ROYAL STATISTICAL SOCIETY
DATA | EVIDENCE | DECISIONS

Journal of the Royal Statistical Society
Statistical Methodology
Series B

J. R. Statist. Soc. B (2015)
77, *Part* 4, *pp.* 853–877

Determinantal point process models and statistical inference

Frédéric Lavancier

University of Nantes, France

and Jesper Møller and Ege Rubak

Aalborg University, Denmark

[Received May 2012. Final revision August 2014]

Summary. Statistical models and methods for determinantal point processes (DPPs) seem largely unexplored. We demonstrate that DPPs provide useful models for the description of spatial point pattern data sets where nearby points repel each other. Such data are usually modelled by Gibbs point processes, where the likelihood and moment expressions are intractable and simulations are time consuming. We exploit the appealing probabilistic properties of DPPs to develop parametric models, where the likelihood and moment expressions can be easily evaluated and realizations can be quickly simulated. We discuss how statistical inference is conducted by using the likelihood or moment properties of DPP models, and we provide freely available software for simulation and statistical inference.

Keywords: Maximum-likelihood-based inference; Point process density; Product densities; Repulsiveness; Simulation; Spectral approach

1. Introduction

Spatial point process models where nearby points in the process repel each other are often used for describing point pattern data sets exhibiting regularity (in contrast with aggregated or clustered point pattern data sets). Gibbs point processes, including Markov point processes and pairwise interaction point processes (Ripley, 1977; Ripley and Kelly, 1977; Stoyan *et al.*, 1995; van Lieshout, 2000; Diggle, 2003; Møller and Waagepetersen, 2004; Illian *et al.*, 2008; Gelfand *et al.*, 2010), are usually used for modelling repulsiveness, but then in general

(a) moments are not expressible in closed form,
(b) likelihoods involve intractable normalizing constants,
(c) rather elaborate Markov chain Monte Carlo methods are needed for simulations and approximate likelihood inference and
(d) when dealing with infinite Gibbs point processes defined on $\mathbb{R}^d$, 'things' become rather complicated, e.g. conditions for existence and uniqueness as well as problems with edge effects;

see Møller and Waagepetersen (2004, 2007) and the references therein. For Gibbs point processes, as maximum likelihood inference is complicated, the most popular and much quicker

Address for correspondence: Jesper Møller, Department of Mathematical Sciences, Aalborg University, Fredrik Bajers Vej 7G, Aalborg DK-9220, Denmark.
E-mail: jm@math.aau.dk

 1369–7412/15/77853

alternative inference procedure is based on pseudolikelihood (Besag, 1977; Jensen and Møller, 1991; Baddeley and Turner, 2000; Gelfand *et al.*, 2010).

In contrast, determinantal point processes (DPPs) are a special class of repulsive point processes which have several appealing properties: considering a DPP defined on $\mathbb{R}^d$ (with $d=2$ in our examples), its distribution is specified by a kernel C defined on $\mathbb{R}^d \times \mathbb{R}^d$ (C is assumed in Section 2.3 to be a continuous complex covariance function). Then:

(a) there are simple conditions for existence of the process and uniqueness of the DPP is ensured when it exists;
(b) moments are known—by the very definition, all orders of moments are described by certain determinants of matrices with entries given by C (Section 2.2);
(c) edge effects are not a problem—the restriction of the DPP to a compact subset $S \subset \mathbb{R}^d$ is also a DPP with its distribution specified by the restriction of C to $S \times S$;
(d) the DPP restricted to S ($S \subset \mathbb{R}^d$ compact) has a density (with respect to a Poisson process)—the density is given by a normalizing constant, with a closed form expression, times the determinant of a matrix with entries given by a certain kernel $\tilde{C}$ which is obtained by a simple transformation of the eigenvalues in a spectral representation of C restricted to $S \times S$ (Section 2.5);
(e) if such a spectral representation is not explicitly known, we can approximate it in practice by a Fourier series (Section 4);
(f) the DPP can easily be simulated, basically because it is a mixture of 'determinantal projection point processes' (Section 2.4).

Because of (a)–(f), modelling and estimation for parametric families of DPPs become tractable as discussed in Sections 3–7.

Fig. 1 shows realizations in the unit square of three stationary DPPs with the same intensity of points: Fig. 1(a) is a simulation of a Poisson process, which is a special case of a DPP with no repulsion (or no interaction); Fig. 1(b) is a simulation of a DPP with moderate repulsion; Fig. 1(c) is a simulation of a DPP for the strongest case of repulsiveness (as discussed in Section 3.4) when the intensity is fixed. The point pattern data sets in Fig. 2, which will be detailed and analysed in Sections 5 and 6, have been selected to illustrate cases of repulsiveness which can be modelled by stationary or inhomogeneous DPPs. Further data sets are analysed by DPPs in the on-line supplementary material.

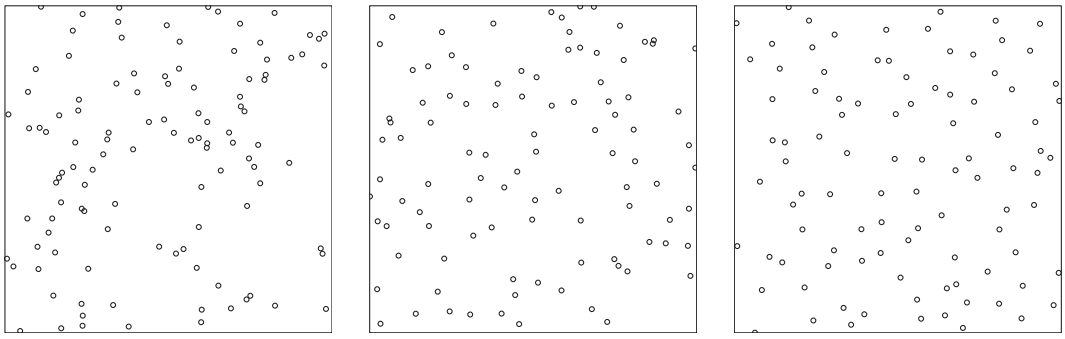

Fig. 1. Realizations of stationary DPPs within a unit square: (a) Poisson process; (b) DPP with moderate repulsion (a Gaussian DPP as described in Section 3.3); (c) stronger repulsive DPP (a jinc-like DPP as described in Section 3.4)

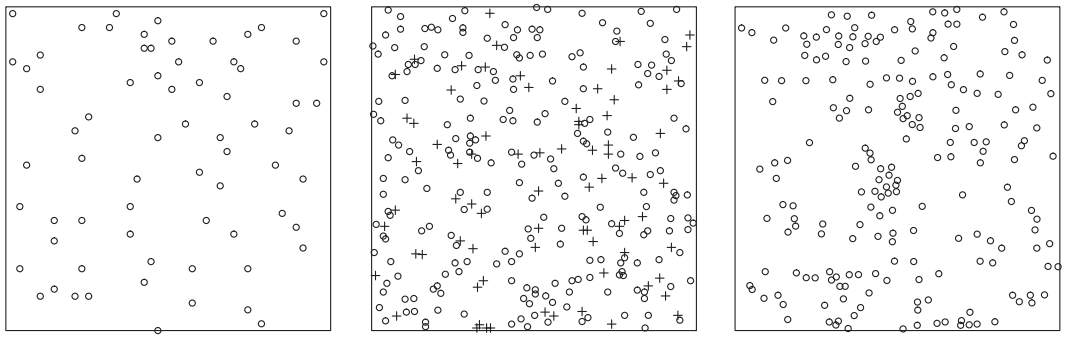

Fig. 2. (a) Locations of 69 Spanish towns in a 40 mile × 40 mile region, (b) locations of 303 cells of two types in a 0.25 mm × 0.25 mm region of a histological section of the kidney of a hamster and (c) locations of 204 seedlings and saplings of Japanese black pines in a 10 m × 10 m region

DPPs are largely unexplored in statistics but have been studied in mathematical physics, combinatorics and random-matrix theory even before the general notion was introduced in Macchi (1975). They have been used to model fermions in quantum mechanics, in classical Ginibre and circular unitary ensembles from random-matrix theory, for examples arising from non-intersecting random walks and random spanning trees, and much more, see section 2 in Soshnikov (2000) and section 4.3 in Hough *et al.* (2009). They can be defined on a locally compact space, where the two most important cases are the d-dimensional Euclidean space $\mathbb{R}^d$ and a discrete state space. Recently, DPPs have been used in machine learning (Kulesza and Taskar, 2012), where the state space is finite (basically a directory for statistical learning), and in wireless communication to model the locations of network nodes (Leonardi and Torrisi, 2013; Miyoshi and Shirai, 2013). In recent years, DPPs have also been much studied in probability theory; see Hough *et al.* (2009) and the references therein. The link between Gibbs point processes and DPPs has been studied in Georgii and Yoo (2005), where the key is the description of the Papangelou conditional intensity for a DPP. From a statistical perspective this link is of limited interest, since, for parametric families of DPPs, the Papangelou conditional intensity is not easier to handle than the likelihood, and the pseudolikelihood is in fact less easy to calculate than the likelihood. Although DPPs may be considered as a subclass of Gibbs point processes, at least when they are defined on a bounded region, we rather think of DPPs as an interesting model class in itself.

In the present paper, we address several statistical problems for DPPs defined on $\mathbb{R}^d$ (or a subregion of $\mathbb{R}^d$). Our main aims are

(i) to provide a short and accessible survey for statisticians on the definition, existence conditions, moment properties, density expressions and simulation procedures for DPPs,
(ii) to clarify when stationary DPPs exist and to develop parametric model classes for stationary DPPs (which later are extended to inhomogeneous DPPs),
(iii) to understand to what extent DPPs can model repulsiveness (or regularity or inhibition) and to demonstrate that DPPs provide useful flexible models for the description of repulsive spatial point processes,
(iv) to construct useful approximations of certain spectral decompositions appearing when dealing with likelihoods and simulations of DPPs,
(v) to discuss how statistical inference is conducted by using the likelihood or moment properties of stationary as well as inhomogeneous DPP models,

(vi) to apply our methodology on real spatial point pattern data sets showing different degrees of repulsiveness and
(vii) to provide freely available software for simulation and statistical inference.

Whereas Hough *et al.* (2009) have provided an excellent and comprehensive survey of the interest in probability theory on DPPs, our survey (item (i) above) is a less technical exposition of DPPs which provides the background material needed for our new contributions (items (ii)–(vii) above).

The paper is organized as follows. Section 2 is our tutorial (see point (i) above). In Section 3 we study stationary DPPs for several purposes: to simplify the general condition for existence of a DPP; to construct useful parametric model classes of DPPs; to understand to what extent they can model repulsiveness. Using a Fourier basis approach, we derive in Section 4 approximations of the spectral representations of the kernels C and $\tilde{C}$ (see items (d) and (e) above) which make simulation and inference feasible for parametric models of DPPs. Section 5 presents statistical inference procedures for parametric models of stationary DPPs, and we fit parsimonious parametric models to the Spanish towns and the cell data sets in Fig. 2. In Section 6, we discuss inference for inhomogeneous DPPs and we fit a model to the Japanese black pines data set in Fig. 2. Section 7 contains our concluding remarks. All proofs of our theoretical results (theorems, etc.) are deferred to the on-line supplementary material.

The statistical analyses have been conducted with R (R Development Core Team, 2011). The software that we have developed is freely available as a supplement to the `spatstat` library (Baddeley and Turner, 2005) enabling users both to simulate and to fit parametric models of DPP models.

2. Definition, existence, simulation and densities for determinantal point processes

The following subsections provide the background material for a general DPP defined on a Borel set $B \subseteq \mathbb{R}^d$ where we shall mainly consider the cases $B = \mathbb{R}^d$ and $B = S$, where S is compact. We aim at a simple exposition, though it is unavoidable at some places to be a little technical. We denote by X a simple locally finite spatial point process defined on B, i.e. we can view realizations of X as locally finite subsets of B (for measure theoretical details, see for example Møller and Waagepetersen (2004) and the references therein). We refer to the elements (or points) of X as events.

2.1. Moments for spatial point processes

Since a DPP X is defined in terms of its product density functions, $\rho^{(n)} : B^n \to [0, \infty)$, $n = 1, 2, \ldots$, we start by recalling this notion. Intuitively, for any pairwise distinct points $x_1, \ldots, x_n \in B$, $\rho^{(n)}(x_1, \ldots, x_n)\, \mathrm{d}x_1 \ldots \mathrm{d}x_n$ is the probability that, for each $i = 1, \ldots, n$, X has a point in an infinitesimally small region around x_i of volume $\mathrm{d}x_i$. For a formal definition of $\rho^{(n)}$, see for example Stoyan *et al.* (1995). We also require that $\rho^{(n)}(x_1, \ldots, x_n) = 0$ if $x_i = x_j$ for some $i \neq j$. This convention becomes consistent with definition 1 below. In particular, $\rho = \rho^{(1)}$ is the intensity function and $g(x, y) = \rho^{(2)}(x, y)/\{\rho(x)\rho(y)\}$ is the pair correlation function, where we set $g(x, y) = 0$ if $\rho(x)$ or $\rho(y)$ is 0. By our convention above, $g(x, x) = 0$ for all $x \in B$. The terminology 'pair correlation function' may be confusing, but it is adapted from physics and is commonly used by spatial statisticians. For a Poisson point process with an intensity function ρ, and for $x \neq y$, we have $g(x, y) = 1$ if $\rho(x) > 0$ and $\rho(y) > 0$.

2.2. Definition

We need the following notation. Let $\mathbb{C}$ denote the complex plane. For a complex number $z = z_1 + \mathrm{i}z_2$ (where $z_1, z_2 \in \mathbb{R}$ and $\mathrm{i} = \sqrt{(-1)}$), denote $\bar{z} = z_1 - \mathrm{i}z_2$ the complex conjugate and $|z| = \sqrt{(z_1^2 + z_2^2)}$ the modulus. For a square complex matrix A, denote $\det(A)$ its determinant. For any function $C : B \times B \to \mathbb{C}$, let $[C](x_1, \ldots, x_n)$ be the $n \times n$ matrix with (i, j)th entry $C(x_i, x_j)$. We refer to C as a kernel. In most examples of applications, the kernel will be real (the Ginibre DPP is an exception).

Definition 1. Suppose that a simple locally finite spatial point process X on B has product density functions

$$\rho^{(n)}(x_1, \ldots, x_n) = \det\{[C](x_1, \ldots, x_n)\}, \qquad (x_1, \ldots, x_n) \in B^n, \qquad n = 1, 2, \ldots. \tag{2.1}$$

Then X is called a DPP with kernel C, and we write $X \sim \mathrm{DPP}_B(C)$.

Remark 1. For $X \sim \mathrm{DPP}_B(C)$ and any Borel set $A \subseteq B$, define $X_A = X \cap A$ and denote its distribution by $\mathrm{DPP}_B(C; A)$. We also write $\mathrm{DPP}_A(C)$ for the distribution of the DPP on A with kernel given by the restriction of C to $A \times A$. Then property (c) in Section 1 follows directly from definition 1, i.e. $\mathrm{DPP}_A(C) = \mathrm{DPP}_B(C; A)$. Further, when $B = \mathbb{R}^d$, we write $\mathrm{DPP}(C)$ for $\mathrm{DPP}_{\mathbb{R}^d}(C)$, and $\mathrm{DPP}(C; A)$ for $\mathrm{DPP}_{\mathbb{R}^d}(C; A)$.

Some further remarks are in order. A Poisson process is the special case where $C(x, y) = 0$ whenever $x \neq y$. Note that $C : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{C}$ needs to be non-negative definite to ensure that $\rho^{(n)} \geqslant 0$ in expression (2.1). Thus C is a complex covariance function if and only if it is Hermitian, i.e. $C(x, y) = \overline{C(y, x)}$ for all $x, y \in \mathbb{R}^d$. If $X \sim \mathrm{DPP}(C)$, then there is no other point process satisfying condition (2.1) (lemma 4.2.6 in Hough *et al.* (2009)). By condition (2.1), the intensity function is

$$\rho(x) = C(x, x), \qquad x \in \mathbb{R}^d, \tag{2.2}$$

and the pair correlation function is

$$g(x, y) = 1 - \frac{C(x, y)\, C(y, x)}{C(x, x)\, C(y, y)} \qquad \text{if } C(x, x) > 0 \text{ and } C(y, y) > 0$$

whereas it is 0 otherwise. Further, repulsiveness is reflected by the following result. If C is Hermitian, then $g \leqslant 1$ and, for any $n = 2, 3, \ldots$,

$$\rho^{(n)}(x_1, \ldots, x_n) \leqslant \rho(x_1) \ldots \rho(x_n)$$

(the inequality is in general sharp and follows from the fact that the determinant of a complex covariance matrix is less than or equal to the product of its diagonal elements). And, if C is continuous, $\rho^{(n)}$ is also continuous and $\rho^{(n)}(x_1, \ldots, x_n) \to 0$ as the Euclidean distance $\|x_i - x_j\|$ goes to 0 for some $i \neq j$; see expression (2.1). Furthermore, for later use, let $R(x, y) = C(x, y)/\{C(x, x)\, C(y, y)\}^{1/2}$, where we set $R(x, y) = 0$ if $C(x, x) = 0$ or $C(y, y) = 0$. Note that, when C is a covariance function, R is its corresponding correlation function and

$$g(x, y) = 1 - |R(x, y)|^2, \qquad x, y \in \mathbb{R}^d. \tag{2.3}$$

Finally, we note the following transformation and independent thinning results for a DPP: if we transform $X \sim \mathrm{DPP}(C)$ by a one-to-one continuous differentiable mapping T, then $T(X) \sim \mathrm{DPP}(C_{\mathrm{trans}})$ where

$$C_{\mathrm{trans}}(x, y) = |J_{T^{-1}}(x)|^{1/2}\, C\{T^{-1}(x), T^{-1}(y)\}\, |J_{T^{-1}}(y)|^{1/2}. \tag{2.4}$$

If $X_1 \sim \mathrm{DPP}(C_1)$ and X_2 is obtained as an independent thinning of X_1 with retention probabilities $p(x)$, $x \in \mathbb{R}^d$, then $X_2 \sim \mathrm{DPP}(C_2)$ with

$$C_2(x, y) = \sqrt{p(x)}\, C_1(x, y) \sqrt{p(y)}. \tag{2.5}$$

2.3. Existence

As argued in the on-line supplementary material it is natural to make the following assumption.

Condition 1. C is a continuous complex covariance function.

Then, if we let $S \subset \mathbb{R}^d$ denote a generic compact set and $L^2(S)$ the space of square integrable functions $h : S \to \mathbb{C}$, we obtain the following result by Mercer's theorem (see for example section 98 in Riesz and Sz.-Nagy (1990)): C restricted to $S \times S$ has a spectral representation

$$C(x, y) = \sum_{k=1}^{\infty} \lambda_k \, \phi_k(x) \, \overline{\phi_k(y)}, \qquad (x, y) \in S \times S, \tag{2.6}$$

with absolute and uniform convergence of the series, and where

(a) the set of eigenvalues $\{\lambda_k\}$ is unique, each $\lambda_k \neq 0$ is real and has finite multiplicity and the only possible accumulation point of the eigenvalues is 0, and
(b) the eigenfunctions $\{\phi_k\}$ form an orthonormal basis of $L^2(S)$, i.e.

$$\int_S \phi_k(x) \, \overline{\phi_l(x)} \, \mathrm{d}x = \begin{cases} 1 & \text{if } k = l, \\ 0 & \text{if } k \neq l, \end{cases} \tag{2.7}$$

and any $h \in L^2(S)$ can be written as $h = \sum_{k=1}^{\infty} \alpha_k \phi_k$, where $\alpha_k \in \mathbb{C}$, $k = 1, 2, \ldots$. Moreover, ϕ_k is continuous if $\lambda_k \neq 0$.

When we need to stress that the eigenvalue λ_k depends on S, we write λ_k^S. We also consider the following condition.

Condition 2. $\lambda_k^S \leqslant 1$ for all compact $S \subset \mathbb{R}^d$ and all k.

Theorem 1. Under condition 1, existence of $\mathrm{DPP}(C)$ is equivalent to condition 2.

Usually, for statistical models of covariance functions, condition 1 is satisfied, and so condition 2 becomes the essential condition. As discussed in Section 3.2, condition 2 simplifies in the stationary case of X. It seems difficult to provide intuition why the eigenvalues need to be bounded by 1, but they appear as probabilities for the simulation algorithm in Section 2.4.

When we are interested in considering a DPP only on a given compact set $S \subset \mathbb{R}^d$, then conditions 1 and 2 can be replaced by the assumption that C is a continuous complex covariance function defined on $S \times S$ such that $\lambda_k^S \leqslant 1$ for all k. The results in Sections 2.4 and 2.5 are then valid for this DPP, even if there is no continuous extension of C to $\mathbb{R}^d \times \mathbb{R}^d$ which satisfies conditions 1 and 2. However, it is convenient to assume conditions 1 and 2 as in Sections 3–5 we consider stationary DPPs.

Though a Poisson process is determinantal from definition 1, it is excluded by our approach where C is continuous. In particular, expression (2.6) does not hold for a Poisson process.

Assumption 1. In the remainder of this paper, $X \sim \mathrm{DPP}(C)$ with C satisfying conditions 1 and 2.

2.4. Simulation

We use the simulation algorithm of Hough *et al.* (2006) in the specific case where we want to

simulate $X_S \sim \mathrm{DPP}(C; S)$ with $S \subset \mathbb{R}^d$ compact. Our implementation of the algorithm becomes more efficient than that in Scardicchio *et al.* (2009), and our description is less technical than that in Hough *et al.* (2006). We consider the spectral representation (2.6) of C restricted to $S \times S$ and note the following definition.

Definition 2 Let $S \subset \mathbb{R}^d$ be compact and assume that all non-zero eigenvalues λ_k^S are 1. Then C restricted to $S \times S$ is called a projection kernel, and X_S is called a *determinantal projection point process*.

Theorem 2. For $k = 1, 2, \ldots$, let B_k be independent Bernoulli variables with mean λ_k. Define the random projection kernel $K : S \times S \to \mathbb{C}$ by

$$K(x, y) = \sum_{k=1}^{\infty} B_k \, \phi_k(x) \, \overline{\phi_k(y)}. \tag{2.8}$$

Then

$$\mathrm{DPP}_S(K) \sim \mathrm{DPP}(C; S). \tag{2.9}$$

Thus, if we first generate the independent Bernoulli variables, and second generate a determinantal projection point process on S with kernel K, the resulting point process follows $\mathrm{DPP}(C; S)$. For convenience, define $B_0 = \lambda_0 = 1$. As verified in the on-line supplementary material, if $N(S) = n(X_S)$ denotes the number of events in S, then

$$N(S) \sim \sum_{k=1}^{\infty} B_k, \qquad E\{N(S)\} = \sum_{k=1}^{\infty} \lambda_k, \qquad \mathrm{var}\{N(S)\} = \sum_{k=1}^{\infty} \lambda_k(1 - \lambda_k). \tag{2.10}$$

The on-line supplementary material exploits the first result in expression (2.10) for simulation of the Bernoulli variables, which is easily done, since $M = \max\{k \geqslant 0 : B_k \neq 0\}$ is finite and can be simulated by inversion, and $B_0, \ldots, B_{m-1}$ conditional on $M = m$ are still independent and identically distributed Bernoulli variables with parameters as before. We therefore concentrate on how to generate a realization from $\mathrm{DPP}_S(K)$ with K given by equation (2.8).

Let $n = \Sigma_{k=1}^{\infty} B_k$ denote the number of non-zero B_ks with $k \geqslant 1$ (as noted in expression (2.10), n is distributed as $N(S)$). If $n = 0$, then $K = 0$ and a realization from $\mathrm{DPP}_S(K)$ is simply equal to the empty point configuration. Assume that $n > 0$ and without loss of generality that $B_1 = \ldots = B_n = 1$. Let $\mathbf{v}(x) = (\phi_1(x), \ldots, \phi_n(x))^{\mathrm{T}}$, where 'T' and '$*$' respectively denote the transpose and conjugate transpose of a vector or a matrix.

Table 1. Algorithm 1: simulation of a determinantal projection point process

Sample X_n from the distribution with density $p_n(x) = \|\mathbf{v}(x)\|^2/n$, $x \in S$
Set $\mathbf{e}_1 = \mathbf{v}(X_n)/\|\mathbf{v}(X_n)\|$
For $i = n - 1$ to $i = 1$
  sample X_i from the distribution with density

$$p_i(x) = \frac{1}{i}\left\{ \|\mathbf{v}(x)\|^2 - \sum_{j=1}^{n-i} |\mathbf{e}_j^* \mathbf{v}(x)|^2 \right\}, \qquad x \in S, \tag{2.11}$$

  set $\mathbf{w}_i = \mathbf{v}(X_i) - \Sigma_{j=1}^{n-i} \{\mathbf{e}_j^* \mathbf{v}(X_i)\}\mathbf{e}_j$, $\mathbf{e}_{n-i+1} = \mathbf{w}_i/\|\mathbf{w}_i\|$
End for
Return $\{X_1, \ldots, X_n\}$

Theorem 3. If $n > 0$ and $K(x, y) = \sum_{k=1}^{n} \phi_k(x)\,\overline{\phi_k(y)}$ for all $x, y \in S$, then $\{X_1, \ldots, X_n\}$ generated by algorithm 1 (Table 1) is distributed as $\mathrm{DPP}_S(K)$.

To implement algorithm 1 we need to sample from the densities p_i, $i = n, \ldots, 1$ (with probability 1, $p_i(x)$ is a density, where we are conditioning on $(X_n, \ldots, X_{i+1})$ if $i < n$). This may simply be done by rejection sampling with a uniform instrumental density and acceptance probability given by $p_i(x)/U$, where U is an upper bound on $p_i(x)$ for $x \in S$. We use the bound $U = \sup_{y \in S} \|\mathbf{v}(y)\|^2 / i$ (see expression (2.11) in Table 1), which simplifies to $U = n/i$ for the Fourier basis that is considered in Section 4. In the on-line supplementary material we discuss rejection sampling for this and other choices of the instrumental distribution. Fig. 5 in Section 4.2 shows the acceptance probability for a uniformly distributed proposal at an intermediate step of algorithm 1.

2.5. Densities

This section briefly discusses the density expression for $X_S \sim \mathrm{DPP}(C; S)$ when $S \subset \mathbb{R}^d$ is compact. Recall that the eigenvalues $\lambda_k = \lambda_k^S$ are assumed to be less than or equal to 1. In general, when some eigenvalues λ_k are allowed to be 1, the density of X_S is not available, but we can condition on the Bernoulli variables B_k from theorem 2 to obtain the conditional density; see the on-line supplementary material. The most interesting case occurs when $\lambda_k < 1$ for all $k = 1, 2, \ldots$, which means that no B_k is almost surely 1. Then the density of X_S exists and is specified in theorem 4 below, where the following considerations and notation are used. If $P\{N(S) = n\} > 0$, then $P\{N(S) = m\} > 0$ for $m = 0, \ldots, n$; see expression (2.10). Thus $P\{N(S) = 0\} = \Pi_{k=1}^{\infty}(1 - \lambda_k)$ is strictly positive, and we can define

$$D = -\log[P\{N(S) = 0\}] = -\sum_{k=1}^{\infty} \log(1 - \lambda_k), \tag{2.12}$$

$$\tilde{C}(x, y) = \sum_{k=1}^{\infty} \tilde{\lambda}_k\, \phi_k(x)\, \overline{\phi_k(y)}, \qquad x, y \in S, \tag{2.13}$$

where

$$\tilde{\lambda}_k = \lambda_k / (1 - \lambda_k), \qquad k = 1, 2, \ldots.$$

Let $|S| = \int_S \mathrm{d}x$, and set $\det\{[\tilde{C}](x_1, \ldots, x_n)\} = 1$ if $n = 0$.

Theorem 4. Assuming $\lambda_k < 1$, $k = 1, 2, \ldots$, then X_S is absolutely continuous with respect to the homogeneous Poisson process on S with unit intensity and has density

$$f(\{x_1, \ldots, x_n\}) = \exp(|S| - D)\, \det\{[\tilde{C}](x_1, \ldots, x_n)\} \tag{2.14}$$

for all $(x_1, \ldots, x_n) \in S^n$ and $n = 0, 1, \ldots$.

Section 4 and the on-line supplementary material discuss efficient ways of approximating $\tilde{C}$ and D when X is stationary.

3. Stationary models

To the best of our knowledge, parametric families of DPP models have not yet been studied in the literature from a statistical perspective. In what follows we focus on the stationary case of DPPs, discuss isotropy (Section 3.1), give a simple condition for the existence of a stationary DPP (Section 3.2) and construct various classes of parametric models (Sections 3.3 and 3.4). Inhomogeneous models of DPPs are discussed in Section 6.

Throughout this section, $X \sim \mathrm{DPP}(C)$ where C is of the form

$$C(x, y) = C_0(x - y), \qquad x, y \in \mathbb{R}^d. \tag{3.1}$$

This condition implies that X is stationary, i.e. its distribution is invariant under translations. If C is real, expression (3.1) is equivalent to the stationarity of X.

We also refer to C_0 as a covariance function. $C_0(0)$, the variance corresponding to C, equals ρ, the intensity of X; see expression (2.2).

3.1. Isotropy

It is often convenient to require that C_0 is isotropic, meaning that $C_0(x) = \rho R_0(\|x\|)$ is invariant under rotations about the origin in $\mathbb{R}^d$. This is a natural simplification, since any stationary anisotropic covariance function can be obtained from some stationary isotropic covariance function by rotation and rescaling; see for example Goovaerts (1997).

Suppose that C_0 is isotropic. Then C_0 is real, and the pair correlation function depends only on the distance between pairs of points, $g(x, y) = g_0(\|x - y\|)$; see expression (2.3). Hence commonly used statistical procedures based on the pair correlation function or the closely related K-function apply (see Ripley (1976, 1977) and Møller and Waagepetersen (2004)). In particular, using the relationship

$$|R_0(r)| = \sqrt{\{1 - g_0(r)\}} \tag{3.2}$$

we can define a 'range of correlation', i.e. a distance $r_0 > 0$ such that $g_0(r) \approx 1$ for $r \geqslant r_0$, as exemplified later in expression (3.6). For many specific models for isotropic covariance, including those studied in Section 3.3, R_0 is a decreasing function. By equation (3.2) g_0 is then an increasing function from 0 to 1.

Examples of stationary isotropic covariance functions are studied in Sections 3.3 and 3.4. However, the following Section 3.2 does not involve an assumption of isotropy, and the approximation of C_0 that is studied in Section 4 is only approximately isotropic when C_0 is isotropic.

3.2. A simple spectral condition for existence

Proposition 1 below simplifies the existence condition 2. For this result we need to recall the notion of a spectral density: for any function $h \in L^2(\mathbb{R}^d)$, we define the Fourier transform $\mathcal{F}(h)$ of h by

$$\mathcal{F}(h)(x) = \int h(y) \exp(-2\pi \mathrm{i} x y)\, \mathrm{d}y, \qquad x \in \mathbb{R}^d,$$

and the inverse Fourier transform $\mathcal{F}^{-1}(h)$ of h by

$$\mathcal{F}^{-1}(h)(x) = \int h(y) \exp(2\pi \mathrm{i} x y)\, \mathrm{d}y, \qquad x \in \mathbb{R}^d.$$

We refer to Stein and Weiss (1971) for more details about the definition and properties of the Fourier transform in $L^2(\mathbb{R}^d)$. Now, if C_0 is a continuous covariance function (as assumed in this paper) and belongs to $L^2(\mathbb{R}^d)$, then $\varphi = \mathcal{F}(C_0)$ is called the spectral density for C_0. By Khinchin's (or Bochner's) theorem, $\varphi \geqslant 0$; see Yaglom (1987).

Proposition 1. Under condition 1 and expression (3.1), if $C_0 \in L^2(\mathbb{R}^d)$, then condition 2 is equivalent to that

$$\varphi \leqslant 1. \tag{3.3}$$

Assumption 2. Henceforth, in addition to condition 1, we assume that $C_0 \in L^2(\mathbb{R}^d)$ and that condition (3.3) holds.

The following corollary becomes useful in Section 3.4 where we discuss a spectral approach for constructing stationary DPPs.

Corollary 1. Under expression (3.1) the following two statements are equivalent.

(a) There exists $\varphi \in L^1(\mathbb{R}^d)$ with $0 \leqslant \varphi \leqslant 1$ and $C_0 = \mathcal{F}^{-1}(\varphi)$.
(b) Conditions 1 and 2 hold and $C_0 \in L^2(\mathbb{R}^d)$.

Remark 2. There is a trade-off between how large the intensity and how repulsive a stationary DPP can be: consider a parametric model for C_0 with parameters ρ and θ. For each fixed value of θ, condition 2 is equivalent to $0 \leqslant \rho \leqslant \rho_{\max}$ where $\rho_{\max} = \rho_{\max}(\theta)$ may depend on θ and is determined by condition (3.3). As exemplified in Section 3.3, $\rho_{\max}$ will be a decreasing function of the range of correlation (which depends only on θ). However, it may be more natural to determine the range of θ in terms of ρ.

3.3. Examples of covariance function models

Numerous examples of stationary and isotropic covariance functions exist (see for example Gelfand *et al.* (2010)), and examples of stationary and anisotropic covariance functions are discussed in De laco *et al.* (2003).

Examples of isotropic covariance functions $C_0(x)$, where the range

$$\delta = \sup\{\|x\| : C_0(x) \neq 0\} \tag{3.4}$$

is finite, are given in Wu (1995) and Gneiting (2002). Such examples can be of particular interest because X_A and X_B are independent DPPs whenever $A, B \subset \mathbb{R}^d$ are separated by a distance larger than δ; see definition 1. However, in what follows, we focus on some classes of stationary isotropic covariance functions with infinite range.

Table 2 specifies C_0 and φ for three important isotropic cases: a Gaussian, Whittle–Matérn or generalized Cauchy covariance function—we refer to a DPP model with these kernels as the Gaussian, Whittle–Matérn or Cauchy model respectively. Each model depends on the intensity parameter $\rho = C_0(0) \geqslant 0$, on a scale parameter $\alpha > 0$ and, for the two latter models, on a shape parameter $\nu > 0$. Note that the Gaussian covariance function with $\alpha = 1/\sqrt{(\pi\rho)}$ is the limit as $\nu \to \infty$ of both the Whittle–Matérn covariance function with $\alpha = 1/\sqrt{(4\pi\nu\rho)}$ and the Cauchy covariance function with $\alpha = \sqrt{\{\nu/(\pi\rho)\}}$. For all three models, the spectral density has maximal value $\varphi(0)$, so by condition (3.3) X exists if and only if $\varphi(0) \leqslant 1$. This means that $\rho_{\max} = M\alpha^{-d}$, where M is given in Table 2. Similarly, given $\rho > 0$ and $\nu > 0$, we denote $\alpha_{\max} = (M/\rho)^{1/d}$. Furthermore, Table 2 gives the expression of the isotropic pair correlation g_0 and, for $d = 2$, Ripley's K-function (Ripley, 1976, 1977) given by

$$K(r) = 2\pi \int_0^r t\, g_0(t)\, \mathrm{d}t, \qquad r \geqslant 0. \tag{3.5}$$

For the Whittle–Matérn model, the integral in expression (3.5) must be evaluated by numerical methods.

In what follows, let $d = 2$. For the Gaussian, Whittle–Matérn and generalized Cauchy models, we use the range of correlation

Table 2. Covariance function and other quantities for the Gaussian, Whittle–Matérn and generalized Cauchy models, where K_ν denotes the modified Bessel function of the second kind, for the K-function, $d = 2$

	Functions for the following models:		
	Gaussian	*Whittle–Matérn*	*Generalized Cauchy*
$C_0(x)$	$\rho \exp(-\|x/\alpha\|^2)$	$\rho \dfrac{2^{1-\nu}}{\Gamma(\nu)} \|x/\alpha\|^\nu K_\nu(\|x/\alpha\|)$	$\rho(1+\|x/\alpha\|^2)^{-\nu-d/2}$
$\varphi(x)$	$\rho(\sqrt{\pi}\alpha)^d \exp(-\|\pi\alpha x\|^2)$	$\rho \dfrac{\Gamma(\nu+d/2)}{\Gamma(\nu)} \dfrac{(2\sqrt{\pi}\alpha)^d}{(1+\|2\pi\alpha x\|^2)^{\nu+d/2}}$	$\rho \dfrac{2^{1-\nu}(\sqrt{\pi}\alpha)^d}{\Gamma(\nu+d/2)} \|2\pi\alpha x\|^\nu K_\nu(\|2\pi\alpha x\|)$
M	$\pi^{-d/2}$	$\dfrac{\Gamma(\nu)}{2^d \pi^{d/2} \Gamma(\nu+d/2)}$	$\dfrac{\Gamma(\nu+d/2)}{\Gamma(\nu)\pi^{d/2}}$
$g_0(r)$	$1-\exp\{-2(r/\alpha)^2\}$	$1-\left\{\dfrac{2^{1-\nu}(r/\alpha)^\nu K_\nu(r/\alpha)}{\Gamma(\nu)}\right\}^2$	$1-\{1+(r/\alpha)^2\}^{-2\nu-d}$
$K(r)$	$\pi r^2 - \dfrac{\pi\alpha^2}{2}\left\{1-\exp\left(\dfrac{-2r^2}{\alpha^2}\right)\right\}$	Numerical quadrature	$\pi r^2 - \dfrac{\pi\alpha^2}{2\nu+1}\left\{1-\left(\dfrac{\alpha^2}{\alpha^2+r^2}\right)^{2\nu+1}\right\}$

$$\left.\begin{aligned} r_0 &= \alpha\sqrt{\{-\log(0.1)\}}, \\ r_0 &= \alpha\sqrt{(8\nu)}, \\ r_0 &= \alpha\sqrt{(0.1^{-1/(\nu+1)}-1)} \end{aligned}\right\} \tag{3.6}$$

respectively. For the Whittle–Matérn model, this is based on an empirical result of Lindgren *et al.* (2011), whereas, for the other two models, $|R_0(r_0)| = 0.1$, i.e. $g_0(r_0) = 0.99$. In each case, r_0 depends linearly on α and, when ν is fixed, $\rho_{\max}$ is proportional to r_0^{-d}. Figs 3(a) and 3(b) show examples of the isotropic pair correlation functions with a fixed range of correlation r_0. In particular the Whittle–Matérn DPPs have several different shapes of pair correlation functions and so they may constitute quite a flexible model class for repulsiveness. It is also evident that the value of $\rho_{\max}$ is of the same order of magnitude for all these models, indicating that the range of interaction has a major effect on the maximal permissible intensity of the model. Moreover, Fig. 3(c) shows $L(r) - r$ for seven DPPs, where $L(r) = \sqrt{\{K(r)/\pi\}}$.

3.4. Spectral approach

As an alternative to specifying a stationary covariance function C_0, involving the need for checking positive semidefiniteness and for controlling the upper bound of its Fourier transform, we may simply specify an integrable function $\varphi : \mathbb{R}^d \to [0, 1]$, which becomes the spectral density; see corollary 1. In fact knowledge about φ is all that we need for the approximate simulation procedure and density approximation in Section 4. However, the disadvantage is that it may then be difficult to determine $C_0 = \mathcal{F}^{-1}(\varphi)$, and hence closed form expressions for g and K may not be available. Furthermore, it may be more difficult to interpret parameters in the spectral domain.

In what follows we propose a model class displaying a higher degree of repulsiveness than the Gaussian model which appears as a special case. For $\rho \geqslant 0$, $\alpha > 0$ and $\nu > 0$, define

$$\varphi(x) = \rho \frac{\Gamma(d/2+1)\alpha^d}{\pi^{d/2}\Gamma(d/\nu+1)} \exp(-\|\alpha x\|^\nu). \tag{3.7}$$

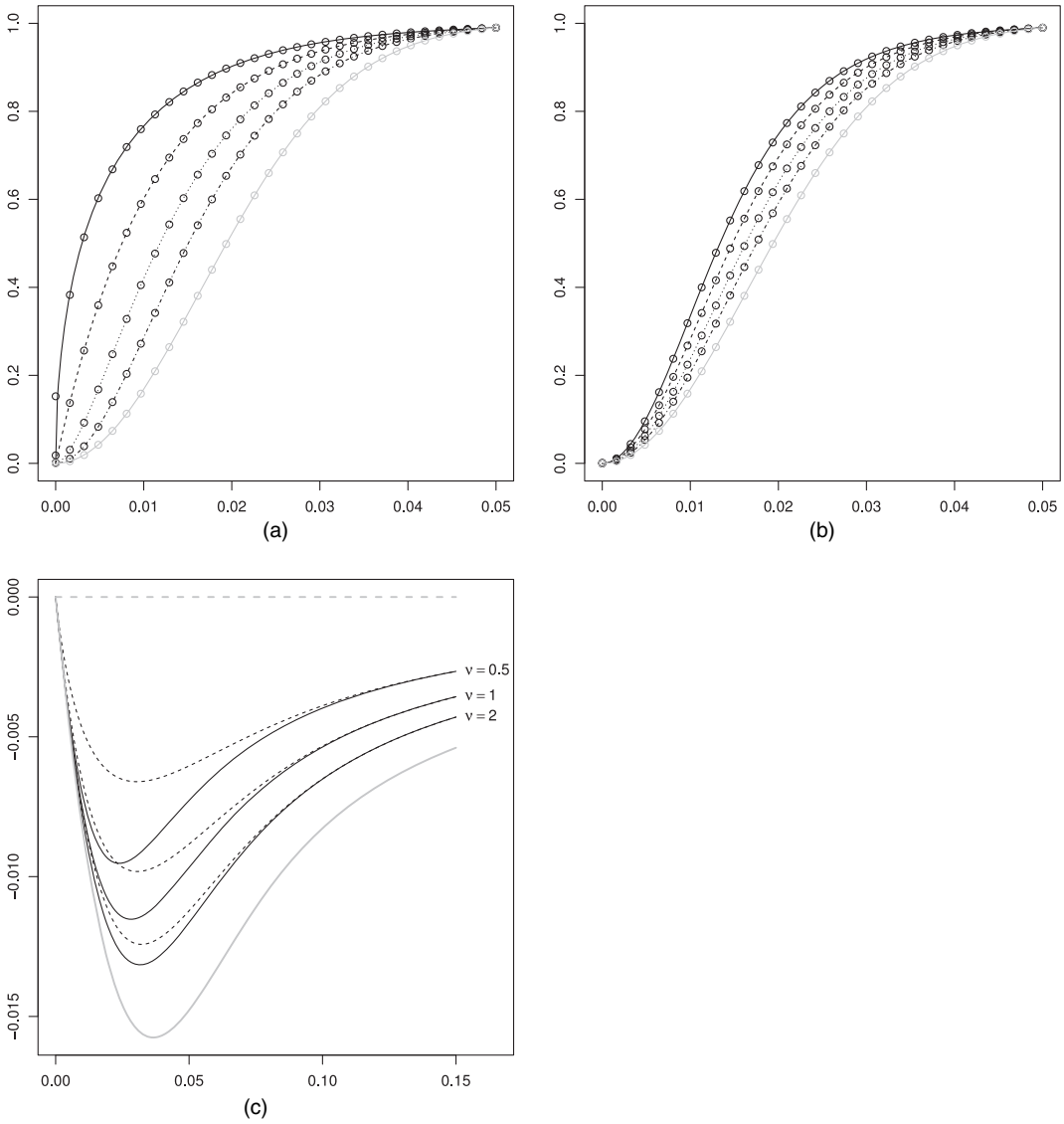

Fig. 3. Isotropic pair correlation functions for (a) the Whittle–Matérn model and (b) the Cauchy model for various values of ν (———, $\nu = 0.25$, $\rho_{\max} = 324$; - - - - - -, $\nu = 0.50$, $\rho_{\max} = 337$; · · · · · ·, $\nu = 1.00$, $\rho_{\max} = 329$; - · - · - ·, $\nu = 2.00$, ———, $\rho_{\max} = 315$; ———, $\nu = \infty$ (Gaussian model), $\rho_{\max} = 293$; ○, values of the approximate isotropic pair correlation function (obtained by using the approximation C_{app} described in Section 4) and where α is chosen such that $r_0 = 0.05$, and (c) plots of $L(r) - r$ *versus* r for the Whittle–Matérn (- - - - - -), Cauchy (———) and Gaussian (———) models with $\alpha = \alpha_{\max}$ and $\rho = 100$ (— — —, $L(r) - r$ for a stationary Poisson process); for the Whittle–Matérn and Cauchy models, $\nu \in \{0.5, 1, 2\}$

We call a DPP model with a spectral density of the form (3.7) a power exponential spectral model. For $\nu = 2$, this is the Gaussian model of Section 3.3. It is easily checked that $\rho = \int \varphi = C_0(0)$, so ρ corresponds to the intensity of the model. Moreover, for fixed ρ and ν, the existence condition (3.3) holds if $\alpha \leqslant \alpha_{\max}$ where $\alpha_{\max}^d = \Gamma(d/\nu + 1)\tau^{-d}$ and $\tau^d = \rho\,\Gamma(d/2 + 1)/\pi^{d/2}$. For the choice $\alpha = \alpha_{\max}$ in equation (3.7), the spectral density of the power exponential spectral model becomes

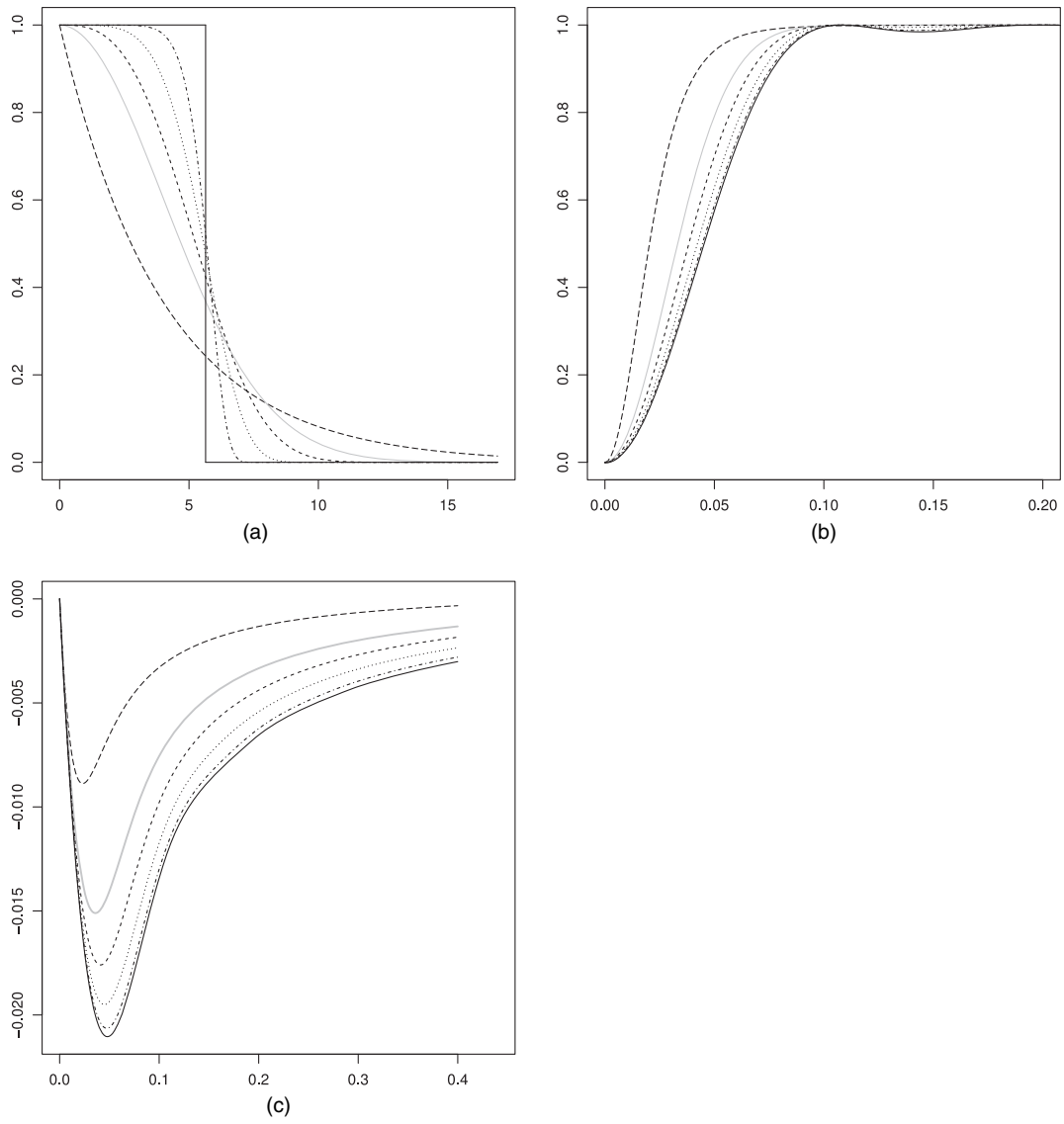

Fig. 4. (a) Isotropic spectral densities, (b) approximate isotropic pair correlation functions and (c) approximate $L(r) - r$ functions for power exponential spectral models with $\rho = 100$, $\nu = 1$(– – –), 2 (———, Gaussian model), 3 (------), 5 (· · · · · ·), 10 (·-·-·-), ∞ (———), and $\alpha = \alpha_{\max}$

$$\varphi(x) = \exp\{-\|\Gamma(d/\nu + 1)^{1/d} x/\tau\|^{\nu}\}. \tag{3.8}$$

This function tends to the indicator function over the set $\{\|x\| \leqslant \tau\}$ as $\nu \to \infty$. This limiting case corresponds in a sense to the most repulsive stationary DPP (see appendix J in the on-line supplementary material). For $d = 2$, as $\nu \to \infty$ we obtain in the limit

$$C_0(x) = \sqrt{(\rho/\pi)}\, J_1\{2\sqrt{(\pi\rho)}\|x\|\}/\|x\| \tag{3.9}$$

which is proportional to a 'jinc-like' function.

Fig. 4 illustrates some properties of the power exponential spectral model when $\alpha = \alpha_{\max}$ and

$\nu = 1, 2, 3, 5, 10, \infty$. Fig. 4(a) shows the spectral density for these cases. Since we are not aware of a closed form expression for $C_0 = \mathcal{F}^{-1}(\varphi)$ when φ is given by equation (3.8), we approximate C_0 by the method discussed in Section 4, leading to approximate pair correlation functions shown in Fig. 4(b). Fig. 4(c) shows the corresponding approximations of $L(r) - r$ (analogously to Fig. 3(c) in Section 3.3).

4. Approximations

Let again $X \sim \mathrm{DPP}(C)$ where $C(x, y) = C_0(x - y)$ (see expression (3.1)), so that X is stationary. By remark 1 in Section 2 the restriction of X to a bounded set S is the finite DPP $X_S \sim \mathrm{DPP}_S(C)$. To simulate and evaluate the density for such a process the spectral representation (2.6) is needed. Unfortunately, analytic expressions for such representations are known in only a few simple cases (see for example Macchi (1975)), which we believe are insufficient to describe the interaction structure in real data sets. In what follows we propose various approximations which are easy to apply for any C_0 with a known spectral density. Throughout this section we consider $S = [-\frac{1}{2}, \frac{1}{2}]^d$, but the methods easily generalize to any rectangular set owing to the transformation property (2.4) (see the on-line supplementary material for more details).

4.1. Approximation of the kernel C

Denoting as before φ the Fourier transform of C_0, we consider $X^{\mathrm{app}} \sim \mathrm{DPP}_S(C_{\mathrm{app}})$ where

$$C_{\mathrm{app}}(x, y) = \sum_{k \in \mathbb{Z}^d} \varphi(k) \exp\{2\pi \mathrm{i} k(x - y)\}, \qquad x, y \in S. \tag{4.1}$$

This is a spectral decomposition of C_{app} on S, as defined in expression (2.6), and X^{app} is well defined since $\varphi \leqslant 1$.

As justified in what follows, for any $x, y \in S$,

$$C(x, y) \approx C_{\mathrm{app}}(x, y) \qquad \text{if } x - y \in S. \tag{4.2}$$

Indeed, if $x - y \in S$, the Fourier expansion of C_0 on S yields

$$C(x, y) = C_0(x - y) = \sum_{k \in \mathbb{Z}^d} \alpha_k \exp\{2\pi \mathrm{i} k(x - y)\}$$

where

$$\alpha_k = \int_S C_0(t) \exp\{-2\pi \mathrm{i} k t\} \, \mathrm{d}t. \tag{4.3}$$

Substituting S by $\mathbb{R}^d$ in this integral, we obtain $\varphi(k)$. Therefore, if $C_0(t) \approx 0$ for $t \in \mathbb{R}^d \setminus S$, then $\alpha_k \approx \varphi(k)$ and consequently $C(x, y) \approx C_{\mathrm{app}}(x, y)$ when $x - y \in S$. This is so for reasonable parameter values of the models, i.e. when the expected number of points in X_S is not very low. For instance, Fig. 3 indicates that the approximation is accurate as the approximate pair correlation functions marked by circles in the plot are very close to the true curves. The quality of the approximation is discussed in more detail in the on-line supplementary material.

If $x, y \in S$ and $x - y \notin S$, then $C_{\mathrm{app}}(x, y)$ is no longer an approximation of $C(x, y)$, but rather an approximation of its periodic extension given by $\Sigma_{k \in \mathbb{Z}^d} \alpha_k \exp\{2\pi \mathrm{i} k(x - y)\}$ for all $x, y \in S$.

4.2. Approximate simulation

It is straightforward to simulate X^{app}, since expression (4.1) is of the form required for the

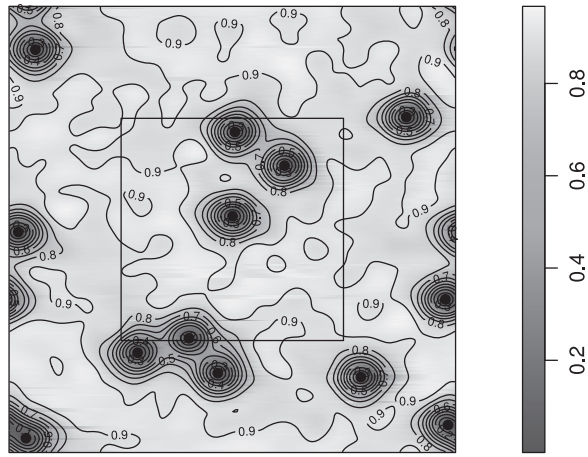

Fig. 5. Acceptance probability for a uniformly distributed proposal at an intermediate step of algorithm 1 (Section 2.4) when simulating a realization of X^{app} on $S = [-\frac{1}{2}, \frac{1}{2}]$ (the interior box is the region $S/2$); ●, previously generated points (the acceptance probability is 0 at these points)

simulation algorithm of Section 2.4. Fig. 5 shows the acceptance probability for a uniformly distributed proposal (used for rejection sampling when simulating from one of the densities p_i) when X^{app} is simulated. The interior region in Fig. 5 is $S/2 = [-\frac{1}{4}, \frac{1}{4}]^d$. If $x, y \in S/2$, then $x - y \in S$ and on the basis of the arguments above we expect that $X^{\text{app}}_{S/2}$ is a good approximation of $X \cap S/2$. In practice it turns out that the approximation works well for the entire region S. This may intuitively be explained by Fig. 5 where the qualitative behaviours of X^{app} and $X^{\text{app}}_{S/2}$ are similar for example in the sense that there are regions at the borders where the acceptance probability is low. For $X^{\text{app}}_{S/2}$, this is due to the influence of points outside $S/2$. For X^{app}, this is created artificially by points at the opposite border. Therefore, we use X^{app} as an approximate simulation of X_S. We refer to this approach as the periodic method of simulation. The online supplementary material studies the alternative border method of simulation which retains only the points inside $S/2$. Although this approximation is mathematically well founded it is computationally more expensive (it needs to simulate 2^d times more points on average). In our experience, the periodic method works equally well in practice.

4.3. Approximation of the density

Consider the density f for X_S as specified in theorem 4 (so we assume that $\varphi < 1$). We may use the approximation $f \approx f^{\text{app}}$, where

$$f^{\text{app}}(\{x_1, \ldots, x_n\}) = \exp(|S| - D_{\text{app}}) \det\{[\tilde{C}_{\text{app}}](x_1, \ldots, x_n)\}, \qquad \{x_1, \ldots, x_n\} \subset S, \tag{4.4}$$

denotes the density of X^{app}, with

$$\tilde{\varphi}(u) = \varphi(u)/\{1 - \varphi(u)\}, \qquad u \in S, \tag{4.5}$$

$$\tilde{C}_{\text{app}}(x, y) = \tilde{C}_{\text{app},0}(x - y) = \sum_{k \in \mathbb{Z}^d} \tilde{\varphi}(k) \exp\{2\pi i k(x - y)\}, \qquad x, y \in S, \tag{4.6}$$

$$D_{\text{app}} = \sum_{k \in \mathbb{Z}^d} \log\{1 + \tilde{\varphi}(k)\}. \tag{4.7}$$

We call f^{app} the periodic approximation of f. To evaluate this approximation in practice we

need to use truncated versions of equations (4.6) and (4.7). For a given integer $N>0$ (the choice of N is discussed in Section 5.1), let $\mathbb{Z}_N=\{-N,-N+1,\ldots,N-1,N\}$ and define

$$D_N=\sum_{k\in\mathbb{Z}_N^d}\log\{1+\tilde{\varphi}(k)\} \tag{4.8}$$

and

$$\tilde{C}_N(u)=\sum_{k\in\mathbb{Z}_N^d}\tilde{\varphi}(k)\exp(2\pi \mathrm{i}ku), \qquad u\in\mathbb{R}^d. \tag{4.9}$$

Whereas it is feasible to evaluate equation (4.8) for large values of N, the evaluation of equation (4.9) is more problematic since it needs to be carried out for every distinct pair of points in $\{x_1,\ldots,x_n\}$. For moderate N (a few hundreds) direct calculation of equation (4.9) can be used. In this case, we can exploit the fact that $\tilde{\varphi}$ often is an even function (corresponding to a real-valued C_0) such that all imaginary parts in equation (4.9) cancel. This allows us to reduce the number of terms in the sum by a factor 2^d, which speeds up calculations considerably when evaluating the approximate density. For large N (hundreds or thousands) we use the fast Fourier transform of $\tilde{\varphi}$. The fast Fourier transform yields values of $\tilde{C}_N$ at a discrete grid of values and we approximate $\tilde{C}_N(x_i-x_j)$ by bilinear interpolation. A simulation study reported in the on-line supplementary material shows that likelihood inference based on this truncated version of f^{app} works well in practice.

5. Inference for stationary models

In this section, we discuss and exemplify how to estimate parameters of stationary DPP models by using the approximate likelihood and how to do model comparison. Section 5.1 briefly describes how to use the density of Section 4.3 for approximate likelihood inference in practice whereas Sections 5.2 and 5.3 describe the statistical analysis of two data sets.

5.1. Maximum-likelihood-based inference

Assume that $\{x_1,\ldots,x_n\}$ is a realization of $X\sim\mathrm{DPP}(C)$ restricted to a compact set S, where $C(x,y)=C_0(x-y)=\rho R_0(\|x-y\|)$ is modelled by one of the parametric models of Sections 3.3 and 3.4, namely the Gaussian, Whittle–Matérn, Cauchy and power exponential spectral models—for short we refer to these as the four parametric models of DPPs. Recall that ρ is the intensity parameter, θ denotes the parameter of the correlation function R_0 and a given value of ρ introduces a bound on the parameter space for θ which is denoted Θ_ρ. In all that follows we shall replace ρ by $\hat{\rho}=n/|S|$; as discussed in the on-line supplementary material this works very well in practice. Further, we consider only the (approximate) maximum likelihood estimate for θ as the value that maximizes the approximate log-likelihood, i.e. the truncated version of the density f^{app} in expression (4.4),

$$l_N(\theta)=\log[\det\{[\tilde{C}_N](x_1,\ldots,x_n)\}]-D_N, \qquad \theta\in\Theta_{\hat{\rho}},$$

where $[\tilde{C}_N](x_1,\ldots,x_n)$ is the $n\times n$ matrix with (i,j)th element $\tilde{C}_N(x_i-x_j)$. Here we suppress in the notation that $\tilde{C}_N$ and D_N depend on (ρ,θ) through $\tilde{\varphi}$; see expressions (4.8) and (4.9).

Concerning the choice of N, note that the sum

$$S_N=\sum_{k\in\mathbb{Z}_N^d}\varphi(k)$$

tends to ρ from below as $N \to \infty$. Hence, for any value of θ, one criterion for choosing N is to require for example $S_N > 0.99\hat{\rho}$. However, this may be insufficient as N also determines the grid resolution when a fast Fourier transform is used, and a high resolution may be required to obtain a good approximation of the likelihood. Therefore, in the fast Fourier transform case we use increasing values of N until the approximate maximum likelihood estimate stabilizes.

When comparing the four parametric models fitted to the same data set, we prefer the model with the largest value of $l_N(\theta)$. The comparison of $l_N(\theta)$ between different models is valid, since the dominating measure is the same.

5.2. Spanish towns data set

The Spanish towns data set (Fig. 2(a)) was first analysed in Glass and Tobler (1971). In a subsequent analysis, Ripley (1988) used a Strauss hard-core model specified by four parameters: a hard-core distance h, an interaction distance R, an abundance parameter β and an interaction parameter γ. The maximum likelihood estimate for h is the minimal observed distance $r_{\min}$, but more commonly $nr_{\min}/(n+1)$ is used. Estimation of R is typically based on an *ad hoc* method such as maximum profile pseudolikelihood (see for example Møller and Waagepetersen (2004)). Conditionally on h and R, the parameters β and γ can for example be estimated by the maximum pseudolikelihood method or much more computationally demanding Markov chain Monte Carlo methods to approximate the likelihood can be used. Following an analysis in Illian *et al.* (2008), we use their estimates $\hat{h} = 0.83$ and $\hat{R} = 3.5$. Further, we estimate β and γ by using the approximate likelihood method of Huang and Ogata (1999) that is available in `spatstat`. The estimates are $\hat{\beta} = 0.13$ and $\hat{\gamma} = 0.48$.

Fig. 6 is used to assess the goodness of fit for the Strauss hard-core model. The broken central curves show non-parametric estimates of $L(r) - r$, the nearest neighbour distribution function $G(r)$ and the empty space function $F(r)$ (for definitions of F and G, see for example Møller and Waagepetersen (2004)). Figs 6(a), 6(c) and 6(e) also show 2.5% and 97.5% pointwise quantiles (the grey curves) for these summary statistics based on 4000 simulations of the fitted Strauss hard-core model. Overall the model appears to provide an acceptable fit, but the characteristic cusp of the envelopes of $L(r) - r$ at $r = \hat{R} = 3.5$ seems to be a somewhat artificial model effect that the data set does not exhibit (see also example 3.14 in Illian *et al.* (2008)). Figs 6(b), 6(d) and 6(f) show the rank envelopes of Myllymäki *et al.* (2013) using their R package `spptest` with a significance level of 5%. In contrast with the pointwise envelopes these are global envelopes and if the non-parametric estimate exits the envelopes the deviation is significant at the 5% level. The rank envelope test also yields an interval for the p-value, and the p-value intervals for all three summary statistics are above 5% (which corresponds to the non-parametric estimates staying within the envelopes).

As an alternative to the Strauss hard-core model, we now consider the four parametric classes of DPP models. The intensity estimate is $\hat{\rho} = 0.043$, and the fitted Whittle–Matérn model (with $\hat{\nu} = 2.7$ and $\hat{\alpha} = 0.819$) has the highest value of the likelihood and it is therefore preferred over the Cauchy and power exponential spectral models which also have three parameters. As the Gaussian model has only two parameters and is a (limiting) special case of the Whittle–Matérn model, we carry out a simulation-based likelihood ratio test as follows. Using simulated realizations under the fitted Gaussian model (with $\hat{\alpha} = 2.7$), we fit both the Gaussian model and the (alternative) Whittle–Matérn model. Then, for each sample, we evaluate $D = -2\log(Q)$ where Q is the ratio of the Gaussian and the Whittle–Matérn likelihoods. We finally compare the distribution of D over the 400 simulated realizations with the observed value of D for the data set. The resulting p-value is 0.03 and we reject the Gaussian model in favour of the Whittle–Matérn model. Again Fig. 6 is used to assess the goodness of fit, which appears to be quite good

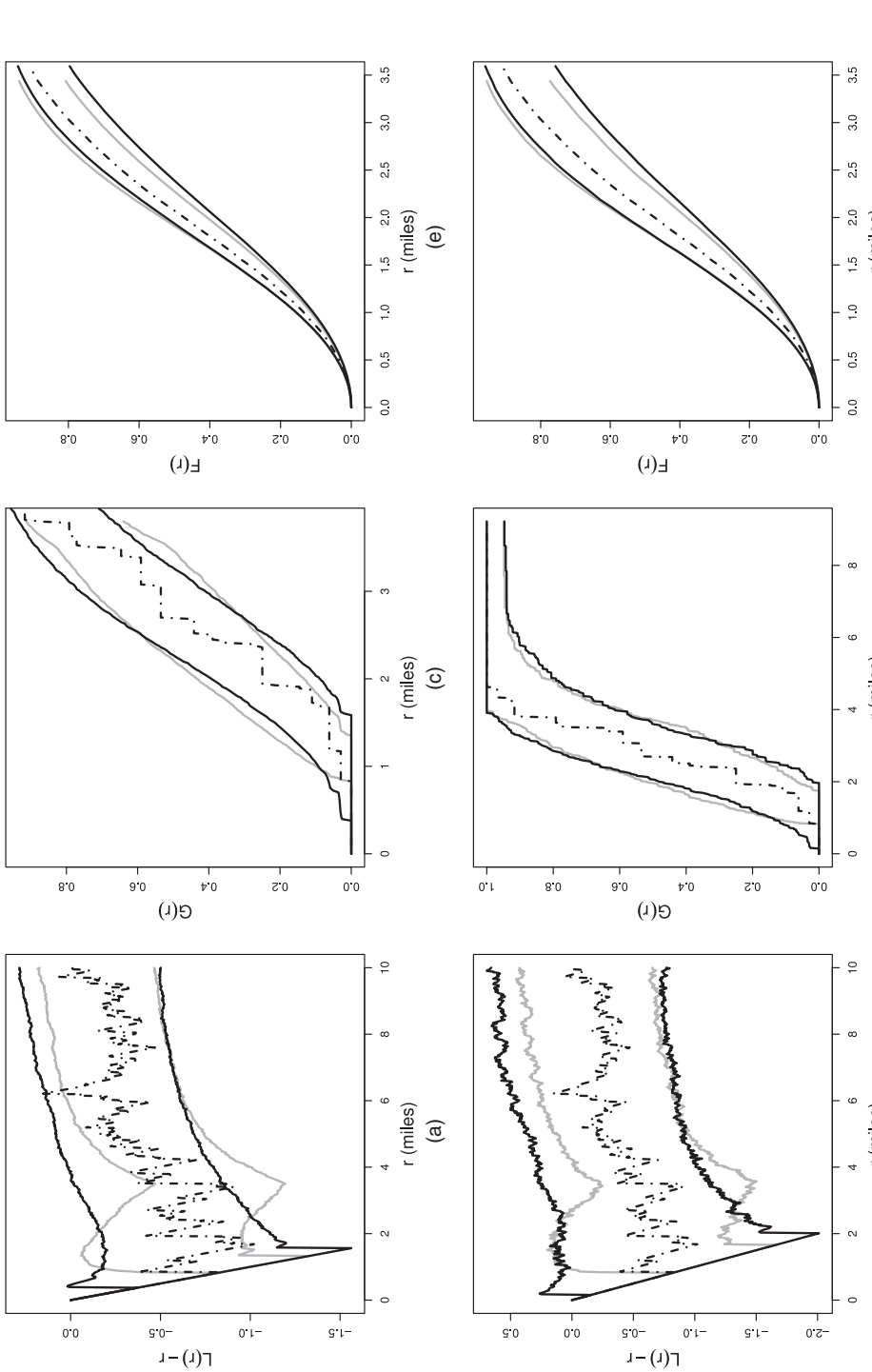

Fig. 6. Non-parametric estimates of (a), (b) $L(r) - r$, (c), (d) $G(r)$ and (e), (f) $F(r)$ for the Spanish towns data set with simulation-based envelopes for both the fitted Strauss hard-core model (———) and the fitted Whittle–Matérn model (— — —) (·····, data) (for both models the envelopes are based on 4000 simulated realizations): for (a), (c), (e) the envelopes are 2.5% and 97.5% pointwise quantiles; for (b), (d) and (f) the envelopes are rank envelopes with a 5% level of significance

as none of the non-parametric estimates exit either the 95% pointwise envelopes or the 5% rank envelopes obtained from simulations under the fitted Whittle–Matérn model.

In summary the Whittle–Matérn model both has fewer parameters and arguably provides a better fit than the Strauss hard-core model. Furthermore, we have direct access to the moments of the Whittle–Matérn model such as the intensity and the pair correlation function which can only be obtained by simulation for the Strauss hard-core model.

5.3. Hamster cells data set

Fig. 2(b) shows a plot of the locations of 303 cells of two types in a $0.25\,\text{mm} \times 0.25\,\text{mm}$ region of a histological section of the kidney of a hamster. The data have been rescaled to a unit square and there are 226 dividing (living) cells marked by circles and 77 pyknotic (dying) cells marked by plus symbols. The data set was analysed in Diggle (2003), section 6.4.1, where it was concluded to be in agreement with independent labelling of a simple sequential inhibition process with hard-core distance $\delta = 0.0012$. As noted by Diggle, this model is not strictly valid, since there are a few pairs of data points violating the hard-core condition. However, Diggle considered the good overall fit as an indication that the simple sequential inhibition model, together with random labelling of cell types, provides a reasonable approximate description of the data.

Under the assumption of random labelling, the two subpoint patterns consisting of respectively the dividing and the pyknotic cells correspond to respectively the retained and thinned points of an independent thinning of the full unmarked point pattern. Using a stationary DPP model with kernel C_0 for the full unmarked point pattern, this implies that each individual subpoint pattern should follow the same type of DPP model with different values of ρ_1 and ρ_2 for the intensities (with $\rho = \rho_1 + \rho_2$), and kernels $(\rho_1/\rho)C_0$ and $(\rho_2/\rho)C_0$; see the thinning result (2.5). We may exploit this to test the hypothesis of random labelling: first we fit a parametric class of DPP models to the full unmarked point pattern; second we fit the same model class to each of the subpoint patterns. If random labelling is true, all fitted models should coincide, up to the intensity.

For the full unmarked point pattern, all four of the parametric classes of DPPs fit well (judged by envelopes of summary statistics that are not shown here) with very similar values of the approximate likelihood. Simulation-based likelihood ratio tests cannot reject the null hypothesis of the Gaussian model against alternatives given by either the Whittle–Matérn, Cauchy or power exponential spectral model. We therefore use the simpler fitted Gaussian model with estimates $(\hat{\rho}, \hat{\alpha}) = (303, 0.0181)$ for the full unmarked point pattern, $(\hat{\rho}_1, \hat{\alpha}_1) = (226, 0.0188)$ for the dividing cells and $(\hat{\rho}_2, \hat{\alpha}_2) = (77, 0.00816)$ for the pyknotic cells. The relevant hypotheses to check for random labelling, based on the thinning characterization that was explained above, is thus H_0: $\alpha_1 = \alpha_2 = \alpha$ against H_1: $\alpha_1 \neq \alpha$ or $\alpha_2 \neq \alpha$. Several test statistics can be proposed to perform this test. We choose to base our decision on $\Pi = |\hat{\alpha} - \hat{\alpha}_1||\hat{\alpha} - \hat{\alpha}_2|$. The distribution of Π under H_0 is evaluated from 1000 realizations of a Gaussian DPP model with $(\rho, \alpha) = (303, 0.0181)$. On one hand, we fit a Gaussian DPP model to each realization to obtain an estimate of α and, on the other hand, we apply an independent thinning with retention probability $\hat{\rho}_1/\hat{\rho} = 0.75$ and fit a Gaussian DPP model to both the retained points and the thinned points to obtain estimates of both α_1 and α_2. The distribution of Π over the 400 simulations is compared with the empirical value of Π for the data set. The resulting p-value is 0.55 and there is no reason to reject the null hypothesis.

6. Inference for non-stationary models

In this section, we use the Japanese pines data set (Fig. 2(c)) from Numata (1964) to illustrate

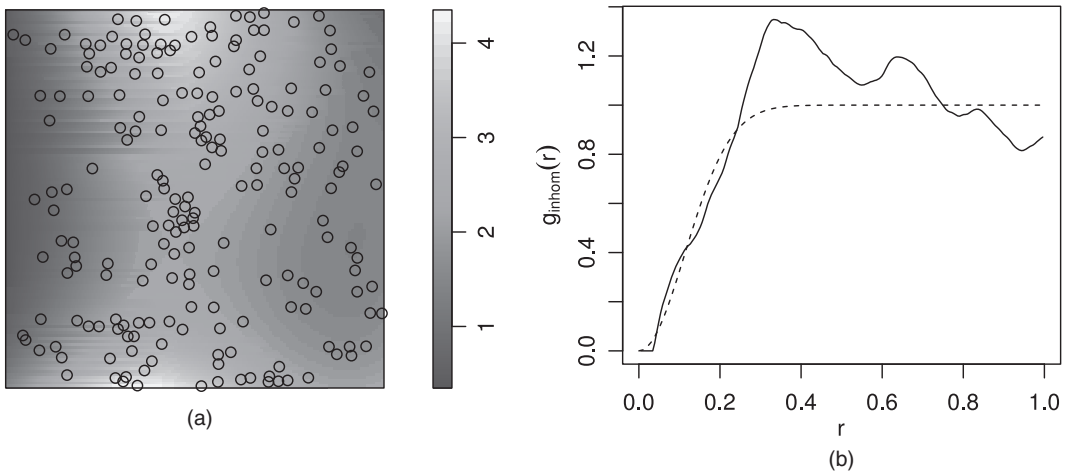

Fig. 7. Japanese pines data set: (a) estimated intensity function $\rho_{\hat{\psi}}$ with the points of the data set overlayed; (b) empirical pair correlation function (———) and theoretical pair correlation function (------) for the fitted Gaussian DPP model

how a non-stationary DPP model can be fitted to a point pattern data set in practice. The data consist of the locations of 204 seedlings and saplings of Japanese black pines (*Pinus Thunbergii*) in a $10\,\mathrm{m} \times 10\,\mathrm{m}$ region. The data have previously been analysed by among others Ogata and Tanemura (1986) by using an inhomogeneous Gibbs model where the logarithm of the first-order term was assumed to be a cubic polynomial in the Cartesian co-ordinates. Since we have no access to external covariates (such as soil quality and type of terrain) we use the Cartesian co-ordinates as an artificial way of accounting for spatial heterogeneity. However, the methodology that is described below works equally well for data sets with external covariates.

There are at least two different possible strategies for constructing a non-stationary DPP model. In Section 6.1, we assume that the DPP is second-order intensity-reweighted stationary (Baddeley *et al.*, 2000) which allows us to devise quite a generally applicable strategy of analysis. Alternatively we could construct a non-stationary DPP model by transforming a stationary DPP, in which case expression (2.4) gives a closed form expression for the kernel of the non-stationary DPP. This approach is exemplified in the on-line supplementary material, but it is very specific to the data set analysed and it appears to be difficult to devise a general strategy for modelling non-stationarity this way. In both approaches, we exploit that the kernel (covariance function) C of a DPP can be written as

$$C(x, y) = \sqrt{\rho(x)}\, R(x, y) \sqrt{\rho(y)} \tag{6.1}$$

where ρ is the intensity and R is the corresponding correlation function to C.

6.1. Second-order intensity-reweighted stationary models

Let X be a DPP on a compact set S with intensity function ρ and correlation function R. Then X is second-order intensity-reweighted stationarity if R is invariant by translation, since this implies that g is translation invariant; see expression (2.3). If we further assume that R is isotropic, we have $R(x, y) = R_0(\|x - y\|)$. Let ρ_b be an upper bound of $\rho(x)$, $x \in S$, and assume that a stationary isotropic DPP X^{dom} with kernel

$$C^{\mathrm{dom}}(x, y) = C_0^{\mathrm{dom}}(\|x - y\|) = \rho_\mathrm{b}\, R_0(\|x - y\|)$$

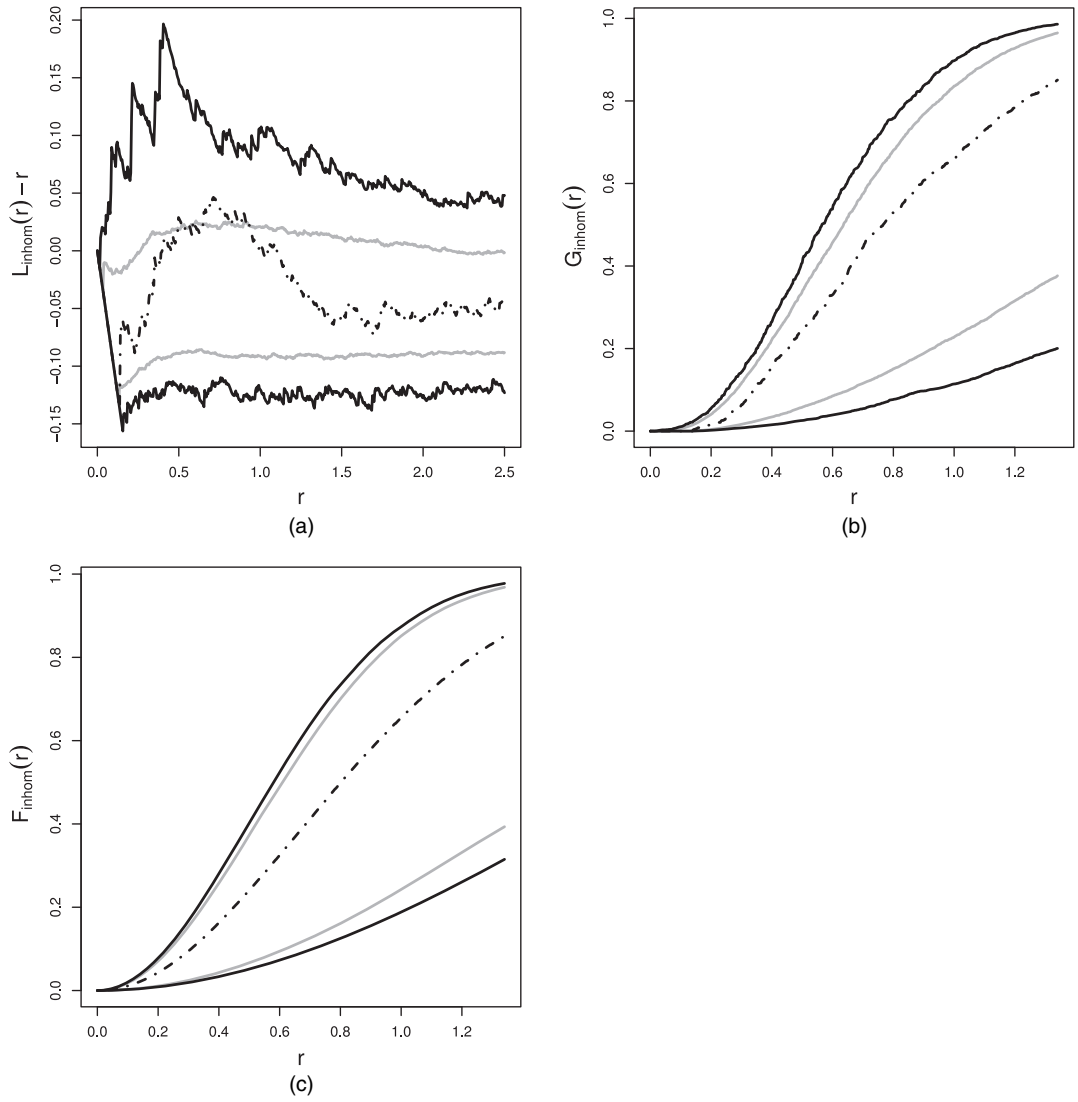

Fig. 8. Non-parametric estimates of (a) $L(r) - r$, (b) $G(r)$ and (c) $F(r)$ for the Japanese pines data set with simulation-based envelopes for the fitted Gaussian DPP model (both sets of envelopes are based on 4000 simulated realizations): ——, 2.5% and 97.5% pointwise quantiles; ———, rank envelopes with a 5% significance level; - - - - -, data

is well defined. By the thinning result (2.5), X corresponds to an independent thinning of X^{dom} with retention probability $\rho(x)/\rho_{\mathrm{b}}$. A parametric model can then be constructed by using the models of Sections 3.3 and 3.4 for R_0 and specifying a parametric model ρ_ψ for ρ, which possibly depends on spatial covariates. We estimate ψ by the Poisson maximum likelihood estimator $\hat{\psi}$ (Schoenberg, 2005). Further, let $R_{0,\theta}$ be parameterized by θ. Then we estimate θ by the minimum contrast estimate $\hat{\theta}$ based on the pair correlation function $g_{0,\theta}(r) = 1 - |R_{0,\theta}(r)|^2, r \geqslant 0$, as detailed below (we could also use the K-function for estimation, but in a simulation study in the on-line supplementary material using g is better).

Specifically, for the Japanese pines data set, we assume that $\log(\rho_\psi)$ is a cubic polynomial in

Cartesian co-ordinates. This is in close analogy with the analysis of Ogata and Tanemura (1986), but we have the advantage of modelling the intensity directly, whereas the intensity is unknown in their analysis. Fig. 7(a) shows the fitted intensity function with the points of the data set overlayed. Further, given $\rho_{\hat{\psi}}$, we define $\hat{\rho}_{\mathrm{b}} = \sup_{x \in S}\{\rho_{\hat{\psi}}(x)\}$. Then the existence condition (3.3) for X^{dom} with kernel $\hat{\rho}_{\mathrm{b}} R_{0,\theta}$ defines the valid parameter space $\Theta_{\hat{\rho}}$. Using $\rho_{\hat{\psi}}$ for reweighting we can non-parametrically estimate the pair correlation function (Baddeley *et al.*, 2000). We then find the minimum contrast estimate as the value of $\theta \in \Theta_{\hat{\rho}}$ which minimizes

$$D(\theta) = \int_{r_l}^{r_u} |\hat{g}_{0,\theta}(r)^q - g_{0,\theta}(r;\theta)^q|^p \, \mathrm{d}r.$$

Here $r_l < r_u$, $p > 0$ and $q > 0$ are user-specified parameters. Following the recommendations in Diggle (2003), we let $q = \frac{1}{2}$, $p = 2$ and r_u be a quarter of the minimal side length of S. It is customary to use $r_l = 0$, which we recommend when the minimum contrast estimate is based on the K-function. However, when the minimum contrast estimate is based on g, we let r_l be 1% of the minimal side length of S to avoid numerical instabilities of the estimate of g close to zero. Finally, to choose between several different models fitted to the same data set, we choose the model with minimal value of $D(\theta)$.

Since we do not have a closed form expression for the pair correlation function of the power exponential spectral model, we have omitted that model from this analysis. The fitted Gaussian model has $\hat{\alpha} = 0.226$ and Fig. 7(b) shows the fitted pair correlation function. The estimates for both the Whittle–Matérn and Cauchy models have very large values of ν indicating that the fitted model is close to the Gaussian model; in fact plots of the fitted pair correlation functions (which are not shown here) coincide with that for the Gaussian model. Also the values of $D(\theta)$ are very similar for all three models. Therefore we prefer the simpler Gaussian model.

Similarly to Fig. 6 in Section 5.2, Fig. 8 is used to assess the goodness of fit for the Gaussian model. Overall the model appears to provide an acceptable fit. The estimate of $L(r) - r$ exits the pointwise envelopes but stays within the rank envelopes, so this deviation is not significant at the 5% level.

7. Concluding remarks

We do not think of a DPP as a mechanistic model, i.e. a model which describes a physical process generating a spatial point process data set. Rather our overall purpose of fitting DPP models is to provide empirical models with a parsimonious parameterization, where we can compare different spatial point pattern data sets by comparing their estimated DPP model parameters, their maximized likelihoods, their intensities and pair correlation functions as well as other summaries. This is also possible when fitting parametric Poisson process models, Poisson cluster process models and Cox process models (see for example Møller and Waagepetersen (2007) and the references therein); however, these are not models for repulsion (or regularity or inhibition) but rather for no interaction in the Poisson case and clustering or aggregation in the other cases. Moreover, as mentioned, for Gibbs point processes it is in general more complicated to use a maximum likelihood approach and it is not possible to find the intensity, the pair correlation function or other moment properties except by using time-consuming simulations.

DPPs cannot be as repulsive as Gibbs hard-core point processes can be. Nevertheless, the jinc-like DPP exhibits strong repulsiveness (Fig. 1(b)) and allows us to fit quite regular data sets; see the on-line supplementary material, where in connection with a regular point pattern data set we compare the jinc-like DPP with a Strauss process (Strauss, 1975; Kelly and Ripley, 1976) which is a standard example of a repulsive Gibbs point process. Not only the moments

of the jinc-like DPP are expressible in closed form, unlike the Strauss model, but it turns out that simulation of the fitted jinc-like DPP is very fast, whereas long Markov chain Monte Carlo simulations are needed for the fitted Strauss process.

Whittle–Matérn and Cauchy models with low values of ν (e.g. $\nu < 0.5$) are very close to Poisson models, and in our experience it requires a rather large point pattern data set before the null Poisson hypothesis is rejected by a likelihood ratio test. Such models (close to Poisson) are difficult to estimate and simulate, since very large values of N will be needed in the truncations that were discussed in Section 5.1 to obtain satisfactory approximations of C and $\tilde{C}$.

In general there is an inverse relationship between the range of correlation and the spread of the spectral density: if C_0 decays rapidly then φ decays slowly, and if C_0 decays slowly then φ decays rapidly. This is in line with the following fact: the (generalized) Fourier transform of the Dirac delta function (over $\mathbb{R}^d$) is 1 and vice versa, and the Dirac function is the kernel of the Poisson process. From an 'end-user' point of view this is very important: everything works well and is fast for DPPs except in the less interesting cases which are close to Poisson processes. In such cases very weakly repulsive Gibbs point processes become interesting competitors to DPPs, unless some other and more efficient approximations of C and $\tilde{C}$ are developed.

In Section 4 we discussed useful approximations of C and $\tilde{C}$ restricted to $R \times R$ when $R \subset \mathbb{R}^d$ is rectangular. Frequently in the spatial point process literature, including the present paper, spatial point pattern data sets observed within a rectangular region are considered. However, applications with non-rectangular observation windows are not uncommon; see for example Harkness and Isham (1983). It remains to clarify how such cases should be handled when fitting DPP models. We could embed a non-rectangular observation window W in a rectangular region R and consider the situation as a missing data problem, since we are missing the events in $R \setminus W$, and at least the 'complete likelihood' can be handled. In our opinion this seems a difficult approach for maximum likelihood. However, moment-based estimation and other simple alternatives to maximum likelihood as discussed in this paper will easily apply.

Large point pattern data sets may exhibit aggregation on the large scale and repulsiveness on the small scale. For this purpose DPP models which depend on spatial covariates may be sufficient in some cases. As another possibility we are currently developing models for dependent thinning of DPPs.

Generalizations of DPPs to weighted DPPs, which also are models for repulsion, and to the closely related permanental and weighted permanental point processes, which are models for attraction, were studied in Shirai and Takahashi (2003) and McCullagh and Møller (2006). Since determinants have a geometric meaning and are multiplicative, and there are faster algorithms for computations, DPPs are much easier to deal with, not least from a statistical and computational perspective. The approximations of C and $\tilde{C}$ by using a Fourier basis approach (Section 4) apply as well for weighted DPPs and weighted permanental point processes, but the practical usefulness of the approximations is yet unexplored in these cases.

Though a DPP is broadly speaking a special case of a Gibbs point process, we are not aware of a simple Hammersley–Clifford–Ripley–Kelly representation (see Ripley and Kelly (1977)) in terms of a product of interaction functions (or, using the terminology of statistical physics, a sum of potentials). Gibbsianness of DPPs has been studied in Georgii and Yoo (2005). In our opinion the Markovian properties of DPPs is still an interesting area of research.

Another open research problem is the development of determinantal space–time point process models. In the continuous time case, formally we are then just dealing with a DPP defined on $\mathbb{R} \times \mathbb{R}^d$ (where $\mathbb{R}$ is considered to be the time axis), but the natural direction on the time axis should be taken into consideration when developing parametric families of space–time covariance functions and understanding how they can be used for modelling repulsion between

events in time or space or both time and space. Also the development of statistical inference procedures for such models is a challenge. Recently, in a discrete time setting, Affandi *et al.* (2012) have constructed a Markov chain of DPPs with a finite state space. It would be interesting to study a similar Markov chain construction for our case with state space $\mathbb{R}^d$.

Acknowledgements

We are grateful to Philippe Carmona, Morten Nielsen and Rasmus Waagepetersen for helpful comments and to Adrian Baddeley for supplying the Japanese pines data set, which Yosihiko Ogata and Masaharu Tanemura kindly granted us permission to use. This work was supported by the Danish Council for Independent Research—Natural Sciences, grant 09-072331, 'Point process modelling and statistical inference', and grant 12-124675, 'Mathematical and statistical analysis of spatial data'. It was also supported by the Centre for Stochastic Geometry and Advanced Bioimaging, funded by a grant from the Villum Foundation.

An alphabetical ordering has been used for the authors' names since all the authors made significant contributions to the paper.

References

Affandi, R., Kulesza, A. and Fox, E. (2012) Markov determinantal point processes. In *Proc. 28th Conf. Uncertainty in Artificial Intelligence* (ed. K. M. N. de Freitas), pp. 26–35. Corvallis: Association for Uncertainty in Artificial Intelligence Press.

Baddeley, A., Møller, J. and Waagepetersen, R. (2000) Non- and semi-parametric estimation of interaction in inhomogeneous point patterns. *Statist. Neerland.*, **54**, 329–350.

Baddeley, A. and Turner, R. (2000) Practical maximum pseudolikelihood for spatial point patterns. *Aust. New Zeal. J. Statist.*, **42**, 283–322.

Baddeley, A. and Turner, R. (2005) Spatstat: an R package for analyzing spatial point patterns. *J. Statist. Softwr.*, **12**, no. 6, 1–42.

Besag, J. (1977) Some methods of statistical analysis for spatial data. *Bull. Int. Statist. Inst.*, **47**, 77–92.

De laco, S., Palma, M. and Posa, D. (2003) Covariance functions and models for complex-valued random fields. *Stoch. Environ. Res. Risk Assessmnt*, **17**, 145–156.

Diggle, P. (2003) *Statistical Analysis of Spatial Point Patterns*, 2nd edn. London: Hodder Arnold.

Gelfand, A. E., Diggle, P. J., Guttorp, P. and Fuentes, M. (2010) *Handbook of Spatial Statistics*. Boca Raton: CRC Press.

Georgii, H.-O. and Yoo, H. J. (2005) Conditional intensity and Gibbsianness of determinantal point processes. *J. Statist. Phys.*, **118**, 617–666.

Glass, L. and Tobler, W. R. (1971) Uniform distribution of objects in a homogeneous field: cities on a plain. *Nature*, **233**, 67–68.

Gneiting, T. (2002) Compactly supported correlation functions. *J. Multiv. Anal.*, **83**, 493–508.

Goovaerts, P. (1997) *Geostatistics for Natural Resources Evaluation*. New York: Oxford University Press.

Harkness, R. D. and Isham, V. (1983) A bivariate spatial point pattern of ants' nests. *Appl. Statist.*, **32**, 293–303.

Hough, J. B., Krishnapur, M., Peres, Y. and Viràg, B. (2006) Determinantal processes and independence. *Probab. Surv.*, **3**, 206–229.

Hough, J. B., Krishnapur, M., Peres, Y. and Viràg, B. (2009) *Zeros of Gaussian Analytic Functions and Determinantal Point Processes*. Providence: American Mathematical Society.

Huang, F. and Ogata, Y. (1999) Improvements of the maximum pseudo-likelihood estimators in various spatial statistical models. *J. Computnl Graph. Statist.*, **8**, 510–530.

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

Jensen, J. L. and Møller, J. (1991) Pseudolikelihood for exponential family models of spatial point processes. *Ann. App. Probab.*, **1**, 445–461.

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

Kulesza, A. and Taskar, B. (2012) Determinantal point processes for machine learning. *Found. Trends Mach. Learn.*, **5**, 123–286.

Leonardi, E. and Torrisi, G. L. (2013) Large deviations of the interference in the Ginibre network model. *Preprint arXiv:1304.2234*.

van Lieshout, M. N. M. (2000) *Markov Point Processes and Their Applications*. London: Imperial College Press.

Lindgren, F., Lindström, J. and Rue, H. (2011) An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach (with discussion). *J. R. Statist. Soc.* B, **73**, 423–498.
Macchi, O. (1975) The coincidence approach to stochastic point processes. *Adv. Appl. Probab.*, **7**, 83–122.
McCullagh, P. and Møller, J. (2006) The permanental process. *Adv. Appl. Probab.*, **38**, 873–888.
Miyoshi, N. and Shirai, T. (2013) A cellular network model with ginibre configurated base stations. *Technical Report*. Department of Mathematical and Computing Sciences, Tokyo Institute of Technology, Tokyo.
Møller, J. and Waagepetersen, R. P. (2004) *Statistical Inference and Simulation for Spatial Point Processes*. Boca Raton: Chapman and Hall–CRC.
Møller, J. and Waagepetersen, R. P. (2007) Modern spatial point process modelling and inference (with discussion). *Scand. J. Statist.*, **34**, 643–711.
Myllymäki, M., Mrkvicka, T., Seijo, H. and Grabarnik, P. (2013) Global envelope tests for spatial processes. *Preprint arXiv:1307.0239*.
Numata, M. (1964) Forest vegetation, particularly pine stands in the vicinity of Choshi—flora and vegetation in Choshi, Chiba prefecture, VI (in Japanese). *Bull. Choshi Mar. Lab.*, no. 6, 27–37.
Ogata, Y. and Tanemura, M. (1986) Likelihood estimation of interaction potentials and external fields of inhomogeneous spatial point patterns. In *Proc. Pacif. Statist. Congr.* (eds I. Francis, B. Manly and F. Lam), pp. 150–154. Amsterdam: Elsevier.
R Development Core Team (2011) *R: a Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.
Riesz, F. and Sz.-Nagy, B. (1990) *Functional Analysis*. New York: Dover Publications.
Ripley, B. D. (1976) The second-order analysis of stationary point processes. *J. Appl. Probab.*, **13**, 255–266.
Ripley, B. D. (1977) Modelling spatial patterns (with discussion). *J. R. Statist. Soc.* B, **39**, 172–212.
Ripley, B. D. (1988) *Statistical Inference for Spatial Processes*. Cambridge: Cambridge University Press.
Ripley, B. D. and Kelly, F. P. (1977) Markov point processes. *J. Lond. Math. Soc.*, **15**, 188–192.
Scardicchio, A., Zachary, C. and Torquato, S. (2009) Statistical properties of determinantal point processes in high-dimensional Euclidean spaces. *Phys. Rev.* E, **79**, no. 4, article 041108.
Schoenberg, F. (2005) Consistent parametric estimation of the intensity of a spatial temporal point process. *J. Statist. Planng Inf.*, **128**, 79–93.
Shirai, T. and Takahashi, Y. (2003) Random point fields associated with certain Fredholm determinants: I, Fermion, Poisson and boson point processes. *J. Functnl Anal.*, **2**, 414–463.
Soshnikov, A. (2000) Determinantal random point fields. *Russ. Math. Surv.*, **55**, 923–975.
Stein, E. M. and Weiss, G. (1971) *Introduction to Fourier Analysis on Euclidean Spaces*. Princeton: Princeton University Press.
Stoyan, D., Kendall, W. S. and Mecke, J. (1995) *Stochastic Geometry and Its Applications*, 2nd edn. Chichester: Wiley.
Strauss, D. J. (1975) A model for clustering. *Biometrika*, **63**, 467–475.
Wu, Z. (1995) Compactly supported positive definite radial functions. *Adv. Computnl Math.*, **4**, 283–292.
Yaglom, A. M. (1987) *Correlation Theory of Stationary and Related Random Functions*. New York: Springer.

Supporting information

Additional 'supporting information' may be found in the on-line version of this article:

'Determinantal point process models and statistical inference: Extended version'.