $\mathbf{X}$ and $\mathbf{Y}$, the resulting mixture process $\mathbf{Z}$ is either equal to $\mathbf{X}$ (with probability p) or equal to $\mathbf{Y}$ (with probability $1-p$) where p is the mixture probability. In a *superposition* of two point processes $\mathbf{X}$ and $\mathbf{Y}$, the superposition process $\mathbf{Z}$ is equal to $\mathbf{X} \cup \mathbf{Y}$, the union of the points of $\mathbf{X}$ and the points of $\mathbf{Y}$. Typically $\mathbf{X}$ and $\mathbf{Y}$ are assumed to be independent.

These three concepts are different. Suppose $\mathbf{X}$ is a hard core process and $\mathbf{Y}$ is a Poisson process. The hybrid of $\mathbf{X}$ and $\mathbf{Y}$ is another hard core process. The mixture of $\mathbf{X}$ and $\mathbf{Y}$ has realizations that are either Poisson or hard core patterns. The superposition of $\mathbf{X}$ and $\mathbf{Y}$ is not a hard core process but does exhibit spatial inhibition. Thus, the hybrid, mixture, and superposition of $\mathbf{X}$ and $\mathbf{Y}$ are three different processes.

Superposition of independent point processes is a useful model in problems where the compo-

nent processes must be separated, for example in order to recognize a minefield in the presence of noise (Cressie and Collins 2001; Walsh and Raftery 2002, 2005). It is not an appropriate description of point pattern data such as Figure 1.

A mixture of two Cox processes is again a Cox process: or in formal terms, the Cox processes are closed under mixtures. Similarly, Cox processes and independent cluster processes are closed under superposition. However, Cox and cluster processes are not closed under hybridization, and it would be useful to know more about hybrids of cluster processes.

Possible extensions

The approach of this paper extends to multitype point process models, where points are classified into several discrete types. This is beyond the scope of the present paper.

A pairwise interaction process (5) whose pair potential is a step function with m jumps, can be represented as the hybrid of m different Strauss process densities (7). Similarly we might wish to represent a general pairwise interaction process as the hybrid of an infinite number of Strauss process densities. This requires a concept of 'continuous' hybridization.

Intuitively it is clear that the class of potential functions $\log h$ (where h ranges over all unnormalized densities that are Ruelle stable) is a cone. Hybridization is a multilinear map on this cone. The map $\log h \mapsto \log \lambda$ (where λ is the conditional intensity) is convex linear. The mathematical groundwork required to establish a theory of continuous hybridization is beyond the scope of this paper.

Computer vision applications

One approach to computer vision is to model the observed scene as a superposition of geometrical objects. Stochastic models of spatial arrangements of geometrical objects, based on point process models, can then be used to impose penalties on the more unlikely configurations (e.g., Baddeley and van Lieshout 1992a,b; van Lieshout and Baddeley 1995). The ability to construct more flexible point process models, especially multi-scale models, is therefore potentially valuable for computer vision, especially multiresolution vision problems.

Acknowledgments

The pattern of people sitting in Gordon Square was first used as a hypothetical example of a point process exhibiting complex multi-scale interactions for teaching purposes by a colleague, Mark Lake (UCL), with further input from Enrico Crema, and we are grateful to both of them for further discussion and suggestions.

We thank Jianbao Chen and Gopalan Nair for informative discussion about the density (26). We thank Mark Berman and the anonymous referees for their insightful comments.

References

Ambler GK (2002). *Dominated Coupling from the Past and Some Extensions of the Area-Interaction Process*. Ph.D. thesis, Department of Mathematics, University of Bristol.

Ambler GK, Silverman BW (2004). "Perfect Simulation of Spatial Point Processes Using Dominated Coupling from the Past with Application to a Multiscale Area-Interaction Process." *Technical report*, Department of Mathematics, University of Bristol. arXiv: 0903.2651v1 [stat.ME], URL http://arXiv.org/abs/0903.2651v1.

Assunção R (2003). "Score Tests for Pairwise Interaction Parameters of Gibbs Point Processes." *Brazilian Journal of Probability and Statistics*, **17**(2), 169–178.

Baddeley A (2010). "Modelling Strategies." In AE Gelfand, PJ Diggle, M Fuentes, P Guttorp (eds.), *Handbook of Spatial Statistics*, chapter 20, pp. 339–369. CRC Press.

Baddeley A, Chang YM, Song Y (2012). "Leverage and Influence Diagnostics for Spatial Point Processes." *Scandinavian Journal of Statistics*, **40**(1), 86–104.

Baddeley A, Møller J, Pakes AG (2008). "Properties of Residuals for Spatial Point Processes." *The Annals of the Institute of Statistical Mathematics*, **60**(3), 627–649.

Baddeley A, Nair G (2012a). "Approximating the Moments of a Spatial Point Process." *Stat*, **1**(1), 18–30.

Baddeley A, Nair G (2012b). "Fast Approximation of the Intensity of Gibbs Point Processes." *Electronic Journal of Statistics*, **6**, 1155–1169.

Baddeley A, øller JM, Waagepetersen R (2000). "Non- and Semiparametric Estimation of Interaction in Inhomogeneous Point Patterns." *Statistica Neerlandica*, **54**(3), 329–350.

Baddeley A, Rubak E, Møller J (2011). "Score, Pseudo-Score and Residual Diagnostics for Spatial Point Process Models." *Statistical Science*, **26**(4), 613–646.

Baddeley A, Turner R (2000). "Practical Maximum Pseudolikelihood for Spatial Point Patterns." *Australian and New Zealand Journal of Statistics*, **42**(3), 283–322.

Baddeley A, Turner R (2005). "**spatstat**: An R Package for Analyzing Spatial Point Patterns." *Journal of Statistical Software*, **12**(6), 1–42. URL http://www.jstatsoft.org/v12/i06/.

Baddeley A, Turner R (2006). "Modelling Spatial Point Patterns in R." In A Baddeley, P Gregori, J Mateu, R Stoica, D Stoyan (eds.), *Case Studies in Spatial Point Pattern Modelling*, number 185 in Lecture Notes in Statistics, pp. 23–74. Springer-Verlag.

Baddeley A, Turner R, Møller J, Hazelton M (2005). "Residual Analysis for Spatial Point Processes." *Journal of the Royal Statistical Society B*, **67**(5), 617–666.

Baddeley AJ, Møller J (1989). "Nearest-Neighbour Markov Point Processes and Random Sets." *International Statistical Review*, **57**(2), 89–121.

Baddeley AJ, van Lieshout MNM (1992a). "ICM for Object Recognition." In Y Dodge, J Whittaker (eds.), *COMPSTAT – Proceedings in Computational Statistics*, pp. 271–286. Physica-Verlag, Heidelberg.

Baddeley AJ, van Lieshout MNM (1992b). "Object Recognition Using Markov Spatial Processes." In *Proceedings of the 11th IAPR International Conference on Pattern Recognition*, volume 2, pp. 136–139. IEEE Computer Society Press. Conference B: Pattern Recognition Methodology and Systems.

Baddeley AJ, van Lieshout MNM (1995). "Area-Interaction Point Processes." *The Annals of the Institute of Statistical Mathematics*, **47**(4), 601–619.

Barndorff-Nielsen OE (1978). *Information and Exponential Families in Statistical Theory*. John Wiley & Sons.

Baxter JC (1970). "Interpersonal Spacing in Natural Settings." *Sociometry*, **33**(4), 444–456.

Berman M, Diggle P (1989). "Estimating Weighted Integrals of the Second-Order Intensity of a Spatial Point Process." *Journal of the Royal Statistical Society B*, **51**(1), 81–92.

Berman M, Turner TR (1992). "Approximating Point Process Likelihoods with GLIM." *Journal of the Royal Statistical Society C*, **41**(1), 31–38.

Berthelsen KK, Møller J (2002). "A Primer on Perfect Simulation for Spatial Point Processes." *Bulletin of the Brazilian Mathematical Society*, **33**, 351–367.

Berthelsen KK, Møller J (2003). "Likelihood and Non-Parametric Bayesian MCMC Inference for Spatial Point Processes Based on Perfect Simulation and Path Sampling." *Scandinavian Journal of Statistics*, **30**, 549–564.

Besag J (1977). "Some Methods of Statistical Analysis for Spatial Data." *Bulletin of the International Statistical Institute*, **47**, 77–91.

Binford LR (1978). "Dimensional Analysis of Behavior and Site Structure: Learning from an Eskimo Hunting Stand." *American Antiquity*, **43**, 330–61.

Blankholm HP (1991). *Intrasite Spatial Analysis in Theory and Practice*. Aarhus University Press, Aarhus.

Coeurjolly JF, Rubak E (2012). "Fast Covariance Estimation for Innovations Computed from a Spatial Gibbs Point Process." *Technical Report 3*, Centre for Stochastic Geometry and Bioimaging, Aarhus, Denmark.

Cressie N, Collins LB (2001). "Patterns in Spatial Point Locations: Local Indicators of Spatial Association in a Minefield with Clutter." *Naval Research Logistics*, **48**(5), 333–347.

Diggle PJ (1983). *Statistical Analysis of Spatial Point Patterns*. Academic Press, London.

Diggle PJ (1985). "A Kernel Method for Smoothing Point Process Data." *Journal of the Royal Statistical Society C*, **34**, 138–147.

Diggle PJ (2003). *Statistical Analysis of Spatial Point Patterns*. 2nd edition. Hodder Arnold, London.

Geyer CJ (1999). "Likelihood Inference for Spatial Point Processes." In OE Barndorff-Nielsen, WS Kendall, MNM van Lieshout (eds.), *Stochastic Geometry: Likelihood and Computation*, chapter 3, pp. 79–140. Chapman and Hall/CRC.

Geyer CJ (2009). "Likelihood Inference in Exponential Families and Directions of Recession." *Electronic Journal of Statistics*, **3**, 259–289.

Geyer CJ, Møller J (1994). "Simulation Procedures and Likelihood Inference for Spatial Point Processes." *Scandinavian Journal of Statistics*, **21**(4), 359–373. ISSN 0303-6898.

Hall ET (1966). *The Hidden Dimension*. Doubleday, New York.

Handcock MS, Morris M (1999). *Relative Distribution Methods in the Social Sciences*. Springer-Verlag.

Heikkinen J, Penttinen A (1999). "Bayesian Smoothing in the Estimation of the Pair Potential Function of Gibbs Point Processes." *Bernoulli*, **5**(6), 1119–1136.

Hendrick C, Giesen M, Coy S (1974). "The Social Ecology of Free Seating Arrangements in a Small Group Interaction Context." *Sociometry*, **37**(2), 262–274.

Hietala HJ (ed.) (1984). *Intrasite Spatial Analysis in Archaeology*. New Directions in Archaeology. Cambridge University Press, Cambridge.

Hodder I, Orton C (1976). *Spatial Analysis in Archaeology*. New Studies in Archaeology. Cambridge University Press, Cambridge.

Huang F, Ogata Y (1999). "Improvements of the Maximum Pseudo-Likelihood Estimators in Various Spatial Statistical Models." *Journal of Computational and Graphical Statistics*, **8**(3), 510–530.

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

Jensen JL, Künsch HR (1994). "On Asymptotic Normality of Pseudo Likelihood Estimates for Pairwise Interaction Processes." *The Annals of the Institute of Statistical Mathematics*, **46**(3), 475–486.

Jensen JL, Møller J (1991). "Pseudolikelihood for Exponential Family Models of Spatial Point Processes." *The Annals of Applied Probability*, **1**(3), 445–461.

Jones JE (1924). "On the Determination of Molecular Fields. II. From the Equation of State of a Gas." *Proceedings of the Royal Society of London A*, **106**(738), 463–477.

Kelly FP, Ripley BD (1976). "A Note on Strauss's Model for Clustering." *Biometrika*, **63**(2), 357–360.

Kendall WS, Møller J (2000). "Perfect Simulation Using Dominating Processes on Ordered Spaces, with Application to Locally Stable Point Processes." *Advances in Applied Probability*, **32**(3), 844–865.

Kendall WS, van Lieshout MNM, Baddeley AJ (1999). "Quermass-Interaction Processes: Conditions for Stability." *Advances in Applied Probability*, **31**(2), 315–342.

Klein W (1982). "Potts-Model Formulation of Continuum Percolation." *Physical Review B*, **26**(5), 2677–2678.

Küchler U, Sørensen M (1997). *Exponential Families of Stochastic Processes*. Springer-Verlag.

Mateu J, Montes F (2001a). "Likelihood Inference for Gibbs Processes in the Analysis of Spatial Point Processes." *International Statistical Review*, **69**(1), 81–104.

Mateu J, Montes F (2001b). "Pseudo-Likelihood Inference for Gibbs Processes with Exponential Families through Generalized Linear Models." *Statistical Inference for Stochastic Processes*, **4**(2), 125–154.

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

Neyman J, Scott EL (1958). "Statistical Approach to Problems of Cosmology." *Journal of the Royal Statistical Society B*, **20**(1), 1–43.

Picard N, Bar-hen A, Mortier F, Chadoeuf J (2009). "The Multi-Scale Marked Area-Interaction Point Process: A Model for the Spatial Pattern of Trees." *Scandinavian Journal of Statistics*, **36**(1), 23–41.

R Core Team (2013). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Ripley BD (1976). "The Second-Order Analysis of Stationary Point Processes." *Journal of Applied Probability*, **13**(2), 255–266.

Ripley BD, Kelly FP (1977). "Markov Point Processes." *Journal of the London Mathematical Society*, **15**(1), 188–192.

Robinson GM (1998). *Methods and Techniques in Human Geography*. John Wiley & Sons, New York.

Rowlinson JS (1980). "Penetrable Sphere Models of Liquid-Vapor Equilibrium." *Advances in Chemical Physics*, **41**, 1–57.

Ruelle D (1969). *Statistical Mechanics: Rigorous Results*. W.A. Benjamin, Reading, Mass.

Sommer R (1966). *Personal Space*. Prentice-Hall, Englewood Cliffs.

Strauss DJ (1975). "A Model for Clustering." *Biometrika*, **62**(2), 467–475.

van Lieshout MNM, Baddeley AJ (1995). "Markov Chain Monte Carlo Methods for Clustering of Image Features." In *Image Processing and its Applications*, Conference Publication no. 410, pp. 241–245. IEE.

van Lieshout MNM, Molchanov IS (1998). "Shot Noise Weighted Processes: A New Family of Spatial Point Processes." *Stochastic Models*, **14**(3), 715–734.

Walsh DCI, Raftery AE (2002). "Detecting Mines in Minefields with Linear Characteristics." *Technometrics*, **44**(1), 34–44.

Walsh DCI, Raftery AE (2005). "Classification of Mixtures of Spatial Point Processes via Partial Bayes Factors." *Journal of Computational and Graphical Statistics*, **14**(1), 139–154.

Wei BC (1998). *Exponential Family Nonlinear Models*. Springer-Verlag.

Widom B, Rowlinson JS (1970). "New Model for the Study of Liquid-Vapor Phase Transitions." *The Journal of Chemical Physics*, **52**(4), 1670–1684.

Yellen JE (1977). *Archaeological Approaches to the Present: Models for Reconstructing the Past*. Academic Press, New York.

Affiliation:

Adrian Baddeley
Centre for Exploration Targeting (M006)
University of Western Australia
35 Stirling Highway
Crawley WA 6009, Australia
and
CSIRO Computational Informatics
Perth, Australia
E-mail: Adrian.Baddeley@uwa.edu.au

Rolf Turner
Department of Statistics
University of Auckland
3A Symonds St
Auckland 1010, New Zealand
E-mail: r.turner@auckland.ac.nz

Jorge Mateu
Departamento de Matemáticas, Campus Riu Sec
Universitat Jaume I
E-12071 Castellón, Spain
E-mail: mateu@mat.uji.es

Andrew Bevan
Institute of Archaeology
University College London
31–34 Gordon Square
London WC1H 0PY, United Kingdom
E-mail: a.bevan@ucl.ac.uk

Journal of Statistical Software http://www.jstatsoft.org/
published by the American Statistical Association http://www.amstat.org/
Volume 55, Issue 11 *Submitted:* 2012-09-24
November 2013 *Accepted:* 2013-04-26