



HAL
open science

Determinantal point process models and statistical inference

Frédéric Lavancier, Jesper Møller, Ege Rubak

► **To cite this version:**

Frédéric Lavancier, Jesper Møller, Ege Rubak. Determinantal point process models and statistical inference. 2012. hal-00698958v3

HAL Id: hal-00698958

<https://hal.science/hal-00698958v3>

Preprint submitted on 29 Jul 2013 (v3), last revised 23 Jun 2014 (v4)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Determinantal point process models and statistical inference

Frédéric Lavancier¹, Jesper Møller² and Ege Rubak^{*2}

¹ Laboratoire de Mathématiques Jean Leray, University of Nantes, France,
Frederic.Lavancier@univ-nantes.fr

²Department of Mathematical Sciences, Aalborg University, jm@math.aau.dk,
rubak@math.aau.dk

Abstract

Statistical models and methods for determinantal point processes (DPPs) seem largely unexplored. We demonstrate that DPPs provide useful models for the description of repulsive spatial point processes, particularly in the ‘soft-core’ case. 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 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, simulation, repulsiveness, spectral approach.

1 Introduction

1.1 Aim of the paper

Determinantal point processes (DPPs) are largely unexplored in statistics, though they possess a number of very attractive properties and 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,

*An alphabetical ordering has been used since all authors have made significant contributions to the paper.

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.

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

- (i) to provide a short and accessible survey for statisticians on the definition, existing 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;
- (iii) to understand to which 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, particularly in the ‘soft-core’ case and to some extent in cases with more repulsion;
- (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 using the likelihood or moment properties of DPP models;
- (vi) to apply our methodology on real spatial point pattern datasets showing different degrees of repulsiveness;
- (vii) to provide freely available software for simulation and statistical inference.

While Hough et al. (2009) provides 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 needed background material for our new contributions (items (ii)-(vii) above).

1.2 Repulsiveness and point pattern datasets

Repulsiveness in a spatial point process means that nearby points in the process are repelling each other. Formal definitions of this concept will be given later in this paper, either in terms of second order properties (the so-called pair correlation function and the K -function, see Sections 2.2 and 3.5) or in terms of the Papangelou conditional intensity (Appendix G).

The following examples of planar point pattern datasets have been selected to illustrate the range of repulsiveness which can be modelled by DPPs. These dataset will be detailed and analysed in Section 6. The five point pattern in Figure 1(a)-(e) will be fitted using parsimonious parametric models of DPPs, while the last dataset

in Figure 1(f) is an example of point pattern that seems too regular to be fitted by a DPP.

For comparison Figure 2 shows realizations in the unit square of three stationary DPPs with the same intensity of points. Figure 2(a) shows a simulation of a Poisson process, which is a special case of a DPP with no repulsion (or no interaction). Figure 2(b) shows a simulation of a DPP with moderate repulsion; in comparison with Figure 2(a) the point pattern looks more regular. Figure 2(c) shows, in a sense made more precise in Section 3.5, a simulation of a DPP for the strongest case of repulsiveness when the intensity is fixed. The point pattern in Figure 2(c) is clearly regular but not to the same extent as in a Gibbs hard-core point process or as in the cell dataset in Figure 1(f).

1.3 Gibbs point processes versus determinantal point processes

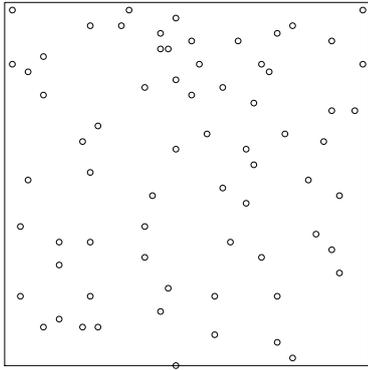
The usual class of point processes used for modelling repulsiveness is the class of Gibbs point processes, including Markov point processes and pairwise interaction point processes (Ripley, 1977; Ripley and Kelly, 1977; Stoyan et al., 1995; Lieshout, 2000; Diggle, 2003; Møller and Waagepetersen, 2004; Illian et al., 2008; Gelfand et al., 2010). In general for Gibbs point processes,

- moments are not expressible in closed form;
- likelihoods involve intractable normalizing constants;
- rather elaborate Markov chain Monte Carlo methods are needed for simulations and approximate likelihood inference;
- 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;

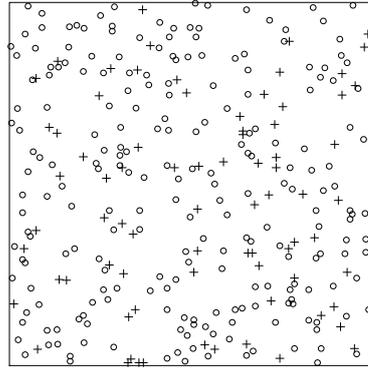
see Møller and Waagepetersen (2004, 2007) and the references therein. For Gibbs point processes, as maximum likelihood inference is complicated, the most popular and much quicker alternative inference procedure is based on pseudo-likelihood Besag (1977a); Jensen and Møller (1991); Baddeley and Turner (2000); Gelfand et al. (2010). The pseudo-likelihood function is specified in terms of the Papangelou conditional intensity which does not depend on the normalizing constant from the likelihood.

In contrast, DPPs possess a number of appealing properties: Considering a DPP defined on \mathbb{R}^d (with $d = 2$ in most of our examples), its distribution is specified by a kernel (which we assume is a continuous complex covariance function) C defined on $\mathbb{R}^d \times \mathbb{R}^d$ and which is properly scaled (these regularity conditions on C are imposed to ensure existence of the process as discussed in Section 2.3). Then

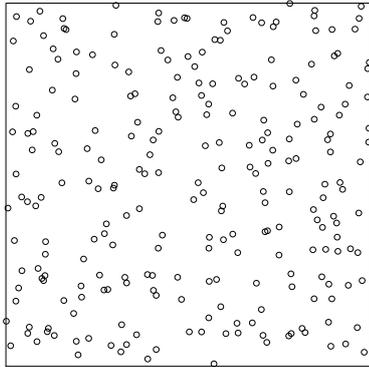
- (a) there are simple conditions for existence of the process and there is no phase transition: uniqueness of the DPP is ensured when it exists;



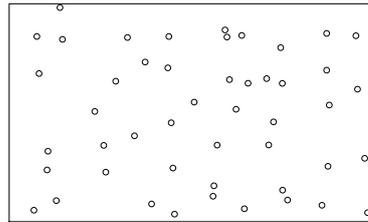
(a) Locations of 69 Spanish towns in a 40 mile by 40 mile region.



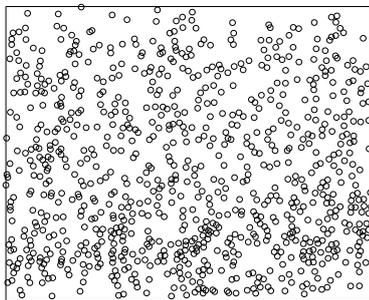
(b) Locations of 303 cells of two types in a 0.25 mm by 0.25 mm region of a histological section of the kidney of a hamster.



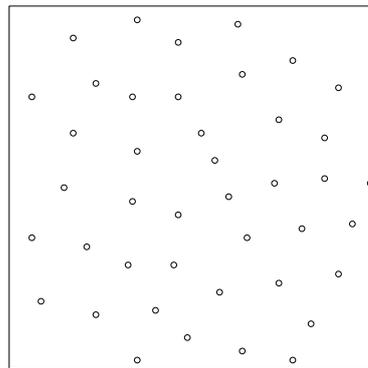
(c) Locations of 244 trees of the species oak and beech in a 80 m by 80 m region.



(d) Locations of 48 termite mounds in a 250 m by 150 m region.



(e) Locations of 876 cells of a mucous membrane in a rectangular region rescaled to unit width and height 0.81.



(f) Locations of 42 cells in a histological section of an insect rescaled to the unit square.

Figure 1: Examples of point patterns.

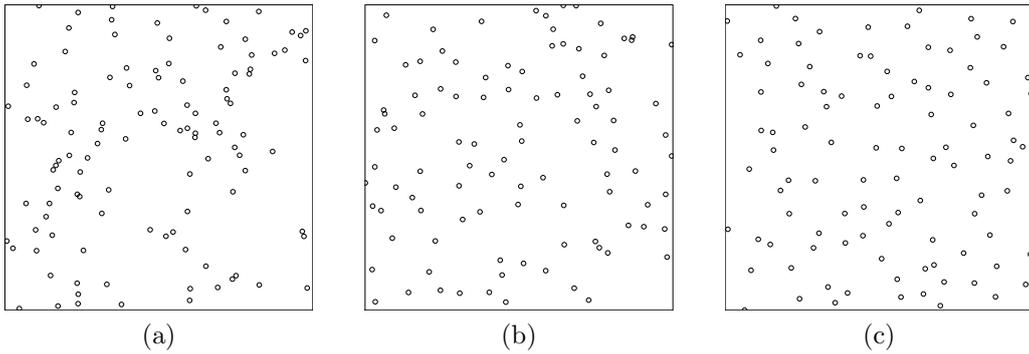


Figure 2: Realizations of stationary DPPs within a unit square: (a) a Poisson process; (b) a DPP with moderate repulsion (a Gaussian DPP as described in Section 3.3); (c) a stronger repulsive DPP (a jinc-like DPP as described in Section 3.4).

- (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 may not be a problem as 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).

Indeed, DPPs possess further useful properties, e.g. a one-to-one smooth transformation or an independent thinning of the DPP is also a DPP (Appendix A); the reduced Palm measure of a DPP is also a DPP (Appendix C). Due to (a)-(f), modelling and estimation for parametric families of DPPs become tractable as discussed in Sections 3-7. In particular, we shall calculate likelihood functions, maximum likelihood estimates, and likelihood ratio statistics for parametric DPP models.

The link between Gibbs point processes and DPPs have 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 pseudo-likelihood 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.

1.4 Software

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

1.5 Outline

The paper is organized as follows. Section 2 is our tutorial (cf. (i) in Section 1.1). 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; and to understand to which 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} (cf. (d)-(e) in Section 1.3) which make simulation and inference feasible for parametric models of DPPs. Section 5 discusses statistical inference procedures for parametric models of DPPs, using either a maximum likelihood approach or simpler moment-based alternatives. In Section 6 we fit parsimonious parametric DPP models to the datasets presented in Section 1.2. Section 7 contains our concluding remarks. Finally, Appendices A-K contain the technical proofs of our results and provide supplementary methods, examples, and remarks to the ones presented in the main text.

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

The following provides the background material needed in this paper on the definition (Sections 2.1-2.2), existence (Section 2.3), simulation (Section 2.4), and density expression for a general DPP defined on a Borel set $B \subseteq \mathbb{R}^d$ (Section 2.5). 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 bit 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 e.g. 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 DPPs are defined in terms of their moment properties as expressed by their so-called product density functions, we start by recalling this notion.

For $n = 1, 2, \dots$, X has n 'th order product density function $\rho^{(n)} : B^n \rightarrow [0, \infty)$ if this function is locally integrable (with respect to Lebesgue measure restricted to B^n) and for any Borel function $h : B^n \rightarrow [0, \infty)$,

$$\mathbb{E} \sum_{x_1, \dots, x_n \in X}^{\neq} h(x_1, \dots, x_n) = \int_B \cdots \int_B \rho^{(n)}(x_1, \dots, x_n) h(x_1, \dots, x_n) dx_1 \cdots dx_n \quad (2.1)$$

where \neq over the summation sign means that x_1, \dots, x_n are pairwise distinct events. See e.g. Stoyan et al. (1995). Intuitively, for any pairwise distinct points $x_1, \dots, x_n \in B$, $\rho^{(n)}(x_1, \dots, x_n) dx_1 \cdots dx_n$ is the probability that for each $i = 1, \dots, n$, X has a point in an infinitesimally small region around x_i of volume dx_i . Clearly, $\rho^{(n)}$ is only uniquely defined up to a Lebesgue nullset. We shall henceforth require that $\rho^{(n)}(x_1, \dots, x_n) = 0$ if $x_i = x_j$ for some $i \neq j$. This convention becomes consistent with Definition 2.1 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 zero. By our convention above, $g(x, x) = 0$ for all $x \in B$. The terminology ‘pair correlation function’ may be confusing, but it 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 $z = z_1 + iz_2$ (where $z_1, z_2 \in \mathbb{R}$ and $i = \sqrt{-1}$), we denote $\bar{z} = z_1 - iz_2$ the complex conjugate and $|z| = \sqrt{z_1^2 + z_2^2}$ the modulus. For any function $C : B \times B \rightarrow \mathbb{C}$, let $[C](x_1, \dots, x_n)$ be the $n \times n$ matrix with (i, j) ’th entry $C(x_i, x_j)$. For a square complex matrix A , let $\det A$ denote its determinant.

Definition 2.1. *Suppose that a simple locally finite spatial point process X on B has product density functions*

$$\rho^{(n)}(x_1, \dots, x_n) = \det[C](x_1, \dots, x_n), \quad (x_1, \dots, x_n) \in B^n, \quad n = 1, 2, \dots \quad (2.2)$$

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

Note that a Poisson process is the special case where $C(x, y) = 0$ whenever $x \neq y$.

Remark 2.2. For $X \sim \text{DPP}_B(C)$ and any Borel set $A \subseteq B$, define $X_A = X \cap A$ and denote its distribution by $\text{DPP}_B(C; A)$. We also write $\text{DPP}_A(C)$ 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 1.3 follows directly from Definition 2.1, i.e. $\text{DPP}_A(C) = \text{DPP}_B(C; A)$. Further, when $B = \mathbb{R}^d$, we write $\text{DPP}(C)$ for $\text{DPP}_{\mathbb{R}^d}(C)$, and $\text{DPP}(C; A)$ for $\text{DPP}_{\mathbb{R}^d}(C; A)$.

Some further remarks are in order. Let $X \sim \text{DPP}(C)$. Then there is no other point process satisfying (2.2) (Lemma 4.2.6 in Hough et al. (2009)). By (2.2), the intensity function of X is

$$\rho(x) = C(x, x), \quad x \in \mathbb{R}^d, \quad (2.3)$$

and the pair correlation function of X is

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

while it is zero otherwise. If C is Hermitian, then $g \leq 1$, showing that the events of X repel each other. Furthermore, if C is continuous, $\rho^{(n)}$ is also continuous and $\rho^{(n)}(x_1, \dots, x_n)$ tends to zero as the Euclidean distance $\|x_i - x_j\|$ goes to zero for some $i \neq j$, cf. (2.2). This once again reflects the repulsiveness of a DPP.

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

$$g(x, y) = 1 - |R(x, y)|^2, \quad x, y \in \mathbb{R}^d, \quad (2.4)$$

and when C is a covariance function, R is its corresponding correlation function.

2.3 Existence

Existence of a DPP on \mathbb{R}^d is ensured by the following assumptions (C1)-(C2) on C .

Note that $C : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{C}$ needs to be non-negative definite to ensure $\rho^{(n)} \geq 0$ in (2.2). 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 $x, y \in \mathbb{R}^d$. As argued below we find it natural to consider the following condition:

$$(C1) \quad C \text{ is a continuous complex covariance function.}$$

Then, if we let $S \subset \mathbb{R}^d$ denote a generic compact set and $L^2(S)$ the space of square-integrable functions $h : S \rightarrow \mathbb{C}$, we obtain the following by Mercer's theorem (see e.g. Section 98 in Riesz and Sz. Nagy (1990)). Under (C1), C restricted to $S \times S$ has a spectral representation,

$$C(x, y) = \sum_{k=1}^{\infty} \lambda_k \phi_k(x) \overline{\phi_k(y)}, \quad (x, y) \in S \times S, \quad (2.5)$$

with absolute and uniform convergence of the series, and where

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

$$\int_S \phi_k(x) \overline{\phi_l(x)} dx = \begin{cases} 1 & \text{if } k = l, \\ 0 & \text{if } k \neq l, \end{cases} \quad (2.6)$$

and any $h \in L^2(S)$ can be written as $h = \sum_{k=1}^{\infty} \alpha_k \phi_k$, where $\alpha_k \in \mathbb{C}$, $k = 1, 2, \dots$. 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 . Finally, we consider the following condition:

$$(C2) \quad \lambda_k^S \leq 1 \text{ for all compact } S \subset \mathbb{R}^d \text{ and all } k.$$

Theorem 2.3. *Under (C1), existence of DPP(C) is equivalent to (C2).*

This result is verified in Appendix B. Usually, for statistical models of covariance functions, (C1) is satisfied, and so (C2) becomes the essential condition. As discussed in Section 3.2, (C2) simplifies in the stationary case of X .

Assumption 2.4. *In the remainder of this paper, $X \sim \text{DPP}(C)$ with C satisfying the conditions (C1) and (C2).*

Remark 2.5. Various comments on (C1) and (C2) are in order.

As noticed in Hough et al. (2009), there are interesting examples of DPPs with non-Hermitian kernels, but they do not possess various general properties, and the results and methods in our paper rely much on the spectral representation (2.5). We therefore confine ourselves to the Hermitian case of C .

We find that (C1) is often a natural condition for several reasons: statisticians are used to deal with covariance functions; as seen in the proof of Theorem 2.3, the situation simplifies when C is assumed to be continuous; continuity of C implies continuity of the intensity function and the pair correlation function; conversely, if C is real and non-negative, continuity of ρ and g implies continuity of C .

When we are only interested in considering a DPP Y on a given compact set $S \subset \mathbb{R}^d$, then (C1)-(C2) can be replaced by the assumption that C is a continuous complex covariance function defined on $S \times S$ such that $\lambda_k^S \leq 1$ for all k . The results in Sections 2.4-2.5 are then valid for Y , even if there is no continuous extension of C to $\mathbb{R}^d \times \mathbb{R}^d$ which satisfies (C1)-(C2). However, it is convenient to assume (C1)-(C2) as we in Sections 3-5 consider stationary DPPs.

Though a Poisson process is determinantal from Definition 2.1, it is excluded by our approach where C is continuous. In particular, (2.5) does not hold for a Poisson process (therefore many of our results as well as those established in Hough et al. (2009) do not hold for a Poisson process).

Definition 2.6. *Let $S \subset \mathbb{R}^d$ be compact and assume all non-zero eigenvalues λ_k^S are one. Then C restricted to $S \times S$ is called a projection kernel, and X_S is called a determinantal projection point process.*

The terminology in Definition 2.6 seems commonly used (e.g. Hough et al. (2006) and Hough et al. (2009)); McCullagh and Møller (2006) call a determinantal projection point process a special DPP because of its special properties as discussed below.

2.4 Simulation

An algorithm for simulating a finite DPP in a very general setup is provided in Hough et al. (2006). There are special cases of DPPs which may be simulated in a different manner, e.g. the Ginibre ensemble, see Section 4.3 in Hough et al. (2009).

We explain and prove the simulation algorithm of Hough et al. (2006) in the specific case where we want to simulate $X_S \sim \text{DPP}(C; S)$ with $S \subset \mathbb{R}^d$ compact. Our implementation of the algorithm becomes more efficient than the one in Scardicchio et al. (2009), and our description and proof use mainly linear algebra and are less technical than that in Hough et al. (2006).

Consider the spectral representation (2.5) of C restricted to $S \times S$. The simulation algorithm is based on the following result (Theorem 7 in Hough et al. (2006); see also Theorem 4.5.3 in Hough et al. (2009)).

Theorem 2.7. *For $k = 1, 2, \dots$, let B_k be independent Bernoulli variables with mean λ_k . Define the random projection kernel $K : S \times S \rightarrow \mathbb{C}$ by*

$$K(x, y) = \sum_{k=1}^{\infty} B_k \phi_k(x) \overline{\phi_k(y)}. \quad (2.7)$$

Then

$$DPP_S(K) \sim DPP(C; S) \quad (2.8)$$

in the sense that if we first generate the independent Bernoulli variables, and second generate a determinantal projection point process on S with kernel K , then the resulting point process follows $DPP(C; S)$.

Note that if $N(S) = n(X_S)$ denotes the number of events in S , then

$$N(S) \sim \sum_{k=1}^{\infty} B_k, \quad \mathbb{E}[N(S)] = \sum_{k=1}^{\infty} \lambda_k, \quad \text{Var}[N(S)] = \sum_{k=1}^{\infty} \lambda_k(1 - \lambda_k). \quad (2.9)$$

The first result in (2.9) follows from (2.8) and Theorem 2.8 below (or from Lemma 4.4.1 in Hough et al. (2009)), and the first result immediately implies the two other results.

2.4.1 Simulation of Bernoulli variables

This and the following section describe a two step simulation procedure based on Theorem 2.7.

Recall that $\mathbb{P}(B_k = 1) = 1 - \mathbb{P}(B_k = 0) = \lambda_k$, $k = 1, 2, \dots$, and define $B_0 = \lambda_0 = 1$. With probability one, $\sum B_k < \infty$, since $\sum \lambda_k = \int_S C(x, x) dx < \infty$ as S is bounded and C is continuous. Consequently, with probability one, the random variable $M = \max\{k \geq 0 : B_k \neq 0\}$ is finite. For any integer $m > 0$, it is easily verified that B_0, \dots, B_{m-1} are independent of the event $\{M = m\}$. Therefore the strategy is first to generate a realization m of M , second independently generate realizations of the Bernoulli variables B_k for $k = 1, \dots, m - 1$ (if $m = 0$ we do nothing), and third set $B_m = 1$ and $B_k = 0$ for $k = m + 1, m + 2, \dots$. Simulation of these Bernoulli variables is of course easily done. For simulation of M , we use the inversion method described in Appendix D.

2.4.2 Simulation of determinantal projection point process

Suppose we have generated a realization of the Bernoulli variables B_k as described in Section 2.4.1 and we now want to generate a realization from $DPP_S(K)$ with K given by (2.7).

Let $n = \sum_{k=1}^{\infty} B_k$ denote the number of non-zero B_k 's with $k \geq 1$ (as foreshadowed in connection to (2.9), n can be considered as a realization of the count $N(S)$).

If $n = 0$, then $K = 0$ and a realization from $\text{DPP}_S(K)$ is simply equal to the empty point configuration. Assume that $n > 0$ and without loss of generality that

$$K(x, y) = \sum_{k=1}^n \phi_k(x) \overline{\phi_k(y)} = \mathbf{v}(y)^* \mathbf{v}(x) \quad (2.10)$$

where $\mathbf{v}(x) = (\phi_1(x), \dots, \phi_n(x))^T$, and where T and $*$ denote the transpose and conjugate transpose of a vector or a matrix. For n -dimensional complex column vectors such as $\mathbf{v}(x)$ and $\mathbf{v}(y)$, we consider their usual inner product $\langle \mathbf{v}(x), \mathbf{v}(y) \rangle = \mathbf{v}(y)^* \mathbf{v}(x)$.

Algorithm 1 Simulation of 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 1 **do**

sample X_i from the distribution with density

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

set $\mathbf{w}_i = \mathbf{v}(X_i) - \sum_{j=1}^{n-i} (\mathbf{e}_j^* \mathbf{v}(X_i)) \mathbf{e}_j$, $\mathbf{e}_{n-i+1} = \mathbf{w}_i/\|\mathbf{w}_i\|$

end for

return $\{X_1, \dots, X_n\}$

The following theorem is proved in Appendix E. It follows from the proof that with probability one, $p_i(x)$ is a density, where we are conditioning on (X_n, \dots, X_{i+1}) if $i < n$.

Theorem 2.8. *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 $\{X_1, \dots, X_n\}$ generated by Algorithm 1 is distributed as $\text{DPP}_S(K)$.*

To implement Algorithm 1 we need to sample from the densities p_i , $i = n, \dots, 1$. This may simply be done by rejection sampling with a uniform instrumental density and acceptance probability $p_i(x)/\sup_{y \in S} p_i(y)$. Appendixes E-F discuss rejection sampling for this and other choices of the instrumental distribution.

2.5 Densities

This section briefly discusses the density expression for $X_S \sim \text{DPP}(C; S)$ 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 one.

In general, when some eigenvalues λ_k are allowed to be one, the density of X_S is not available. But we can condition on the Bernoulli variables B_k from Theorem 2.7, or just condition on $K(x, y)$ for all $x, y \in S$, to obtain the conditional density. Note that the trace $\text{tr}_S(K) = \int_S K(x, x) dx = \sum_{k=1}^{\infty} B_k$ is almost surely finite.

Conditional on K , when $\text{tr}_S(K) = n > 0$, the ordered n -tuple of events of the determinantal projection point process X_S has density

$$p(x_1, \dots, x_n) = \det[K](x_1, \dots, x_n)/n!, \quad (x_1, \dots, x_n) \in S^n,$$

as verified in (E.2). Moreover, by Algorithm 1 and Theorem 2.8,

$$p_n(x) = K(x, x)/n, \quad x \in S,$$

is the density for an arbitrary selected event of X_S . This is in agreement with the simple fact that in the homogeneous case, i.e. when the intensity $K(x, x)$ is constant on S , any event of X_S is uniformly distributed on S .

The most interesting case occurs when $\lambda_k < 1$ for all $k = 1, 2, \dots$, which means that no B_k is almost surely one. Then the density of X_S exists and is specified in Theorem 2.9 below, where the following considerations and notation are used. If $P(N(S) = n) > 0$, then $P(N(S) = m) > 0$ for $m = 0, \dots, n$, cf. (2.9). Thus

$$P(N(S) = 0) = \prod_{k=1}^{\infty} (1 - \lambda_k)$$

is strictly positive, and we can define

$$D = -\log P(N(S) = 0) = -\sum_{k=1}^{\infty} \log(1 - \lambda_k). \quad (2.12)$$

Further, define $\tilde{C} : S \times S \rightarrow \mathbb{C}$ by

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

where

$$\tilde{\lambda}_k = \lambda_k / (1 - \lambda_k), \quad k = 1, 2, \dots$$

Let $|S| = \int_S dx$, and set $\det[\tilde{C}](x_1, \dots, x_n) = 1$ if $n = 0$. Then we have the following result, cf. Appendix G.

Theorem 2.9. *Assuming $\lambda_k < 1$, $k = 1, 2, \dots$, then X_S is absolutely continuous with respect to the homogeneous Poisson process on S with unit intensity, and has density*

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

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

Section 4 and Appendix K 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 yet not been studied in the literature from a statistical perspective. In the sequel 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), construct various classes of parametric models (Sections 3.3-3.4), and study to which extend these are models for repulsiveness (Section 3.5).

Throughout this section, $X \sim \text{DPP}(C)$ is assumed to be stationary, i.e. its distribution is invariant under translations, or equivalently, C is of the form

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

We also refer to C_0 as a covariance function. Note that $C_0(0)$, the variance corresponding to C , equals ρ , the intensity of X , cf. (2.3).

In light of Propositions A.1-A.2, as inhomogeneous DPPs can be obtained by transforming or thinning X , stationarity is not a very restrictive assumption. For example, by (A.1), if we transform X by a one-to-one continuous differentiable mapping T such that its Jacobian matrix is invertible, then $T(X)$ is a DPP with kernel

$$C_{\text{trans}}(x, y) = |J_{T^{-1}}(x)|^{1/2} C_0(T^{-1}(x) - T^{-1}(y)) |J_{T^{-1}}(y)|^{1/2}. \quad (3.2)$$

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 and anisotropic covariance function can be obtained from some stationary and isotropic covariance function using some rotation followed by some rescaling, see e.g. Goovaerts (1997).

Suppose C_0 is isotropic. Then C_0 is real, and the pair correlation function depends only on the distance between pairs of points, $g(x, y) = g_0(\|x - y\|)$, cf. (2.4). 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 relation

$$|R_0(r)| = \sqrt{1 - g_0(r)} \quad (3.3)$$

we can define a ‘range of correlation’, i.e. a distance $r_0 > 0$ such that $\sqrt{1 - g_0(r)}$ is considered to be negligible for $r \geq r_0$, as exemplified later in (3.14). For many specific models for isotropic covariance, including those studied in Section 3.3, R_0 is a decreasing function. By (3.3) g_0 is then an increasing function from zero to one, and informally such cases are called soft-core models (see e.g. Illian et al. (2008)).

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

3.2 A simple spectral condition for existence

The following Proposition 3.1 simplifies condition (C2). As it involves the spectral density for C_0 , we start by recalling this and related notions.

For any number $p > 0$ and Borel set $B \subseteq \mathbb{R}^d$, let $L^p(B)$ be the class of p -integrable functions $h : B \rightarrow \mathbb{C}$, i.e. $\int_B |h(x)|^p dx < \infty$. Denote \cdot the usual inner product in \mathbb{R}^d . For any Borel function $h : \mathbb{R}^d \rightarrow \mathbb{C}$, define the Fourier transform $\mathcal{F}(h)$ of h by

$$\mathcal{F}(h)(x) = \int h(y) e^{-2\pi i x \cdot y} dy, \quad x \in \mathbb{R}^d,$$

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

$$\mathcal{F}^{-1}(h)(x) = \int h(y) e^{2\pi i x \cdot y} dy, \quad x \in \mathbb{R}^d,$$

provided the integral exists. For instance, if $h \in L^1(\mathbb{R}^d)$, then $\mathcal{F}(h)$ and $\mathcal{F}^{-1}(h)$ are well-defined. Recall that $L^2(\mathbb{R}^d)$ is a Hilbert space with inner product

$$\langle h_1, h_2 \rangle = \int h_1(x) \overline{h_2(x)} dx$$

and the Fourier and inverse Fourier operators initially defined on $L^1(\mathbb{R}^d) \cap L^2(\mathbb{R}^d)$ extend by continuity to $\mathcal{F} : L^2(\mathbb{R}^d) \rightarrow L^2(\mathbb{R}^d)$ and $\mathcal{F}^{-1} : L^2(\mathbb{R}^d) \rightarrow L^2(\mathbb{R}^d)$. Furthermore, these are unitary operators that preserve the inner product, and \mathcal{F}^{-1} is the inverse of \mathcal{F} . See e.g. Stein and Weiss (1971).

By Khinchin's (or Bochner's) theorem, since C_0 is a continuous covariance function, a spectral distribution function F exists, i.e. F defines a finite measure so that

$$C_0(x) = \int e^{2\pi i x \cdot y} dF(y), \quad x \in \mathbb{R}^d.$$

If F is differentiable, then the derivative $\varphi(x) = dF(x)/dx$ is the spectral density for C_0 . In this case, φ is non-negative, $\varphi \in L^1(\mathbb{R}^d)$, and $C_0 = \mathcal{F}^{-1}(\varphi)$. On the other hand, if $C_0 \in L^1(\mathbb{R}^d)$ and C_0 is continuous (as assumed in this paper), then the spectral density necessarily exists (equivalently F is differentiable), $\varphi = \mathcal{F}(C_0)$, and φ is continuous and bounded. See e.g. pages 331-332 in Yaglom (1987).

Alternatively, if $C_0 \in L^2(\mathbb{R}^d)$ and C_0 is continuous, the spectral density φ also exists, since we can define $\varphi = \mathcal{F}(C_0)$ in $L^2(\mathbb{R}^d)$ as explained above. In this case, φ is non-negative, belongs to $L^1(\mathbb{R}^d) \cap L^2(\mathbb{R}^d)$, but is not necessarily continuous or bounded. Note that if $C_0 \in L^1(\mathbb{R}^d)$, then $C_0 \in L^2(\mathbb{R}^d)$ by continuity of C_0 .

The following is proved in Appendix H.

Proposition 3.1. *Under (C1) and (3.1), if $C_0 \in L^2(\mathbb{R}^d)$, then (C2) is equivalent to that*

$$\varphi \leq 1. \tag{3.4}$$

Assumption 3.2. *Henceforth, in addition to (C1), we assume that $C_0 \in L^2(\mathbb{R}^d)$ and that (3.4) holds.*

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

Corollary 3.3. *Under (3.1) the following two statements are equivalent.*

- (i) *There exists $\varphi \in L^1(\mathbb{R}^d)$ with $0 \leq \varphi \leq 1$ and $C_0 = \mathcal{F}^{-1}(\varphi)$.*
- (ii) *Conditions (C1) and (C2) hold and $C_0 \in L^2(\mathbb{R}^d)$.*

For later purposes, when considering a parametric model for C_0 with parameters ρ and θ , notice the following. For each fixed value of θ , $0 \leq \rho \leq \rho_{\max}$ where $\rho_{\max} = \rho_{\max}(\theta)$ may depend on θ and is determined by (3.4). As exemplified in Section 3.3, ρ_{\max} will be a decreasing function of the range of correlation (which only depends on θ). On the other hand, it may be more natural to determine the range of θ in terms of ρ and a given maximal range of correlation. Finally, in order to work with the density given in Theorem 2.9, we may require that $\varphi < 1$.

3.3 Examples of covariance function models

Numerous examples of stationary and isotropic covariance functions exist (see e.g. Gelfand et al., 2010), while examples of stationary and anisotropic covariance functions are discussed in De Iaco et al. (2003). This section starts by considering the simple example of the circular covariance function and continues with a brief discussion of the broad class of stationary isotropic covariance functions obtained by scaling in normal-variance mixture distributions, where a few specific examples of such models are considered in more detail. Section 3.4 discusses further examples based on a spectral approach.

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

$$\delta = \sup\{\|x\| : C_0(x) \neq 0\} \tag{3.5}$$

is finite are given in Wu (1995) and Gneiting (2002). Then, by Definition 2.1, X_A and X_B are independent DPPs if $A, B \subset \mathbb{R}^d$ are separated by a distance larger than δ . In this paper we only consider the circular covariance function to understand well the quality of our approximations in Section 4. For $d = 2$, the circular covariance function with finite range $\delta > 0$ is given by

$$C_0(x) = \rho \frac{2}{\pi} \left(\arccos(\|x\|/\delta) - \|x\|/\delta \sqrt{1 - (\|x\|/\delta)^2} \right), \quad \|x\| < \delta. \tag{3.6}$$

Note that $\pi\delta^2 C_0(x)/(4\rho)$ is the area of the intersection of two discs, each with diameter δ , and with distance $\|x\|$ between the centers. Since this area is equal to the autoconvolution of the indicator function of the disc with center at the origin and with diameter δ , the associated spectral density becomes

$$\varphi(x) = (\rho/\pi)(J_1(\pi\delta\|x\|)/\|x\|)^2$$

where J_1 is the Bessel function of the first kind with parameter $\nu = 1$. This spectral density has maximal value $\varphi(0) = \rho\pi\delta^2/4$, so by (3.4), a stationary and isotropic DPP with kernel (3.6) exists if and only if $0 \leq \rho \leq \rho_{\max}$, where $\rho_{\max} = 4/(\pi\delta^2)$. Therefore, we require

$$\rho\delta^2 \leq 4/\pi. \tag{3.7}$$

In the sequel we focus on more interesting classes of covariance functions. Let Z be a d -dimensional standard normally distributed random variable, and W be a strictly positive random variable with $E(W^{-d/2}) < \infty$, where Z and W are independent. Then $Y = \sqrt{W}Z$ follows a normal-variance mixture distribution, with density

$$h(x) = E [W^{-d/2} \exp(-\|x\|^2/(2W))] / (2\pi)^{d/2}, \quad x \in \mathbb{R}^d.$$

Note that $h(0) = \sup h$, and define

$$C_0(x) = \rho h(x)/h(0), \quad x \in \mathbb{R}^d.$$

The Fourier transform of C_0 is

$$\varphi(x) = \rho E [\exp(-2\pi^2\|x\|^2W)] / h(0), \quad x \in \mathbb{R}^d$$

which is positive, showing that C_0 is a stationary and isotropic covariance function. Note that φ is given by the Laplace transform of W . By (3.4), a stationary DPP with kernel C_0 exists if $0 \leq \rho \leq \rho_{\max}$, where ρ is the intensity and

$$\rho_{\max} = h(0) = E(W^{-d/2}) / (2\pi)^{d/2}.$$

Gneiting (1997) presents several examples of pairs h and $\mathcal{F}(h)$ in the one-dimensional case $d = 1$, and these examples can be generalized to the multivariate case. Here we restrict attention to the following three examples, where Y follows either a multivariate normal distribution or two special cases of the multivariate generalized hyperbolic distribution (Barndorff-Nielsen, 1977, 1978). We let $\Gamma(a, b)$ denote the Gamma-distribution with shape parameter $a > 0$ and scale parameter $b > 0$.

First, taking $\sqrt{2W} = \alpha$, where $\alpha > 0$ is a parameter, we obtain the Gaussian (or squared exponential) covariance function

$$C_0(x) = \rho \exp(-\|x/\alpha\|^2), \quad x \in \mathbb{R}^d, \quad (3.8)$$

and

$$\varphi(x) = \rho(\sqrt{\pi}\alpha)^d \exp(-\|\pi\alpha x\|^2), \quad x \in \mathbb{R}^d.$$

Hence

$$\rho_{\max} = (\sqrt{\pi}\alpha)^{-d} \quad (3.9)$$

is a decreasing function of α .

Second, suppose that $W \sim \Gamma(\nu + d/2, 2\alpha^2)$ where $\nu > 0$ and $\alpha > 0$. Then

$$h(x) = \frac{\|x/\alpha\|^\nu K_\nu(\|x/\alpha\|)}{2^{\nu+d-1}(\sqrt{\pi}\alpha)^d \Gamma(\nu + d/2)}, \quad x \in \mathbb{R}^d,$$

where K_ν is the modified Bessel function of the second kind (see Appendix J). Hence

$$C_0(x) = \rho \frac{2^{1-\nu}}{\Gamma(\nu)} \|x/\alpha\|^\nu K_\nu(\|x/\alpha\|), \quad x \in \mathbb{R}^d, \quad (3.10)$$

is the Whittle-Matérn covariance function, where for $\nu = 1/2$, $C_0(x) = \rho \exp(-\|x\|/\alpha)$ is the exponential covariance function. Moreover,

$$\varphi(x) = \rho \frac{\Gamma(\nu + d/2)}{\Gamma(\nu)} \frac{(2\sqrt{\pi}\alpha)^d}{(1 + \|2\pi\alpha x\|^2)^{\nu+d/2}}, \quad x \in \mathbb{R}^d,$$

so

$$\rho_{\max} = \frac{\Gamma(\nu)}{\Gamma(\nu + d/2)(2\sqrt{\pi}\alpha)^d} \quad (3.11)$$

is a decreasing function of ν as well as of α .

Third, suppose that $1/W \sim \Gamma(\nu, 2\alpha^{-2})$ where $\nu > 0$ and $\alpha > 0$. Then

$$h(x) = \frac{\Gamma(\nu + d/2)}{\Gamma(\nu)(\sqrt{\pi}\alpha)^d (1 + \|x/\alpha\|^2)^{\nu+d/2}}, \quad x \in \mathbb{R}^d,$$

is the density of a multivariate t -distribution, and

$$C_0(x) = \frac{\rho}{(1 + \|x/\alpha\|^2)^{\nu+d/2}}, \quad x \in \mathbb{R}^d, \quad (3.12)$$

is the generalized Cauchy covariance function. Furthermore,

$$\varphi(x) = \frac{\rho(\sqrt{\pi}\alpha)^d 2^{1-\nu}}{\Gamma(\nu + d/2)} \|2\pi\alpha x\|^\nu K_\nu(\|2\pi\alpha x\|), \quad x \in \mathbb{R}^d,$$

so

$$\rho_{\max} = \frac{\Gamma(\nu + d/2)}{\Gamma(\nu)(\sqrt{\pi}\alpha)^d} \quad (3.13)$$

is an increasing function of ν and a decreasing function of α .

For later use, notice that the Gaussian covariance function (3.8) with $\alpha = 1/\sqrt{\pi\rho}$ is the limit of both

- (i) the Whittle-Matérn covariance function (3.10) with $\alpha = 1/\sqrt{4\pi\nu\rho}$, and
- (ii) the Cauchy covariance function (3.12) with $\alpha = \sqrt{\nu/(\pi\rho)}$

as $\nu \rightarrow \infty$.

We refer to a DPP model with kernel (3.8), (3.10), or (3.12) as the Gaussian, Whittle-Matérn, or Cauchy model, respectively. In all three models, α is a scale parameter of C_0 , and for the Whittle-Matérn and Cauchy models, ν is a shape parameter of C_0 . Their isotropic pair correlation functions are as follows.

For the Gaussian model: $g_0(r) = 1 - \exp(-2(r/\alpha)^2)$, $r \geq 0$.

For the Whittle-Matérn model: $g_0(r) = 1 - [2^{1-\nu}(r/\alpha)^\nu K_\nu(r/\alpha)/\Gamma(\nu)]^2$, $r \geq 0$.

For the Cauchy model: $g_0(r) = 1 - [1 + (r/\alpha)^2]^{-2\nu-d}$, $r \geq 0$.

In each case, $g_0(r)$ is a strictly increasing function from zero to one.

In the sequel, let $d = 2$. For a given model as above, we choose the range of correlation r_0 such that $g_0(r_0) = 0.99$, whereby the isotropic correlation function given by (3.3) has absolute value 0.1. While it is straightforward to determine r_0 for the Gaussian and Cauchy model, r_0 is not expressible on closed form for the Whittle-Matérn model, and in this case we use the empirical result of Lindgren et al. (2011).

The ranges of correlation for the Gaussian, Whittle-Matérn and Cauchy models are then

$$r_0 = \alpha\sqrt{-\log(0.1)}, \quad r_0 = \alpha\sqrt{8\nu}, \quad r_0 = \alpha\sqrt{0.1^{-1/(\nu+1)} - 1}, \quad (3.14)$$

respectively. In each case, r_0 depends linearly on α , and when ν is fixed, ρ_{\max} decreases as r_0 increases, since ρ_{\max} is proportional to r_0^{-d} , cf. (3.9), (3.11), and (3.13). There is a similar trade-off between how large the intensity and the range of the circular covariance function can be, cf. (3.7).

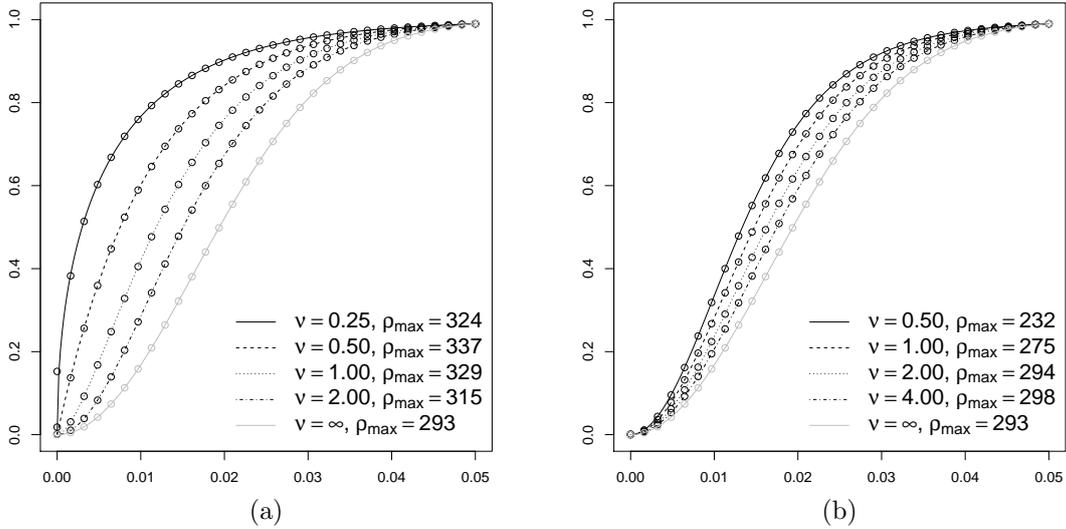


Figure 3: Isotropic pair correlation functions for (a) the Whittle-Matérn model and (b) the Cauchy model. Each black line corresponds to a different value of the shape parameter ν . The pair correlation function for the Gaussian model ($\nu = \infty$) is shown in gray in both plots. For each model, the scale parameter α is chosen such that the range of correlation is fixed at $r_0 = 0.05$, and the corresponding value of ρ_{\max} is reported in the legend. The circles show values of the approximate isotropic pair correlation function obtained by using the approximation C_{app} described in Section 4.

Figure 3 shows examples of the isotropic pair correlation functions with a fixed range of correlation. In particular the Whittle-Matérn DPPs have several different shapes of pair correlation functions and so they may constitute a quite flexible model class for soft-core-ness. From the figure it is also evident that the value of ρ_{\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.

Ripley's K -function (Ripley, 1976, 1977) is for $d = 2$ given by

$$K(r) = 2\pi \int_0^r t g_0(t) dt, \quad r \geq 0, \quad (3.15)$$

and we obtain the following.

$$\text{For the Gaussian model: } K(r) = \pi r^2 - \frac{\pi\alpha^2}{2} \left(1 - \exp\left(\frac{-2r^2}{\alpha^2}\right)\right).$$

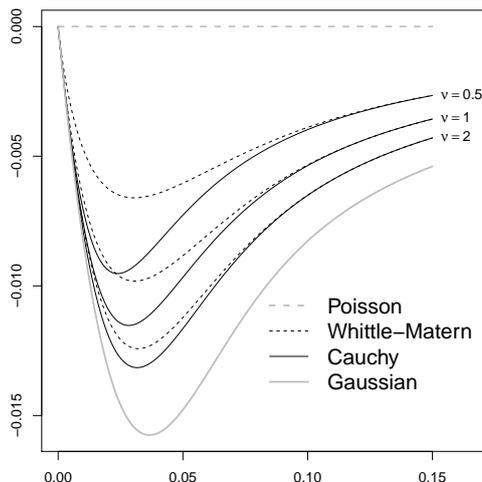


Figure 4: Plots of $L(r) - r$ vs. r for the Whittle-Matérn, Cauchy, and Gaussian model with $\alpha = \alpha_{\max}$ and $\rho = 100$. For the Whittle-Matérn and Cauchy models, $\nu \in \{0.5, 1, 2\}$. The horizontal line at zero is $L(r) - r$ for a stationary Poisson process.

For the Cauchy model: $K(r) = \pi r^2 - \frac{\pi \alpha^2}{2\nu + d - 1} \left(1 - \left(\frac{\alpha^2}{\alpha^2 + r^2} \right)^{2\nu + d - 1} \right)$.

For the Whittle-Matérn model: The integral in (3.15) has to be evaluated by numerical methods.

We consider the variance stabilizing transformation of the K -function, $L(r) = \sqrt{K(r)/\pi}$ (Besag, 1977b), and recall that $L(r) = r$ for a stationary Poisson process. Figure 4 shows $L(r) - r$ for seven different DPPs. Figures 3 and 4 illustrate the dependence between the degree of repulsiveness and ν , which will be discussed in more detail in Section 3.5.

3.4 Spectral approach

As an alternative of specifying a stationary covariance function C_0 , involving the need for checking positive semi-definiteness, we may simply specify an integrable function $\varphi : \mathbb{R}^d \rightarrow [0, 1]$, which becomes the spectral density, cf. Corollary 3.3. In fact knowledge about φ is all 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 the following we first describe a general method for constructing isotropic models via the spectral approach. Second, this method is used to construct a model class displaying a higher degree of repulsiveness (in a sense made precise in Section 3.5) than the Gaussian model which appears as a special case.

Let $f : [0, \infty) \rightarrow [0, \infty)$ be any Borel function such that $\sup f < \infty$ and $0 < c <$

∞ , where

$$c = \int_{\mathbb{R}^d} f(\|x\|) dx = \frac{d\pi^{d/2}}{\Gamma(d/2 + 1)} \int_0^\infty r^{d-1} f(r) dr. \quad (3.16)$$

Then we can define the spectral density of a stationary and isotropic DPP model as

$$\varphi(x) = \rho f(\|x\|)/c, \quad x \in \mathbb{R}^d, \quad (3.17)$$

where ρ is the intensity parameter. The model is well-defined whenever

$$\rho \leq \rho_{\max} = c / \sup f. \quad (3.18)$$

Below we give an example of a parametric model class for such functions f , where the integral in (3.16) and the supremum in (3.18) can be evaluated analytically.

Assume $Y \sim \Gamma(\gamma, \beta)$ and let f denote the density of $Y^{1/\nu}$, where $\gamma > 0$, $\beta > 0$, and $\nu > 0$ are parameters. Let $\alpha = \beta^{-1/\nu}$, then by (3.16) and (3.17),

$$c = \frac{d\pi^{d/2}\Gamma(\gamma + \frac{d+1}{\nu})}{\Gamma(d/2 + 1)\Gamma(\gamma)} \alpha^{1-d}$$

and

$$\varphi(x) = \rho \frac{\Gamma(d/2 + 1)\nu\alpha^d}{d\pi^{d/2}\Gamma(\gamma + \frac{d-1}{\nu})} \|\alpha x\|^{\gamma\nu-1} \exp(-\|\alpha x\|^\nu). \quad (3.19)$$

We have $\rho_{\max} = 0$ if $\gamma\nu < 1$, and

$$\rho_{\max} = \frac{c}{f((\gamma - 1/\nu)^{1/\nu})} = \frac{d\pi^{d/2}\alpha^{-d}\Gamma(\gamma + \frac{d-1}{\nu}) \exp(\gamma - 1/\nu)}{\Gamma(d/2 + 1)\nu(\gamma - 1/\nu)^{\gamma-1/\nu}} \quad \text{if } \gamma\nu \geq 1. \quad (3.20)$$

We call a DPP model with a spectral density of the form (3.19) a generalized gamma model. For $\gamma\nu > 1$, the spectral density (3.19) attains its maximum at a non-zero value, which makes it fundamentally different from the other models considered so far where the maximum is attained at zero.

In the remainder of this section, we consider the special case $\gamma = 1/\nu$, so

$$\varphi(x) = \rho \frac{\Gamma(d/2 + 1)\nu\alpha^d}{d\pi^{d/2}\Gamma(d/\nu)} \exp(-\|\alpha x\|^\nu). \quad (3.21)$$

We call a DPP model with a spectral density of the form (3.21) a power exponential spectral model. For $\nu = 2$, this is the Gaussian model of Section 3.3.

For the power exponential spectral model, for fixed ρ and ν , α has an upper limit α_{\max} given by $\alpha_{\max}^d = \Gamma(d/\nu + 1)r^{-d}$, where $r^d = \rho d \Gamma(d/2) / (2\pi^{d/2})$. For the choice $\alpha = \alpha_{\max}$ in (3.21), the spectral density of the power exponential spectral model becomes

$$\varphi(x) = \exp(-\|\Gamma(d/\nu + 1)^{1/d} x / r\|^\nu). \quad (3.22)$$

Note that this function tends to the indicator function over the set $\{\|x\| \leq r\}$ as ν tends to ∞ . This limiting case corresponds to a 'most repulsive possible stationary DPP' as discussed in detail in Section 3.5.

Figure 5 illustrates some properties of the power exponential spectral model when $\alpha = \alpha_{\max}$ and $\nu = 1, 2, 3, 5, 10, \infty$. Recall that $\nu = 2$ is the Gaussian model.

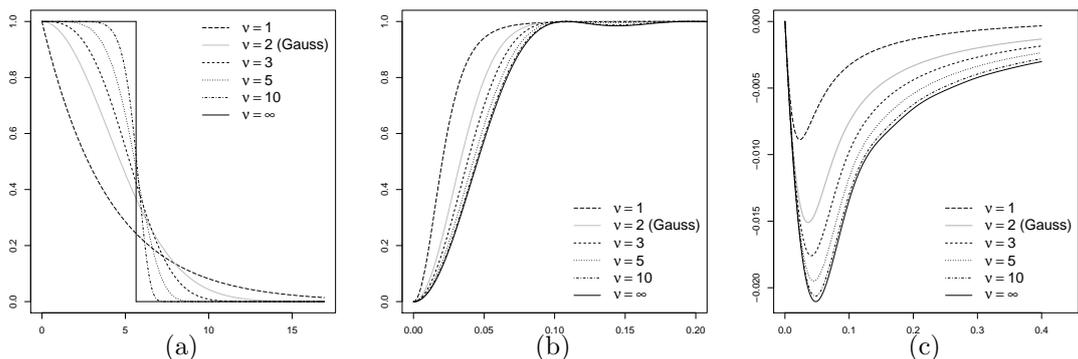


Figure 5: (a) Isotropic spectral densities, (b) approximate isotropic pair correlation functions, and (c) approximate $L(r) - r$ functions for power exponential spectral models with $\rho = 100$, $\nu = 1, 2, 3, 5, 10, \infty$, and $\alpha = \alpha_{\max}$ the maximal permissible value determined by (3.20).

Figure 5(a) shows the spectral density for these cases. Note that the spectral density approaches an indicator function as $\nu \rightarrow \infty$. Since we are not aware of a close form expression for $C_0 = \mathcal{F}^{-1}(\varphi)$ when φ is given by (3.22), we approximate C_0 by the periodic method discussed in Section 4, leading to approximating pair correlation functions shown in Figure 5(b). Figure 5(c) shows the corresponding approximations of $L(r) - r$ (analogously to Figure 4 in Section 3.3). Figure 5 is discussed in further detail in Section 3.5.

3.5 Quantifying and comparing repulsiveness

We now discuss different criteria to quantify repulsiveness. These criteria are used to compare the DPP models introduced in Sections 3.3-3.4.

Recall that $\rho K(r)$ is the conditional expectation of the number of further points of X in a ball of radius r centred at x given that X has a point at x . As a first criterion for repulsiveness, for two stationary DPPs with kernels C_1 and C_2 , common intensity ρ , and corresponding K -functions K_1 and K_2 , we may say that $\text{DPP}(C_1)$ exhibits stronger repulsiveness than $\text{DPP}(C_2)$ if $K_1(r) \leq K_2(r)$ for all $r \geq 0$. If the corresponding pair correlation functions g_1 and g_2 are isotropic, i.e. $g_i(x, y) = g_{i0}(\|x - y\|)$, $i = 1, 2$, then

$$K_1 \leq K_2 \quad \text{if and only if} \quad g_{10} \leq g_{20}. \quad (3.23)$$

In this sense, within each class of the Gaussian, Whittle-Matérn, and Cauchy models introduced in Section 3.3, when ν is fixed, the degree of repulsiveness increases as α increases. However, the increased degree of repulsiveness comes at the cost of a decreased maximal intensity cf. (3.9), (3.11), and (3.13). For fixed ρ and ν , the upper limit α_{\max} of α is given by

$$\alpha_{\max} = 1/\sqrt{\pi\rho}, \quad \alpha_{\max} = 1/\sqrt{4\pi\nu\rho}, \quad \alpha_{\max} = \sqrt{\nu/(\pi\rho)} \quad (3.24)$$

for the Gaussian, Whittle-Matérn, and Cauchy models, respectively. Letting $\alpha = \alpha_{\max}$, the degree of repulsiveness of both the Whittle-Matérn and the Cauchy models grows as ν grows, and the limit is the Gaussian case, cf. (i)-(ii) in Section 3.3.

On the other hand, the power exponential spectral model of Section 3.4 contains the Gaussian model as a special case when $\nu = 2$, and it provides examples of more and more repulsive DPPs as ν increases from zero to infinity, cf. Figure 5.

However, superposing the two plots in Figure 3 or considering Figure 4, the comparison between a Whittle-Matérn model and a Cauchy model is not always possible with our criterion for repulsiveness based on the K -functions.

Instead, for any stationary point process defined on \mathbb{R}^d , with distribution P , constant intensity $\rho > 0$, and pair correlation function $g(x, y) = g(x - y)$ (with a slight abuse of notation), we suggest

$$\mu = \rho \int [1 - g(x)] dx \quad (3.25)$$

as a rough measure for repulsiveness provided the integral exists. Denote o the origin of \mathbb{R}^d and note that the function $x \mapsto \rho g(o, x) = \rho g(x)$ is the intensity function for the reduced Palm distribution $P_o^!$ (intuitively, this is the conditional distribution of all remaining events when we condition on that o is an event, cf. Appendix C). Therefore, μ is the limit as $r \rightarrow \infty$ of the difference between the expected number of events within distance r from o under respectively P and $P_o^!$. For a stationary Poisson process, $\mu = 0$. For any stationary point process, we always have $\mu \leq 1$ (see e.g. (2.5) in Kuna et al. (2007)). When $g \leq 1$ (as in the case of a DPP), we clearly have $\mu \geq 0$, so that $0 \leq \mu \leq 1$.

Especially, for a stationary DPP,

$$\mu = \rho \int [1 - g(x)] dx = \frac{1}{\rho} \int |C_0(x)|^2 dx = \frac{1}{\rho} \int |\varphi(x)|^2 dx$$

where the second equality follows from (2.3) and (2.4), and the last equality follows from Parseval's identity. Using an obvious notation, we say that $\text{DPP}(C_1)$ is more repulsive than $\text{DPP}(C_2)$ if $\rho_1 = \rho_2$ and $\mu_1 \geq \mu_2$. In the isotropic case, this is in agreement with our former definition of repulsiveness: if $\rho_1 = \rho_2$, then $K_1 \leq K_2$ implies that $\mu_1 \geq \mu_2$, cf. (3.23).

A stationary DPP with intensity ρ and a maximal value of μ can be specified as follows. Since $0 \leq \varphi(x)^2 \leq \varphi(x) \leq 1$, we have $\mu = 1$ if and only if $\int \varphi(x)^2 dx = \int \varphi(x) dx = \rho$. So μ is maximal if φ is an indicator function with support on a Borel subset of \mathbb{R}^d of volume ρ . An obvious choice is

$$\varphi(x) = \begin{cases} 1 & \text{if } \|x\| \leq r \\ 0 & \text{otherwise} \end{cases} \quad (3.26)$$

where $r^d = \rho d \Gamma(d/2) / (2\pi^{d/2})$. For $d = 1$, C_0 is then proportional to a sinc function:

$$C_0(x) = \sin(\pi \rho x) / (\pi x) \quad \text{if } d = 1. \quad (3.27)$$

For $d = 2$, C_0 is then proportional to a 'jinc-like' function:

$$C_0(x) = \sqrt{\rho/\pi} J_1(2\sqrt{\pi\rho}\|x\|) / \|x\| \quad \text{if } d = 2. \quad (3.28)$$

As already noticed in Section 3.4, the indicator function (3.26) corresponds to the limit of (3.22) when ν tends to infinity. Thus the power exponential spectral model

contains a 'most repulsive possible stationary DPP' as a limiting case. Figure 5 illustrates how this limiting case is approached as ν increases. Notice in particular the slightly oscillating nature of g in Figure 5(b) for $\nu > 2$, indicating a stronger degree of repulsiveness than in a soft-core model. For Gibbs hard-core point processes, oscillation in the pair correlation function is also seen, but at the hard-core distance, the pair correlation function jumps from zero to a value larger than one (see e.g. Illian et al. (2008)).

4 Approximations

Let again $X \sim \text{DPP}(C)$ be stationary so that $C(x, y) = C_0(x - y)$, cf. (3.1). It would have been desirable if we could compute explicitly the spectral representation (2.5) for a given parametric family of C_0 and for at least some cases of compact sets S (e.g. closed rectangles). Unfortunately, analytic expressions for such representations are only known in a few simple cases (see e.g. Macchi (1975)), which we believe are insufficient to describe the interaction structure in real spatial point process datasets. Numerical approximations of the eigenfunctions and eigenvalues can be obtained for a given covariance function C_0 . However, in the simulation algorithm of Section 2.4 we may need to evaluate the eigenfunctions at several different locations to generate each point of the simulation, and the need for numerical approximation at each step can be computationally costly. On the other hand, the Fourier approximation (4.6) given below is very easy to apply. This requires that the spectral density associated to C_0 is available, which is the case for the examples given in Sections 3.3-3.4.

Throughout this section, $S = [-1/2, 1/2]^d$, $S/2 = [-1/4, 1/4]^d$, $2S = [-1, 1]^d$, and we discuss approximations of the kernel C and the DPP X restricted to S or sometimes to $S/2$ or a more general rectangular set depending on the circumstances. Section 4.1 concentrates on approximations of the kernel, Sections 4.2-4.3 on how to simulate an approximation of the DPP, and Section 4.4 on how to approximate the density of the DPP. We consider the orthonormal Fourier basis of $L^2(S)$ given by

$$\phi_k(x) = e^{2\pi i k \cdot x}, \quad k \in \mathbb{Z}^d, \quad x \in S, \quad (4.1)$$

where \mathbb{Z} denotes the set of integers. For $u \in S$, the Fourier expansion of $C_0(u)$ is

$$C_0(u) = \sum_{k \in \mathbb{Z}^d} \alpha_k e^{2\pi i k \cdot u}$$

where

$$\alpha_k = \int_S C_0(t) e^{-2\pi i k \cdot t} dt. \quad (4.2)$$

Note that for $(x, y) \in S \times S$,

$$C(x, y) = \sum_{k \in \mathbb{Z}^d} \alpha_k e^{2\pi i k \cdot (x-y)} = \sum_{k \in \mathbb{Z}^d} \alpha_k \phi_k(x) \overline{\phi_k(y)} \quad \text{if } x - y \in S$$

but since this expansion is not in general true when $x - y \notin S$, it is not a spectral representation for C restricted to $S \times S$.

4.1 Approximation of the kernel C

Substituting the finite integral in (4.2) by the infinite integral

$$\varphi(k) = \int C_0(t) e^{-2\pi i k \cdot t} dt \quad (4.3)$$

leads us to approximate C_0 on S by

$$C_0(u) \approx C_{\text{app},0}(u), \quad u \in S, \quad (4.4)$$

where

$$C_{\text{app},0}(u) = \sum_{k \in \mathbb{Z}^d} \varphi(k) e^{2\pi i k \cdot u}, \quad u \in S.$$

Since $x - y \in S$ if $x, y \in S/2$, we then obtain the approximation

$$C(x, y) \approx C_{\text{app}}(x, y), \quad x, y \in S/2, \quad (4.5)$$

where $C_{\text{app}}(x, y) = C_{\text{app},0}(x - y)$, $x, y \in S/2$.

Comparing (4.2) and (4.3) we see that the error of the approximations (4.4) and (4.5) is expected to be small if $C_0(t) \approx 0$ for $t \in \mathbb{R}^d \setminus S$. In particular, for covariance functions with finite range $\delta < 1/2$ (see (3.5)), $C_0(t) = 0$ for $t \in \mathbb{R}^d \setminus S$, and so $C_0(u) = C_{\text{app},0}(u)$ for $u \in S$, i.e. the approximations (4.4) and (4.5) are then exact. For instance, considering the circular covariance function (3.6) and the existence condition (3.7), we have $\delta < 1/2$ if $\rho > 16/\pi$, which indeed is not a restrictive requirement in practice.

Appendix J studies the accuracy of the approximation $C_0(u) \approx C_{\text{app},0}(u)$, $u \in S$, for the Whittle-Matérn model introduced in Section 3.3, and show that the error is small provided the intensity ρ is not too small. Furthermore, Figure 3 in Section 3.3 indicates that the approximation is accurate for the examples in the figure as the approximate pair correlation functions marked by circles in the plot are very close to the true curves.

For later purpose, we consider the periodic kernel defined on $S \times S$ as

$$C^{\text{per}}(x, y) = C_0^{\text{per}}(x - y), \quad x, y \in S,$$

where C_0^{per} is the periodic extension of C_0 from S to $2S$ defined by

$$C_0^{\text{per}}(u) = \sum_{k \in \mathbb{Z}^d} \alpha_k e^{2\pi i k \cdot u}, \quad u \in 2S.$$

Evaluating $C^{\text{per}}(x, y)$ for $x, y \in S$ corresponds to wrapping $[-1/2, 1/2]^d$ on a torus and evaluating $C(x^{(t)}, y^{(t)})$, where $x^{(t)}$ and $y^{(t)}$ are the points on the torus corresponding to x and y .

Finally, following (4.4), we use the approximation $C_0^{\text{per}}(u) \approx C_{\text{app},0}^{\text{per}}(u)$, $u \in 2S$, where $C_{\text{app},0}^{\text{per}}(u) = \sum_{k \in \mathbb{Z}^d} \varphi(k) e^{2\pi i k \cdot u}$, $u \in 2S$, which leads to the approximation of C^{per} on $S \times S$: $C^{\text{per}}(x, y) \approx C_{\text{app}}^{\text{per}}(x, y)$ where

$$C_{\text{app}}^{\text{per}}(x, y) = \sum_{k \in \mathbb{Z}^d} \varphi(k) \phi_k(x) \overline{\phi_k(y)}, \quad x, y \in S. \quad (4.6)$$

Note that for $x, y \in S/2$, $C^{\text{per}}(x, y) = C(x, y)$ and $C_{\text{app}}^{\text{per}}(x, y) = C_{\text{app}}(x, y)$.

4.2 The border method for simulation

Consider $X^{\text{per}} \sim \text{DPP}_S(C_{\text{app}}^{\text{per}})$, which is well-defined since $\varphi \leq 1$. We can think of X^{per} as a DPP on the torus with a kernel approximately corresponding to C on the torus. Furthermore, we can approximate $X_{S/2} \sim \text{DPP}(C; S/2)$ by $X_{S/2}^{\text{app}} = X^{\text{per}} \cap S/2 \sim \text{DPP}_S(C_{\text{app}}^{\text{per}}; S/2)$. Thus to approximately simulate $X_{S/2}$ we need to be able to simulate X^{per} , which is straightforward since (4.6) is of the form required for the simulation algorithm of Section 2.4. Recall that for finite range covariance functions with $\delta < 1/2$ (see (3.5)), the simulation is exact (or perfect) as $X_{S/2}$ and $X_{S/2}^{\text{app}}$ are identically distributed.

More generally suppose we want to simulate X_R where $R \subset \mathbb{R}^d$ is a rectangular set. We then define an affine transformation $T(x) = Ax + b$ such that $T(R) = S/2$. Then $Y = T(X)$ is a stationary DPP, with kernel given by (3.2) and spectral density $\varphi_Y(x) = \varphi(A^T x)$. Let Y^{per} be the DPP on S with kernel (4.6) where φ is replaced by φ_Y . Then we simulate Y^{per} and return $T^{-1}(Y^{\text{per}} \cap S/2)$ as an approximate simulation of X_R . We refer to this simulation procedure as the border method for simulating X_R .

4.3 The periodic method for simulation

From our practical experience it appears that $\text{DPP}_S(C_{\text{app}}^{\text{per}})$ is also a good approximation of $\text{DPP}(C; S)$, which may be harder to understand from a purely mathematical point of view. Intuitively, this is due to the fact that the periodic behaviour of $C_{\text{app},0}^{\text{per}}$ mimics the influence of points outside S . To illustrate this, Figure 6(a) shows the acceptance probability for a uniformly distributed proposal (used for rejection sampling when simulating from one of the densities p_i , see Remark E.3) when X^{per} is simulated by the algorithm in Section 2.4. The qualitative behavior of the process in S and in the interior region $S/2$ are similar e.g. in the sense that there are regions at the borders where the acceptance probability is low. For the process on $S/2$, this is due to the influence of points outside $S/2$. For the process on S , this is created artificially by points at the opposite border.

This approximation gives us an alternative way of approximately simulating X_R where $R \subset \mathbb{R}^d$ is a rectangular set as follows. We simply redefine the affine transformation above such that $T(R) = S$. Then $Y = T(X)$ is again a stationary DPP with spectral density $\varphi_Y(x) = \varphi(A^T x)$, and $T^{-1}(Y_S^{\text{per}})$ is an approximate simulation of X_R . We call this the periodic method for simulating X_R .

The advantage of the periodic method is that we on average only need to generate $\rho|R|$ points, whereas the border method requires us to generate $4\rho|R|$ points on average. To increase the efficiency of the border method we could of course use a modified affine transformation such that $T(R) = S'$ with $S/2 \subset S' \subset S$, but we will not go into the details of this, since it is our experience that the periodic method works very well. In particular we have compared the two methods for simulating DPPs with kernels given by circular covariance functions. In this case the border method involves no approximation and comparison of plots of the empirical distribution of various summary statistics revealed almost no difference between the two methods (these plots are omitted to save space).

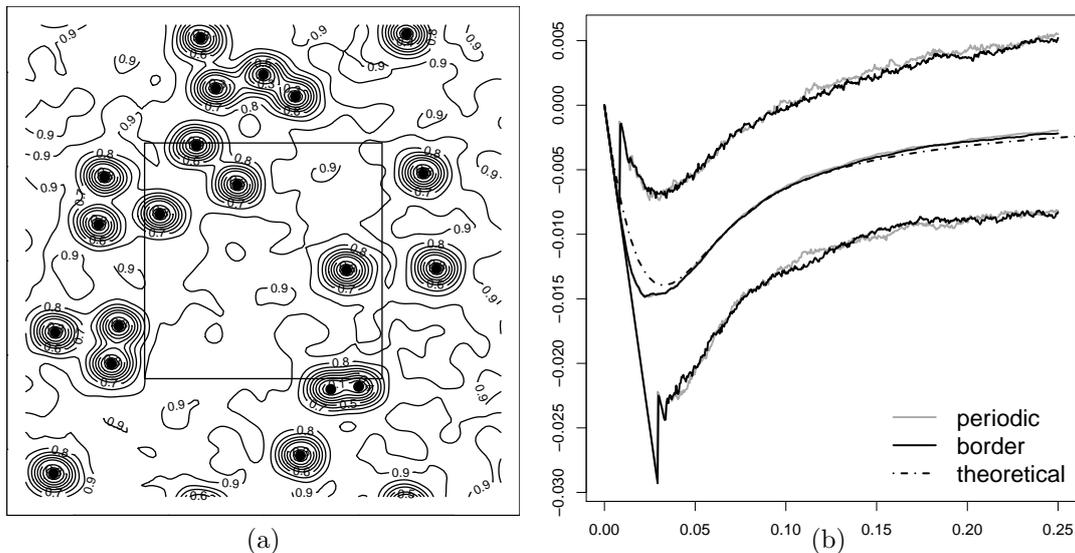


Figure 6: (a) Acceptance probability for a uniformly distributed proposal at an intermediate step of Algorithm 1 (Section 2.4) when simulating a realization of X^{per} on $S = [-1/2, 1/2]^d$. The interior box is the region $S/2 = [-1/4, 1/4]^d$. The black points represent previously generated points, and the acceptance probability is zero at these points. (b) Empirical means and 2.5% and 97.5% pointwise quantiles of $L(r) - r$ using either the periodic method (gray lines) or the border method (black lines), and based on 1000 realizations of a Gaussian model with $\rho = 100$ and $\alpha = 0.05$. The dashed line corresponds to the theoretical $L(r) - r$ function for this Gaussian model.

For a Gaussian covariance function, Figure 6(b) shows empirical means and 2.5% and 97.5% pointwise quantiles of $L(r) - r$ using either the periodic method (gray lines) or the border method (black lines), and based on 1000 realizations of a Gaussian model with $\rho = 100$ and $\alpha = 0.05$. The corresponding curves for the two methods are in close agreement, which suggests that the two methods generate realizations of nearly the same DPPs. This was also concluded when considering other covariance functions and functional summary statistics (plots not shown here). In Figure 6(b) the empirical means of $L(r) - r$ are close to the theoretical $L(r) - r$ function for the Gaussian model, indicating that the two approximations of the Gaussian model are appropriate.

The computational efficiency of the periodic method makes it our preferred method of simulation. The 1000 realizations used in Figure 6(b) were generated in approximately three minutes on a laptop with a dual core processor.

4.4 Approximation of the density f

First, consider the density f for X_S as specified in Theorem 2.9 (so we assume $\varphi < 1$). We use the approximation $f \approx f^{\text{per}}$, where f^{per} denotes the density of X^{per} . Letting

$$\tilde{\varphi}(u) = \varphi(u)/(1 - \varphi(u)), \quad u \in S, \quad (4.7)$$

$$\tilde{C}_{\text{app}}^{\text{per}}(x, y) = \tilde{C}_{\text{app},0}^{\text{per}}(x - y) = \sum_{k \in \mathbb{Z}^d} \tilde{\varphi}(k) e^{2\pi i k \cdot (x - y)}, \quad x, y \in S, \quad (4.8)$$

and

$$D_{\text{app}}^{\text{per}} = \sum_{k \in \mathbb{Z}^d} \log(1 + \tilde{\varphi}(k)) \quad (4.9)$$

we have

$$f^{\text{per}}(\{x_1, \dots, x_n\}) = \exp(|S| - D_{\text{app}}^{\text{per}}) \det[\tilde{C}_{\text{app}}^{\text{per}}](x_1, \dots, x_n), \quad \{x_1, \dots, x_n\} \subset S. \quad (4.10)$$

This density can be approximated in practice by truncating the infinite sums defining $\tilde{C}_{\text{app}}^{\text{per}}$ and $D_{\text{app}}^{\text{per}}$. Furthermore, the speed of calculation can be increased by using a fast Fourier transform (FFT) to evaluate $\tilde{C}_{\text{app},0}^{\text{per}}$. The details of the truncation and use of FFT are given in Section 5.1.

Second, consider the density of X_R , where $R \subset \mathbb{R}^d$ is rectangular. Then we use the affine transformation from above with $T(R) = S$ to define $Y = T(X)$. If f_Y^{per} denotes the approximate density of Y as specified by the right hand side of (4.10), we can approximate the density of X_R by

$$f^{\text{per}}(\{x_1, \dots, x_n\}) = |R|^{-n} \exp(|R| - |S|) f_Y^{\text{per}}(T(\{x_1, \dots, x_n\})), \quad \{x_1, \dots, x_n\} \subset R.$$

We call f^{per} the periodic approximation of f . The simulation study in Section 5 shows that likelihood inference based on f^{per} works well in practice for the examples in this paper. Appendix K introduces a convolution approximation of the density which in some cases may be computationally faster to evaluate. However, as discussed in Appendix K, this approximation appears to be poor in some situations and in general we prefer the periodic approximation.

5 Inference procedures

In this section, we discuss how to estimate parameters of parametric DPP models and how to do model comparison. Section 5.1 focuses on maximum likelihood based inference, Section 5.2 considers alternative ways of performing inference, and Section 5.3 discusses a simulation study of the approaches of Sections 5.1-5.2. Examples of the estimation and model comparison procedures when modelling real datasets are given in Section 6 where we also discuss model checking. Furthermore, in Sections 5.1.1 and 5.3, we discuss the commonly used non-parametric intensity estimate $\hat{\rho} = n/|S|$ in comparison to the maximum likelihood estimate (MLE) of ρ .

For simplicity we assume that X is a stationary DPP, so the kernel C_0 could e.g. be given by one of the parametric covariance models described in Sections 3.3-3.4. The inhomogeneous case will be discussed in Section 6.5. Further, C_0 is assumed to be parametrized by the intensity ρ and an additional parameter θ for the correlation function $R_{0,\theta}$ corresponding to C_0 , i.e. $C_0(x) = \rho R_{0,\theta}(x)$. Recall that ρ introduces a bound on the parameter space for θ , since (using an obvious notation) $\rho_{\max}(\theta) \geq \rho$, cf. Sections 3.3-3.4. Given ρ , we denote Θ_ρ the parameter space of θ . Furthermore, we let $\mathbf{x} = \{x_1, \dots, x_n\}$ denote a realization of X_S , where S is a bounded rectangular region and we refer to \mathbf{x} as the data.

5.1 Maximum likelihood based inference

In order to perform maximum likelihood based inference, we approximate the likelihood function with respect to θ by the density f^{per} of Section 4, where we use truncation and FFT to evaluate f^{per} . This is described in more detail in the following.

For a given integer $N > 0$ (the choice of N is discussed below), let $\mathbb{Z}_N = \{-N, -N+1, \dots, N-1, N\}$ and define truncated versions of (4.8) and (4.9) by

$$D_N = \sum_{k \in \mathbb{Z}_N^d} \log(1 + \tilde{\varphi}(k)) \quad (5.1)$$

and

$$\tilde{C}_N(u) = \sum_{k \in \mathbb{Z}_N^d} \tilde{\varphi}(k) e^{2\pi i k \cdot u}, \quad u \in \mathbb{R}^d. \quad (5.2)$$

Here $\tilde{\varphi}$ is given by (4.7) and we suppress in the notation that $\tilde{\varphi}$ depends on (ρ, θ) where ρ is replaced by $\hat{\rho} = n/|S|$. The (approximate) MLE is the value of θ which maximizes the (approximate) log-likelihood

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

where $[\tilde{C}_N](x_1, \dots, x_n)$ is the $n \times n$ matrix with (i, j) 'th element $\tilde{C}_N(x_i - x_j)$. If θ is one dimensional, the maximum of $\ell_N(\theta)$ can be determined by a simple search algorithm, otherwise the simplex algorithm by Nelder and Mead (1965) can be used. Note that these methods do not require explicit knowledge of the derivatives of $\ell_N(\theta)$.

While it is feasible to evaluate (5.1) for large values of N , the evaluation of (5.2) is more problematic since it needs to be carried out for every pair of points in \mathbf{x} . For

moderate N (few hundreds) direct calculation of (5.2) can be used. In this case and if C_0 is a real function, we can exploit the fact that $\tilde{\varphi}$ is an even function such that all imaginary parts in (5.2) cancel. This allows to reduce the number of terms in the sum by a factor 2^d , which speeds up calculations considerably when evaluating the periodic density. For large N (hundreds or thousands) we use the FFT of $\tilde{\varphi}$. The FFT yields values of \tilde{C}_N at a discrete grid of values and we simply approximate $\tilde{C}_N(x_i - x_j)$ by the value at the closest grid point.

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 tends to infinity. Hence, for any value of θ , one criterion for choosing N may be to require e.g. $S_N > 0.99\hat{\rho}$. However, this may be insufficient as N also determines the grid resolution when FFT is used, and a high resolution may be required to obtain a good approximation of the likelihood. Therefore, we use increasing values of N until the approximate MLE stabilizes.

When comparing several different models fitted to the same dataset (e.g. Gauss, Whittle-Matérn, and Cauchy), we prefer the model with the largest value of $\ell_N(\theta)$. The comparison of $\ell_N(\theta)$ between different model classes is valid, since the dominating measure is the same for all the models.

5.1.1 MLE for the intensity

Rather than fixing the estimate of the intensity to $\hat{\rho} = n/|S|$, we may estimate both ρ and θ by maximum likelihood. This has been done for the simulated Gaussian model given in Section 5.3, where we observed that the MLE of ρ is very close to $n/|S|$. This has further been done for each DPP model fitted to the real datasets in Section 6, where the largest relative difference between the non-parametric estimate and the MLE of ρ was 4%.

When ρ is not too close to ρ_{\max} , the fact that the MLE appears to be close to $n/|S|$ may be understood in the following way. By applying the convolution approximation in Appendix K.1, rough approximations $\tilde{C}(x, y) \approx C_0(x - y)$ and $D \approx |S|\rho$ are obtained by considering only the first terms in (K.1) and (K.3). Hence a rough approximation of the log-likelihood is

$$\ell(\rho, \theta; \{x_1, \dots, x_n\}) \approx -|S|\rho + n \log \rho + \log \det[C^\dagger](x_1, \dots, x_n)$$

where $C^\dagger(x, y) = C_0(x - y)/\rho$ depends only on θ and not on ρ . The maximum point of this approximate log-likelihood has $\rho = n/|S|$.

On the other hand, for very repulsive DPPs, the number of points has a small variance, and so we may expect the intensity to be close to the observed $n/|S|$. In particular, the most repulsive DPP is a determinantal projection process and the intensity is then non-random.

5.2 Alternative approaches for inference

Given a parametric DPP model there are several feasible approaches for inference which are not based on maximum likelihood. For example, parameter estimation can

be based on composite likelihood, Palm likelihood, generalized estimating equations, or minimum contrast methods. See Møller and Waagepetersen (2007), Prokešová and Jensen (2013), and the references therein. Here we only briefly recall how the minimum contrast estimate (MCE) (Diggle and Gratton, 1984) is calculated.

Let $s(r; \theta)$, $r \geq 0$, denote a functional summary statistic for which we have a closed form expression, where $\theta \in \Theta_{\hat{\rho}}$ and $\hat{\rho} = n/|S|$. In our examples, this will be either the pair correlation function g or the K -function. Further, let $\hat{s}(r)$ be a non-parametric estimate of s based on the data \mathbf{x} . The MCE based on the functional summary statistic s is the value of θ which minimizes

$$D(\theta) = \int_{r_l}^{r_u} |\hat{s}(r)^q - s(r; \theta)^q|^p dr$$

where the limits of integration $r_l < r_u$ and the exponents $p > 0$ and $q > 0$ are user-specified parameters. Following the recommendations in Diggle (2003), we let $q = 1/2$, $p = 2$, and r_u be one quarter of the minimal side length of S . It is customary to use $r_l = 0$ and we do this when the MCE is based on the K -function. However, when the MCE is based on g , we let r_l be one percent of the minimal side length of S , since in our simulation experiments it turned out to be a better choice. To minimize $D(\theta)$ we use the same method as was used for maximizing $\ell_N(\theta)$ in Section 5.1, which avoids the use of derivatives of $D(\theta)$.

Finally, when several different models are fitted to the same dataset, the one with minimal value of $D(\theta)$ is preferred.

5.3 Simulation study

We have generated 500 realizations in the unit square of the following five models: Gaussian, Whittle-Matérn with $\nu = 0.5$, Whittle-Matérn with $\nu = 1$, Cauchy with $\nu = 0.5$, and Cauchy with $\nu = 1$. For all models, $\rho = 200$ and $\alpha = \alpha_{\max}/2$, where α_{\max} is given by (3.24). In our experience it is difficult to identify the parameters ν and α simultaneously, which is a well-known issue for the Whittle-Matérn covariance function (see e.g. Lindgren et al., 2011). Here we consider ν known such that the remaining parameter to estimate is one dimensional, i.e. $\theta = \alpha$.

Table 1 provides the empirical means and standard deviations of the MCE based on K , the MCE based on g , and the MLE, where for each model, the MLE is calculated for several different values of N . In general, we see that as long as the truncation is sufficiently large the MLE outperforms the MCE since the former has smaller biases and smaller standard deviations.

The quality of the likelihood approximation is closely related to the decay rate of the spectral density of the model, or equivalently to the rate of convergence of S_N . Figure 7 shows S_N for different values of N for each of the five models. It is clear that the two Whittle-Matérn models approach the theoretical limit $\rho = 200$ at a slower rate than the other models, and this makes the likelihood approximation inaccurate for small N leading to bias in the estimates shown in Table 1.

For the Gaussian model above, we have tried to include ρ as a freely varying parameter when maximizing the likelihood with $N = 2048$. For each realization the MLE of ρ was very close to the non-parametric estimate $n/|W|$, and the largest

Table 1: Empirical means and standard deviations (in parentheses) of parameter estimates based on 500 simulated datasets for each of 5 different models with intensity $\rho = 200$. Model 1: Gauss; Model 2: Whittle-Matérn ($\nu = 0.5$); Model 3: Whittle-Matérn ($\nu = 1$); Model 4: Cauchy ($\nu = 0.5$); Model 5: Cauchy ($\nu = 1$). The columns from left to right are: The true value of α , MCE based on the K -function, MCE based on g , MLE with $N = 256$, MLE with $N = 512$, MLE with $N = 1024$, and MLE with $N = 2048$. All entries are multiplied by 100 to make the table more compact.

α	K	g	MLE256	MLE512	MLE1024	MLE2048	
1	2.00	2.05 (0.58)	1.99 (0.51)	1.42 (0.25)	2.01 (0.43)	2.01 (0.43)	2.01 (0.43)
2	1.40	1.59 (0.88)	1.48 (0.92)	1.77 (0.11)	1.62 (0.56)	1.55 (0.63)	1.52 (0.67)
3	1.00	1.02 (0.46)	0.95 (0.54)	0.97 (0.18)	1.00 (0.36)	1.00 (0.37)	1.00 (0.37)
4	1.40	1.48 (0.68)	1.30 (0.87)	1.39 (0.23)	1.37 (0.54)	1.38 (0.54)	1.38 (0.55)
5	2.00	2.07 (0.83)	1.91 (0.97)	1.69 (0.29)	2.01 (0.61)	2.01 (0.62)	2.02 (0.61)

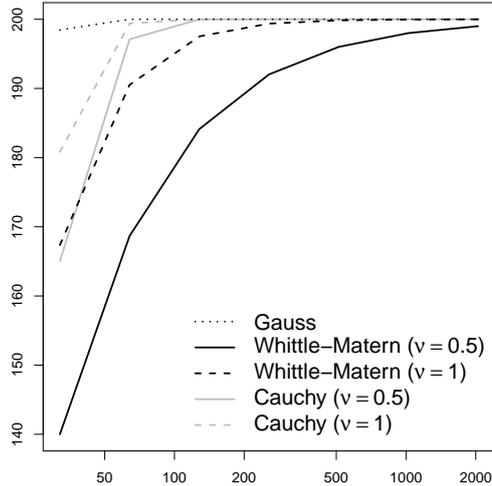


Figure 7: S_N as a function of N .

relative difference between the two estimates of ρ for the 500 realizations was 0.3%. This fits well with Section 5.1.1 where heuristic arguments suggest that the MLE for ρ is close to $n/|W|$.

6 Real data examples

In this section, we analyse the six datasets presented in Figure 1 using the parametric DPP models of Sections 3.3-3.4, namely the Gaussian, Whittle-Matérn, Cauchy, and power exponential spectral models—for short we refer to these as the four parametric classes of DPPs. Except for Section 6.5, we use the computationally simple intensity estimate $\hat{\rho} = n/|S|$ and the approximate MLE for $\theta \in \Theta_{\hat{\rho}}$, cf. Section 5.

6.1 Spanish towns dataset

The Spanish towns dataset displayed in Figure 1(a) was first analysed in Glass and Tobler (1971). In a subsequent analysis Ripley (1988) recommended using a Strauss hard-core model, with four parameters: A hard-core distance h , an interaction distance R , an abundance parameter β and an interaction parameter γ . The MLE for h is the minimal observed distance, r_{\min} , but more commonly $\frac{nr_{\min}}{n+1}$ is used. Estimation of R is typically based on an ad-hoc method such as maximum profile pseudo-likelihood (see e.g. Møller and Waagepetersen (2004)). Conditionally on h and R , the parameters β and γ can e.g. be estimated by the maximum pseudo-likelihood 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 using the approximate likelihood method of Huang and Ogata (1999) available in `spatstat`. The estimates are $\hat{\beta} = 0.11$ and $\hat{\gamma} = 0.85$.

Figure 8 is used to assess the goodness of fit for the Strauss hard-core model. It shows non-parametric estimates of $L(r) - r$, the nearest neighbour distribution function $G(r)$, the empty space function $F(r)$, and $J(r) = (1 - G(r))/(1 - F(r))$, together with 2.5% and 97.5% pointwise quantiles (gray lines) for these summary statistics based on 400 simulations of the fitted Strauss hard-core model (for definitions of F and G , see e.g. Møller and Waagepetersen (2004)). Overall the model appears to provide an acceptable fit, but the characteristic cusp of the envelopes of $L(r) - r$ at $r = \hat{R} = 3.5$ seem to be a somewhat artificial model effect that the dataset does not exhibit.

As an alternative to the Strauss hard-core model, we now consider the four parametric classes of DPP models. The intensity estimate is $\hat{\rho} = n/|W| = 0.043$, 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 400 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 dataset. The resulting p -value is 0.03 and we reject the Gaussian model in favor of the Whittle-Matérn model. Again Figure 8 is used to assess the goodness of fit which appears to be quite good as none of the non-parametric estimates exit the 95% pointwise envelopes obtained from simulations under the fitted model.

In summary the Whittle-Matérn model both has less 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.

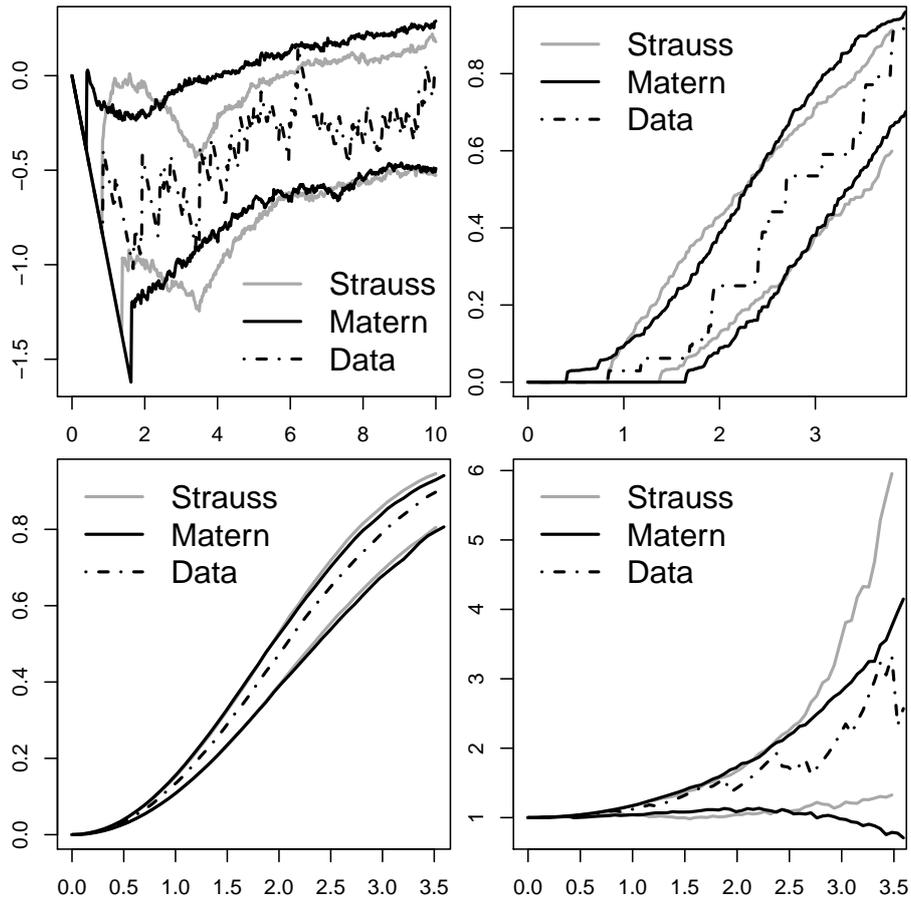


Figure 8: Clockwise from top left: Non-parametric estimate of $L(r) - r$, $G(r)$, $J(r)$, and $F(r)$ for the Spanish towns dataset. All four plots contain simulation based 2.5% and 97.5% pointwise quantiles for both the fitted Strauss hard-core model (gray lines) and the fitted Whittle-Matérn model (black lines). For both models the quantiles are based on 400 simulated realizations.

6.2 Hamster cells dataset

Figure 1(b) shows a plot of the locations of 303 cells of two types in a 0.25 mm by 0.25 mm region of a histological section of the kidney of a hamster. The data has been rescaled to a unit square and there are 226 dividing (living) cells marked by circles and 77 pyknotic (dying) cells marked by pluses. The dataset 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 (SSI) process with hardcore 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 hardcore condition. However, Diggle considered the good overall fit as an indication that the SSI model, together with random labelling of cell types, provides a reasonable approximate description of the data.

Under the assumption of random labelling, the two sub-point 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 sub-point 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 $(\rho_2/\rho)C_0$, cf. Proposition A.2. 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 sub-point 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 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 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 400 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 get 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 to the empirical value of Π for the dataset. The resulting p -value is 0.62 and there is no reason to reject the null hypothesis.

6.3 Oak and beech trees dataset

Figure 1(c) shows a plot of the locations of 244 trees of the species oak and beech. The dataset originates from Pommerening (2002) and was analysed in Mecke and Stoyan (2005) and Illian et al. (2008). In the following analysis we ignore the species type. In this case Mecke and Stoyan (2005) noted that only powerful tests can reject the null hypothesis of a homogeneous Poisson model.

Among the four parametric classes of DPPs, the Whittle-Matérn model with $\hat{\rho} = n/|W| = 0.038$, $\hat{\nu} = 0.4$, and $\hat{\alpha} = 2.28$ has the highest likelihood and shows a reasonable fit, cf. Figure 9. The estimate $\hat{\nu} = 0.4$ could indicate that we are close to the Poisson model. However, when we performed a simulation based likelihood-ratio test similar to the one in Section 6.1, all 400 simulated values of the test statistic were below the observed value, so the null hypothesis of a homogeneous Poisson model is clearly rejected.

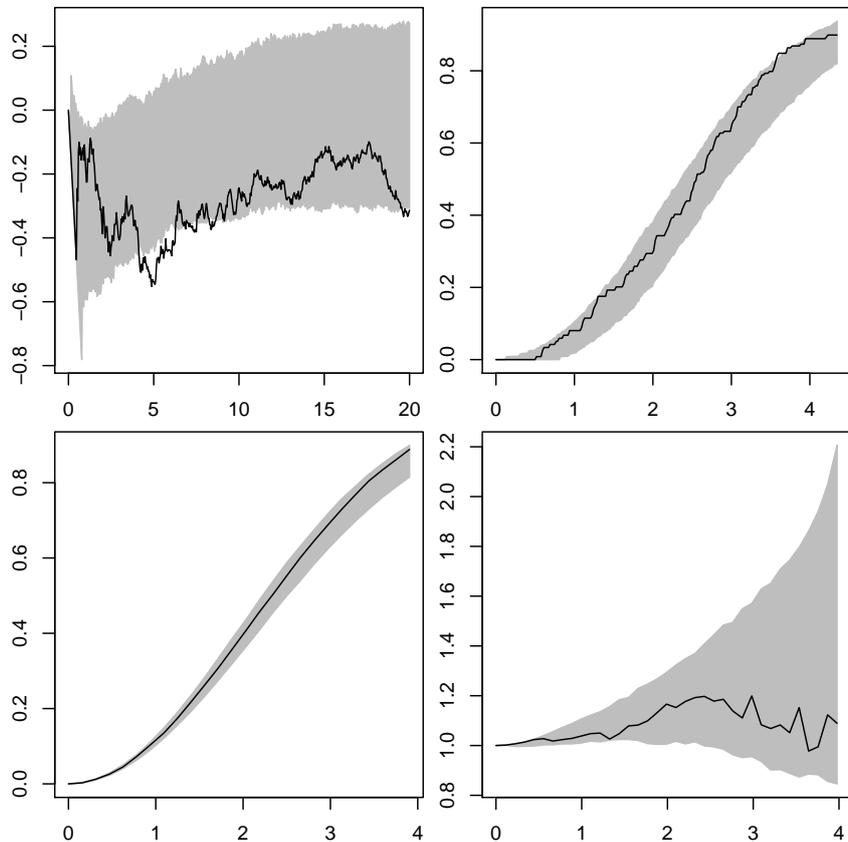


Figure 9: Clockwise from top left: Non-parametric estimate of $L(r) - r$, $G(r)$, $J(r)$, and $F(r)$ for the oak and beech dataset. All four plots contain simulation based 2.5% and 97.5% pointwise quantiles obtained from 400 simulated realizations of the fitted Whittle-Matérn model.

6.4 Termite mounds dataset

Figure 1(d) shows a plot of the locations of 48 termite mounds. The full dataset also contains an associated pattern of palm tree locations which is omitted here. The

dataset was analysed in Illian et al. (2008) where the focus was on the interaction between palms and termite mounds. The palm locations were modelled conditional on the mound locations assuming that the palm locations constitute a cluster process with the mounds as cluster centers. It may therefore be interesting to model and analyse the locations of the mounds separately to know which process generated the cluster centers.

When we fit the four parametric classes of DPPs, the power exponential spectral DPP model has the highest likelihood and an estimate $(\hat{\rho}, \hat{\alpha}, \hat{\nu}) = (0.00128, 48, 15)$. This is on the border of the parameter space and the fitted model is close to the jinc-like DPP with kernel (3.28). The fitted model is judged to provide a good fit, see Figure 10. The figure also indicates that the Gaussian model cannot be used for this strength of repulsion.

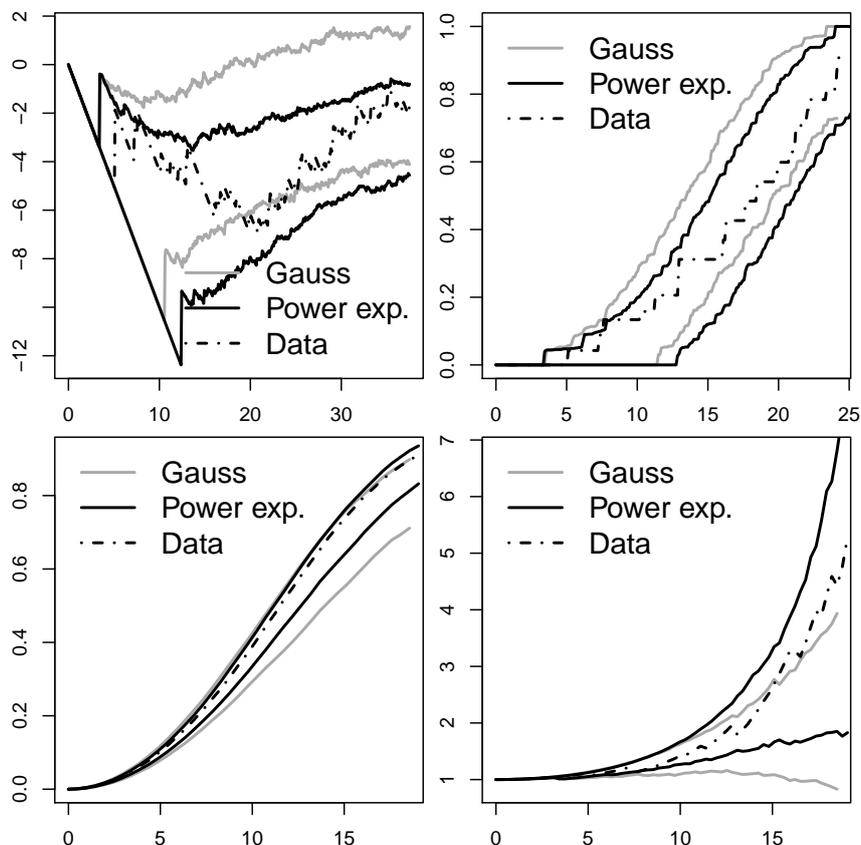


Figure 10: Clockwise from top left: Non-parametric estimate of $L(r) - r$, $G(r)$, $J(r)$, and $F(r)$ for the termites dataset. All four plots contain simulation based 2.5% and 97.5% pointwise quantiles for both the fitted Gaussian model and the fitted power exponential spectral model. For both models the quantiles are based on 400 simulated realizations.

6.5 Mucous membrane dataset

The mucous membrane dataset shown in Figure 1(e) consists of the most abundant type of cell in a bivariate point pattern analysed in Møller and Waagepetersen (2004).

We use this point pattern to illustrate how an inhomogeneous DPP can be fitted to a real dataset. We shall exploit that when the kernel C of a DPP is a covariance function it can be written as

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

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

One way of constructing an inhomogeneous DPP model is to assume that ρ is spatially varying while R is invariant by translation. This is equivalent to assume second-order intensity-reweighted stationarity (Baddeley et al., 2000). A parametric model can then be constructed using the models of Sections 3.3 and 3.4 for R and specifying a parametric model for ρ possibly depending on spatial covariates. Our approximation of the likelihood presented in Section 4.4 is not adapted to this non-stationary setting, but e.g. a minimum contrast procedure as in Section 5.2 may be used to estimate all parameters of the model. However, we do not use this approach for the present dataset, because a comparison (not reported here) of the pair correlation function estimated from the bottom-half and from the top-half of the window indicates that the dataset is not second-order intensity-reweighted stationary.

Instead we propose another way of handling inhomogeneity in a DPP model. Assume X is a planar DPP with kernel C and observed within a rectangular window $W = [0, A] \times [0, B]$, where $A > 0$ and $B > 0$. Suppose X_W has a separable intensity function: $\rho(x_1, x_2) = \rho_1(x_1)\rho_2(x_2)$, where ρ_1 (resp. ρ_2) is positive and integrable on $[0, A]$ (resp. on $[0, B]$). Let $T_1(x_1) = \int_0^{x_1} \rho_1(u) du$, $0 \leq x_1 \leq A$, and $T_2(x_2) = \int_0^{x_2} \rho_2(u) du$, $0 \leq x_2 \leq B$. Define $T(x_1, x_2) = (T_1(x_1), T_2(x_2))$ for $(x_1, x_2) \in W$. The transformed point process $Y = T(X_W)$ is a DPP defined on $T(W) = [0, T_1(A)] \times [0, T_2(B)]$ and its kernel C_Y is deduced from Proposition A.1:

$$C_Y(x, y) = R(T^{-1}(x), T^{-1}(y)), \quad x, y \in T(W), \quad (6.2)$$

with R as in (6.1). Note that for $x \in W$, Y has intensity $C_Y(x, x) = R(T^{-1}(x), T^{-1}(x)) = 1$, since R is a correlation function. Now, assume that Y corresponds to the restriction of a stationary DPP Z to $T(W)$, i.e. $Y = Z_{T(W)}$ and for $x, y \in \mathbb{R}^2$, $C_Y(x, y) = C_{Y,0}(y - x)$. Thus by (6.2), for $x, y \in W$, $R(x, y) = C_{Y,0}(T(y) - T(x))$ and the kernel of X_W follows from (6.1):

$$C(x, y) = \sqrt{\rho(x)} C_{Y,0}(T(y) - T(x)) \sqrt{\rho(y)}, \quad x, y \in W. \quad (6.3)$$

Note that X_W has pair correlation function $g(x, y) = 1 - |C_Y(T(y) - T(x))|^2$. This implies in particular that X is not second-order intensity-reweighted stationary, which is in agreement with our remark above.

In summary we fit an inhomogeneous DPP X_W with separable intensity as follows.

- Fit the intensity function ρ restricted to W (e.g. by the Poisson maximum likelihood estimator, see Schoenberg (2005)).
- Apply the transformation T introduced above to obtain $Y = T(X_W)$.

- Fit a stationary DPP model for Z based on $Y = Z_{T(W)}$, and use (6.3) to obtain the kernel of X_W .

We apply this procedure to the mucous membrane dataset, where $A = 1$ and $B = 0.81$. Considering Figure 1(e) it seems reasonable to assume horizontal homogeneity, i.e. $\rho(x_1, x_2) = \rho_1 \rho_2(x_2)$ with ρ_1 a positive constant. We simply model ρ_2 as piecewise constant on the nine intervals $[0.09(i-1), 0.09i]$, $i = 1, \dots, 9$ (though ρ is then not continuous, X_W becomes a DPP according to Definition 2.1). Thus T_1 is linear and T_2 is piecewise linear making the transformation T very simple. Note that we can choose any positive value for ρ_1 as this choice just amounts to rescaling ρ_2 . We fix $\hat{\rho}_1 = \sqrt{n}/A$, where $n = 876$ is the number of points, whereby T_1 is determined. If T_2 determined by an estimate of ρ_2 satisfies $T_1(A)T_2(B) = n$, then $T(W) = [0, \sqrt{n}]^2$. We therefore estimate ρ_2 on each interval by the frequency of points with first coordinate in the interval divided by $0.09\sqrt{n}$. This gives the estimates 34, 59, 46, 39, 34, 36, 32, 34, 16. Figure 11(a) shows the fitted piecewise constant intensity $\hat{\rho}$. The dataset transformed by T is shown in Figure 12(a) and is modelled as the restriction of a stationary DPP using each of the four parametric classes of DPPs.

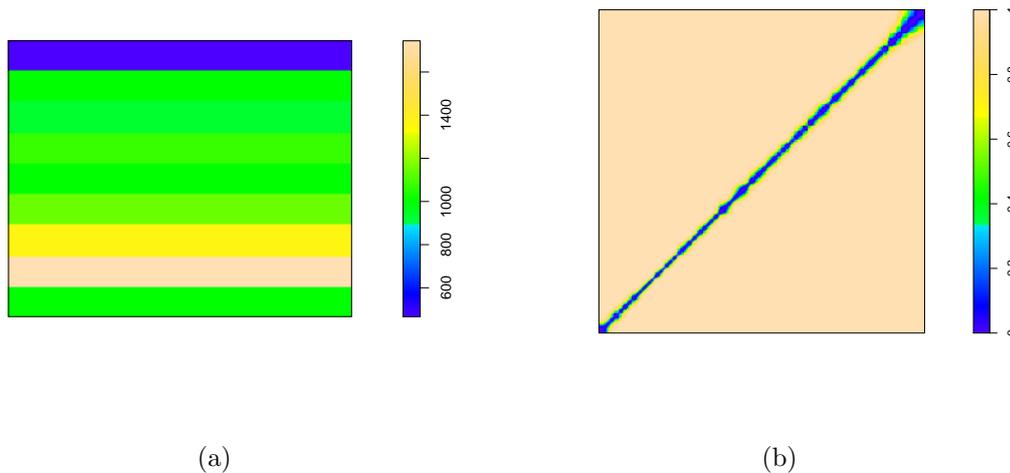


Figure 11: Mucous membrane dataset: (a) Fitted piecewise constant intensity function $\hat{\rho}$. (b) Fitted pair correlation function with x_2 along the abscissa, y_2 along the ordinate, and $y_1 - x_1 = 0$ (see the text).

The fitted power exponential spectral model has the highest likelihood of the four models, but the likelihood value is only slightly larger than for the Gaussian model. The simulated 95% envelopes in Figure 12(b) indicate that these two models are very close in terms of the considered functional summary statistics. A simulation based likelihood-ratio test comparing the Gaussian null model with the power exponential spectral model yielded a p -value of 0.10, and we thus prefer the Gaussian model for Z with fitted parameter $\hat{\alpha} = 0.48$ (and intensity one as imposed by the transformation T).

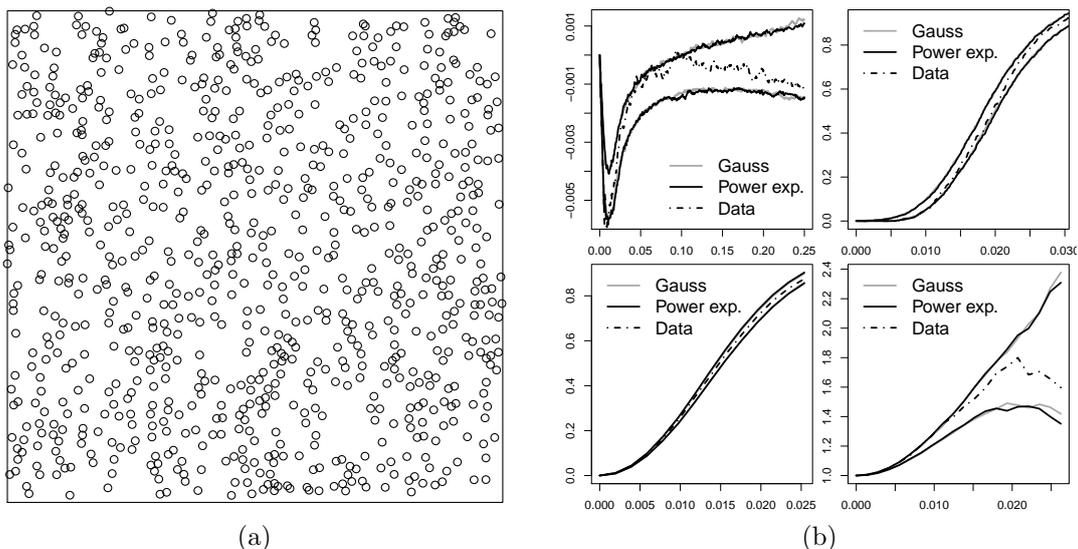


Figure 12: (a) Mucous membrane dataset in $[0, \sqrt{n}]^2$ after transformation; (b) Clock-wise from top left: Non-parametric estimate of $L(r) - r$, $G(r)$, $J(r)$, and $F(r)$ for the transformed mucous membrane dataset. All four plots contain simulation based 2.5% and 97.5% pointwise quantiles for both the power exponential spectral model (black lines) and the Gaussian model (grey lines). For both models the quantiles are based on 400 simulated realizations.

According to the third step of the procedure, we deduce the fitted kernel (6.3) of the mucous dataset and in particular its fitted pair correlation function. The latter is hard to visualize since it is not invariant by translation and depends on two two-dimensional vectors $x = (x_1, x_2)$ and $y = (y_1, y_2)$. However, in our case it only depends on $y_1 - x_1$, x_2 , and y_2 because T_1 is linear. Figure 11(b) shows the fitted pair correlation function as a function of (x_2, y_2) when $y_1 - x_1 = 0$.

6.6 Insect cells dataset

The dataset displayed in Figure 1(f) shows the locations of the centres of 42 cells in a histological section of an insect. These data were analysed in Crick and Lawrence (1975), and presented in Ripley (1977) and Diggle (2003) as an example of a very regular point pattern.

Among the four parametric classes of DPPs the power exponential spectral model has the highest likelihood and estimates $\hat{\rho} = 42$, $\hat{\nu} = 10$, and $\hat{\alpha} = \alpha_{\max} = 0.26$. For numerical stability the estimation of ν is in fact confined to be ≤ 10 , in which case the power exponential spectral model looks like the 'most repulsive possible stationary DPP', that is the jinc-like model (see Section 3.5). Thus the fitted DPP model can be viewed as the jinc-like model with $\hat{\rho} = 42$. This is confirmed in Figure 13, which shows 95% envelopes of summary statistics for the fitted power exponential spectral model and the fitted jinc-like model. The empirical values computed from the cells data and compared with these envelopes indicate that this dataset is too regular to be fitted by a DPP.

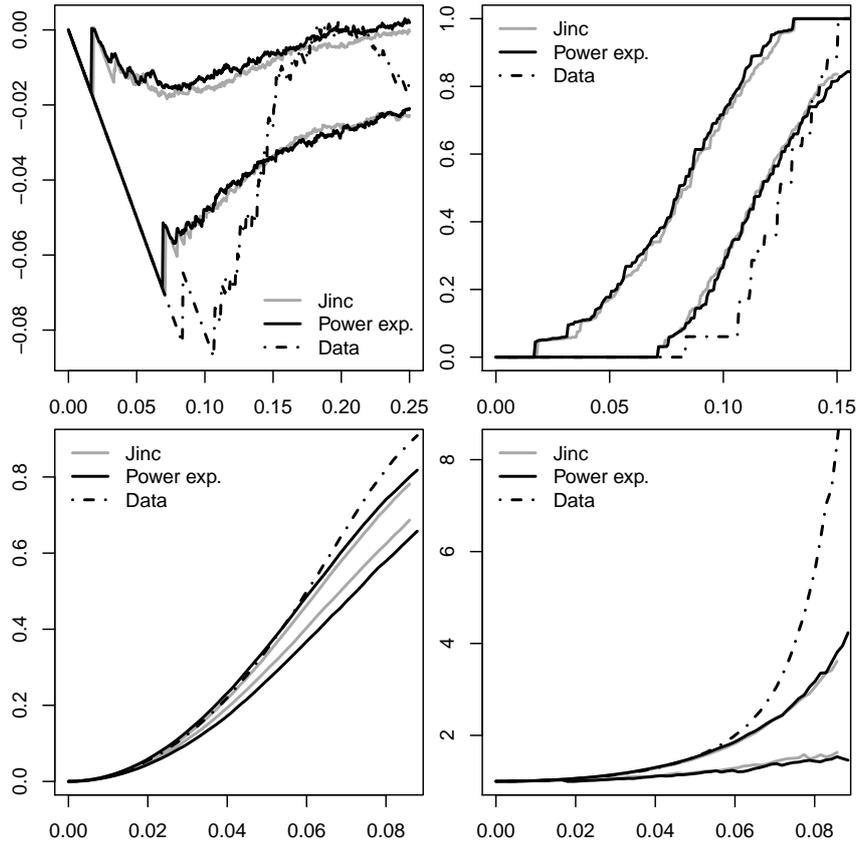


Figure 13: Clockwise from top left: Non-parametric estimate of $L(r) - r$, $G(r)$, $J(r)$, and $F(r)$ for the cells dataset. All four plots contain simulation based 2.5% and 97.5% pointwise quantiles for both the power exponential spectral model fitted via MLE and the jinc-like DPP (grey lines). For both models the quantiles are based on 400 simulated realizations.

7 Concluding remarks

In this paper we have introduced several stationary parametric models for DPPs and discussed to which degree they can model repulsiveness. In analogy with a zero-mean Gaussian process, the law of a DPP is determined by a function, viz. the kernel (or covariance function) C , which as illustrated in our examples of applications can be chosen in many ways. We have also derived approximations which allow us in practice to deal with likelihoods and simulation for DPP models, and we have demonstrated how likelihood and moment based inference procedures work for simulated and real point pattern datasets. In comparison to general Gibbs point processes, DPPs are much easier to handle.

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 dataset. Rather our overall purpose of fitting DPP models is to provide empirical models with a parsimonious parametrization, where we can compare different spatial point pattern datasets 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 e.g. 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.

At several places we discussed repulsiveness of DPPs and realized that a wide range of soft-core cases and to some extent more repulsive cases are well covered by the DPP models introduced in Section 3. However, there is a trade-off between how large the intensity and how repulsive a stationary DPP can be, cf. Section 3.3. In Section 3.5 we suggested the quantity μ in (3.25) as a rough way of quantifying repulsiveness for stationary point processes with a pair correlation function not greater than one, but it may be worthwhile to consider other ways of quantifying repulsiveness in DPPs. In particular, we have characterized the 'most repulsive stationary DPP' with fixed intensity ρ , the kernel of which is a jinc-like function in the planar case, see (3.28).

DPPs cannot be as repulsive as Gibbs hard-core point processes, and this limitation has been demonstrated with the cells dataset of Section 6.6. Nevertheless, the jinc-like DPP model exhibits strong repulsiveness as seen in Figure 2(c) and allows us to fit quite regular datasets such as the termite mounds dataset, see Section 6.4. For comparison consider a Strauss process (Strauss, 1975; Kelly and Ripley, 1976) which is a standard example of a repulsive Gibbs point process. Ignoring edge effects, the Strauss process restricted to a bounded window has a density with respect to the unit rate Poisson process which is proportional to $\beta^{n(\mathbf{x})}\gamma^{s_R(\mathbf{x})}$, where $\beta > 0$, $0 \leq \gamma \leq 1$, $R > 0$, $n(\mathbf{x})$ is the number of points in \mathbf{x} , and $s_R(\mathbf{x})$ is the number of (unordered) R -close pairs of points in \mathbf{x} . The Strauss process fitted by maximum pseudo-likelihood to the termite mounds dataset (Section 6.4) gives $\hat{R} = 23.4$,

$\hat{\beta} = 0.019$, and $\hat{\gamma} = 0.18$. The fitted range of repulsion ($\hat{R} = 23.4$) is in agreement with the practical range of repulsion of the jinc-like DPP, which roughly corresponds to $0.96/\sqrt{\hat{\rho}}$ —see Figure 5(b) where $r_0 = 0.96/\sqrt{\hat{\rho}}$ is the first r -value such that for the jinc-like DPP, $g_0(r) = 0.99$ —that is, $\hat{r}_0 = 0.96/\sqrt{\hat{\rho}} = 26.8$ for the termite mounds data when the estimate $\hat{\rho} = n/|S|$ is used. Comparing the fitted Strauss process with the fitted jinc-like DPP, not only the moments of the jinc-like DPP can be expressible in closed form, unlike the Strauss model, but it turns out that simulation of this jinc-like DPP is very fast, while long Markov chain Monte Carlo simulations are needed for this Strauss process.

Whittle-Matérn and Cauchy models with low values of ν (e.g. $\nu < 0.5$) are very close to Poisson, and in our experience it requires a rather large point pattern dataset before the null hypothesis of Poisson 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 discussed in Section 5.1 in order 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 one 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. In such cases very weakly repulsive Gibbs point processes (e.g. a Strauss process with γ close to one) become interesting competitors to DPPs, unless some other and more efficient approximations of C and \tilde{C} are developed (we leave this problem for future research).

We have mainly considered the stationary case of DPPs, and though this is a natural starting point for the present paper, and stationarity is often assumed in the spatial point process literature, inhomogeneity (or non-stationarity) is a topic of much recent interest (see e.g. Møller and Waagepetersen (2007) and Gelfand et al. (2010)). We have only discussed the inhomogeneous case in Section 6.5, where we exploited that the inhomogeneity is along one of the axes of the rectangular observation window, and hence after a simple transformation we could view the situation as if we were in the stationary case. In other more complicated cases of inhomogeneity, there is a need for developing non-stationary covariance models, and we expect that estimation procedures based on the intensity function and the pair correlation function (e.g. composite likelihoods based on these functions) should play a role. In this connection second-order intensity-reweighted stationarity (Baddeley et al., 2000) could be a useful property, as briefly discussed in Section 6.5. Furthermore, an independent thinning of a DPP is a DPP (Proposition A.2), and this could possibly be exploited for model checking in the inhomogeneous case. For instance, by using thinning probabilities such that homogeneity is obtained, and then applying model checking techniques for the stationary case (see e.g. Section 6).

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 datasets observed within a rect-

angular region are considered. However, applications with non-rectangular observation windows are not uncommon, see e.g. 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 into 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, minimum contrast estimation and other simple alternatives to maximum likelihood as discussed in Section 5.2 will easily apply.

Generalizations of DPPs to weighted DPPs, which also are models for repulsion, and to the closely related permanent and weighted permanent point processes, which are models for attraction, are studied in Shirai and Takahashi (2003) and McCullagh and Møller (2006). Since determinants have a geometric meaning, are multiplicative, and there are algorithms for fast computations, DPPs are much easier to deal with, not at least from a statistical and computational perspective. The approximations of C and \tilde{C} using a Fourier basis approach (Section 4) apply as well for weighted DPPs and weighted permanent 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), see Remark G.2. 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 .

Acknowledgments

We are grateful to Philippe Carmona, Morten Nielsen, and Rasmus Waagepetersen for helpful comments. 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”. Supported by the Centre for Stochastic Geometry and Advanced Bioimaging, funded by a grant from the Villum Foundation.

Appendices

A Smooth transformations and independent thinning of DPPs

Proposition A.1. *Let $U \subseteq \mathbb{R}^d$ be an open set and $T : \mathbb{R}^d \rightarrow U$ a diffeomorphism such that its inverse T^{-1} has a non-zero Jacobian determinant $J_{T^{-1}}(x)$ for all $x \in U$. If $X_1 \sim \text{DPP}(C_1)$ and $X_2 = T(X_1)$, then $X_2 \sim \text{DPP}_U(C_2)$ with*

$$C_2(x, y) = |J_{T^{-1}}(x)|^{1/2} C_1(T^{-1}(x), T^{-1}(y)) |J_{T^{-1}}(y)|^{1/2}. \quad (\text{A.1})$$

Proof. Follows immediately from (2.1) and (2.2). \square

Proposition A.2. *If $X_1 \sim \text{DPP}(C_1)$ and X_2 is obtained as an independent thinning of X_1 with retention probabilities $p(x)$, $x \in \mathbb{R}^d$, then $X_2 \sim \text{DPP}(C_2)$ with $C_2(x, y) = \sqrt{p(x)} C_1(x, y) \sqrt{p(y)}$.*

Proof. Let $U = \{U(x) : x \in \mathbb{R}^d\}$ be a random field of independent Bernoulli variables where $P(U(x) = 1) = p(x)$ and U is independent of X_1 . Then X_2 is distributed as $\{x \in X_1 : U(x) = 1\}$, so from (2.1) and (2.2) it is clear that $X_2 \sim \text{DPP}(C_2)$. \square

B Proof of Theorem 2.3

Let the situation be as in Theorem 2.3 (a slightly different result where the eigenvalues are strictly less than one was first given in Theorem 12 of Macchi (1975)). Recall that C is of local trace class if

$$\text{tr}_S(C) = \sum_{k=1}^{\infty} |\lambda_k^S| < \infty \quad \text{for all compact } S \subset \mathbb{R}^d.$$

We apply Theorem 4.5.5 in Hough et al. (2009), where $C : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{C}$ is Hermitian, locally square integrable, of local trace class, and, as (2.5) may not hold on a Lebesgue nullset, that C is simply given by (2.5). Then existence of $\text{DPP}(C)$ is equivalent to that for all compact $S \subset \mathbb{R}^d$, $0 \leq \lambda_k^S \leq 1$, $k = 1, 2, \dots$. When C is continuous, this nullset vanishes and local square integrability is satisfied. When C is Hermitian and non-negative definite, the eigenvalues are non-negative, and so continuity of C implies the local trace class assumption, since the trace $\sum_{k=1}^{\infty} |\lambda_k^S| = \sum_{k=1}^{\infty} \lambda_k^S = \int_S C(x, x) dx$ is finite. Thereby Theorem 2.3 follows.

C Reduced Palm distributions for DPPs

Recall that for any simple locally finite spatial point process Y on \mathbb{R}^d with intensity function ρ , there exist unique reduced Palm distributions P_x^1 for Lebesgue almost all $x \in \mathbb{R}^d$ with $\rho(x) > 0$, which are determined by that

$$\mathbb{E} \sum_{x \in Y} h(x, Y \setminus \{x\}) = \int \int \rho(x) h(x, \mathbf{x}) dP_x^1(\mathbf{x}) dx$$

for any non-negative Borel function h , where \mathbf{x} denotes a locally finite subset of \mathbb{R}^d . See e.g. Stoyan et al. (1995) and Appendix C.2 in Møller and Waagepetersen (2004). Intuitively, $P_x^!$ is the conditional distribution of $Y \setminus \{x\}$ given that Y has an event at x . When all n 'th order product density functions $\rho^{(n)}$ of Y exist, $n = 1, 2, \dots$, then for Lebesgue almost all $x \in \mathbb{R}^d$ with $\rho(x) > 0$, $P_x^!$ has n 'th order product density function

$$\rho_x^{(n)}(x_1, \dots, x_n) = \rho^{(n+1)}(x, x_1, \dots, x_n) / \rho(x) \quad (\text{C.1})$$

and otherwise we can take $\rho_x^{(n)}(x_1, \dots, x_n) = 0$. See e.g. Lemma 6.4 in Shirai and Takahashi (2003).

For $X \sim \text{DPP}(C)$, using (2.2) it can be shown that for all $x \in \mathbb{R}^d$ with $C(x, x) > 0$, we can take $P_x^! = \text{DPP}(C_x^!)$ where

$$C_x^!(u, v) = \det[C](u, x; v, x) / C(x, x), \quad u, v \in \mathbb{R}^d,$$

and where $[C](x_1, x_2; y_1, y_2)$ is the 2×2 matrix with entries $C(x_i, y_j)$, $i, j = 1, 2$. See Theorem 6.5 in Shirai and Takahashi (2003) (where their condition A is implied by the conditions in our Theorem 2.4). Moreover, (C.1) holds whenever $C(x, x) > 0$.

D Simulation of M

Let the situation be as in Section 2.4.1. For $m = 0, 1, 2, \dots$, let

$$p_m = P(M = m) = \lambda_m \prod_{i>m} (1 - \lambda_i).$$

Note that $m' = \sup\{k \geq 0 : \lambda_k = 1\}$ is finite, and $p_m = 0$ whenever $m < m'$. For $m \geq m'$, the p_m 's can be computed using the recursion

$$p_{m'} = \prod_{k=m'+1}^{\infty} (1 - \lambda_k), \quad p_{m+1} = \frac{\lambda_{m+1}}{\lambda_m(1 - \lambda_{m+1})} p_m, \quad m = m', m' + 1, \dots$$

The calculation of $p_{m'}$ may involve numerical methods. Let F denote the distribution function of M and introduce

$$q_m = F(m) = P(M \leq m) = \sum_{k=0}^m p_k.$$

The inversion method for simulation of M is based on the fact that $F^-(U) = \min\{m : q_m \geq U\}$ is distributed as M if U is uniformly distributed on $(0, 1)$.

E Proof of Theorem 2.8 and related remarks

Let the situation be as in Theorem 2.8. In the sequel, for ease of presentation, we ignore null sets.

We start by proving by induction that for $i = n, \dots, 1$, (2.11) is a probability density and $\mathbf{v}(X_n), \dots, \mathbf{v}(X_i)$ are linearly independent (considering complex scalars).

For $i = n$ and $x \in S$, we have $p_n(x) = \|\mathbf{v}(x)\|^2/n \geq 0$ for all $x \in S$, and

$$\int_S p_n(x) dx = \frac{1}{n} \int_S \|\mathbf{v}(x)\|^2 dx = \frac{1}{n} \int_S \sum_{k=1}^n |\phi_k(x)|^2 dx = 1.$$

Hence p_n is a probability density. Clearly, $p_n(x) = 0$ whenever $\mathbf{v}(x) = \mathbf{0}$, so as X_n is generated from p_n , $\mathbf{v}(X_n) \neq \mathbf{0}$ (almost surely). Thus the induction hypothesis is verified for $i = n$.

Suppose $1 \leq i < n$. By the induction hypothesis, H_i as defined by (E.4) has dimension $n - i$. Let P_i be the matrix of the orthogonal projection from \mathbb{C}^n onto H_i^\perp . By (E.5), for all $x \in S$, $p_i(x) = \|P_i \mathbf{v}(x)\|^2/i \geq 0$ and

$$p_i(x) = 0 \quad \text{whenever } \mathbf{v}(x) \in H_i. \quad (\text{E.1})$$

By the spectral theorem, $P_i = U \Lambda_i U^*$, where U is unitary and Λ_i is diagonal with the first i diagonal elements equal to one and the rest zero. Let u_{kj} denote the (k, j) 'th entry of U . Then

$$p_i(x) = \frac{1}{i} \mathbf{v}(x)^* U \Lambda_i U^* \mathbf{v}(x) = \frac{1}{i} \|\Lambda_i U^* \mathbf{v}(x)\|^2$$

where the j 'th entry of $\Lambda_i U^* \mathbf{v}(x)$ is $\sum_{k=1}^n u_{kj} \phi_k(x)$ if $j \leq i$, and 0 otherwise, so

$$\begin{aligned} \int_S p_i(x) dx &= \frac{1}{i} \int_S \sum_{j=1}^i \sum_{k=1}^n \sum_{l=1}^n u_{kj} \phi_k(x) \bar{u}_{lj} \overline{\phi_l(x)} dx \\ &= \frac{1}{i} \sum_{j=1}^i \sum_{k=1}^n \sum_{l=1}^n u_{kj} \bar{u}_{lj} \int_S \phi_k(x) \overline{\phi_l(x)} dx = \frac{1}{i} \sum_{j=1}^i \sum_{k=1}^n |u_{kj}|^2 \int_S |\phi_k(x)|^2 dx = 1. \end{aligned}$$

Thus p_i is a probability density. Finally, it follows immediately from (E.1) and the induction hypothesis that $\mathbf{v}(X_n), \dots, \mathbf{v}(X_{i+1}), \mathbf{v}(X_i)$ are linearly independent with probability one.

Hence, the induction hypothesis is verified for all $i = n, \dots, 1$.

Now, for iteration $i < n$, write $P_i = P_i(X_n, \dots, X_{i+1})$ and $H_i^\perp = H_i^\perp(X_n, \dots, X_{i+1})$ to emphasize the dependence on the previously generated variables. For $i = n$, set $P_i(X_n, \dots, X_{i+1}) = I_n$ and $H_i^\perp(X_n, \dots, X_{i+1}) = \mathbb{C}^n$. Let

$$\Omega = \{(x_1, \dots, x_n) \in S^n : \mathbf{v}(x_1), \dots, \mathbf{v}(x_n) \text{ are linearly independent}\}$$

be the support of (X_1, \dots, X_n) . Since $p_i(x) = \|P_i \mathbf{v}(x)\|^2/i$, (X_1, \dots, X_n) has density

$$p(x_1, \dots, x_n) = \frac{1}{n!} \prod_{i=1}^n \|P_i(x_n, \dots, x_{i+1}) \mathbf{v}(x_i)\|^2, \quad (x_1, \dots, x_n) \in \Omega.$$

This product is exactly the square of the volume of the parallelepiped determined by the vectors $\mathbf{v}(x_1), \dots, \mathbf{v}(x_n)$, which is equal to the determinant of the $n \times n$ Gram matrix with (i, j) 'th entry $\mathbf{v}(x_i)^* \mathbf{v}(x_j)$, which in turn is equal to the matrix $[K](x_1, \dots, x_n)$. Thus, for $(x_1, \dots, x_n) \in \Omega$,

$$p(x_1, \dots, x_n) = \frac{1}{n!} \det[K](x_1, \dots, x_n). \quad (\text{E.2})$$

Moreover, if $(x_1, \dots, x_n) \in S^n \setminus \Omega$, $\det[K](x_1, \dots, x_n) = |\det[\mathbf{v}(x_1) \dots \mathbf{v}(x_n)]|^2 = 0$. Hence (E.2) is valid for all $(x_1, \dots, x_n) \in S^n$.

Viewing $\{X_1, \dots, X_n\}$ as a point process, the number of points is fixed and equal to n , and hence by definition of $\rho^{(n)}$ for $\{X_1, \dots, X_n\}$,

$$\rho^{(n)}(x_1, \dots, x_n) = n!p(x_1, \dots, x_n) = \det[K](x_1, \dots, x_n), \quad (x_1, \dots, x_n) \in S^n. \quad (\text{E.3})$$

This completes the proof of Theorem 2.8.

Remark E.1. Let $n > 0$, and define $H_n = \{\mathbf{0}\}$ and for $i = n - 1, \dots, 1$,

$$H_i = \text{span}_{\mathbb{C}}\{\mathbf{v}(X_n), \dots, \mathbf{v}(X_{i+1})\} = \left\{ \sum_{j=i+1}^n \alpha_j \mathbf{v}(X_j) : \alpha_j \in \mathbb{C} \right\}. \quad (\text{E.4})$$

With probability one, $\mathbf{v}(X_n), \dots, \mathbf{v}(X_i)$ are linearly independent, cf. the proof above. Thus, almost surely, H_i is a subspace of \mathbb{C}^n of dimension $n - i$. For $i = n - 1, \dots, 1$, by the Gram-Schmidt procedure employed in Algorithm 1, $\mathbf{e}_1, \dots, \mathbf{e}_{n-i}$ is an orthonormal basis of H_i . Further, for $i = n, \dots, 1$, $ip_i(x)$ is the square norm of the orthogonal projection of $\mathbf{v}(x)$ onto H_i^\perp (the orthogonal complement to H_i).

Remark E.2. According to the previous remark,

$$ip_i(x) = \|P_i \mathbf{v}(x)\|^2 \quad (\text{E.5})$$

where P_i is the matrix of the orthogonal projection from \mathbb{C}^n onto H_i^\perp . Denoting by I_n the $n \times n$ identity matrix, we have for $i < n$,

$$P_i = \prod_{k=n}^{i+1} \left(I_n - \frac{\mathbf{v}(x_k) \mathbf{v}(x_k)^*}{K(x_k, x_k)} \right). \quad (\text{E.6})$$

This provides an alternative way to calculate the density $p_i(x)$, where P_i is obtained recursively. This idea was used in Scardicchio et al. (2009) but, as noticed there, the successive multiplication of matrices leads to numerical instabilities. Some corrections must then be applied at each step to make P_i a proper projection matrix when $n - i$ is large. In contrast, the calculation of $p_i(x)$ in Algorithm 1 is straightforward and numerically stable.

Remark E.3. Note that for x such that $\mathbf{v}(x) \in H_i^\perp$, $p_i(x) = \|\mathbf{v}(x)\|^2/i$. Thus for small values of i , simulation of X_i by rejection sampling with respect to a uniform density may be inefficient. However, the computation of $p_i(x)$ is fast, so this is not a major drawback in practice. For the examples in this paper, we have just been using rejection sampling with a uniform instrumental distribution. Appendix F discusses other choices of the instrumental distribution.

F Close upper bounds on the conditional distributions of Algorithm 1

In Remark E.3 we discussed rejection sampling from the densities p_i , $i = n, \dots, 1$, using uniform instrumental distributions. For intensive simulations purposes, for

each i , it is desirable to construct an unnormalized instrumental density which is larger than and close to p_i as well as easy to simulate from.

To find such an unnormalized density, we first notice the following. It follows from Remark E.2 that $ip_i(x)$ is the norm of a vector obtained after $n - i$ successive orthogonal projections of $\mathbf{v}(x)$. These projections commute, so that $ip_i(x)$ is lower than the norm of any projection of $\mathbf{v}(x)$ of a lower order. By (E.5) and (E.6), if $i + 1 \leq k \leq n$, then

$$p_i(x) \leq \frac{1}{i} \left\| \left(I_n - \frac{\mathbf{v}(x_k)\mathbf{v}(x_k)^*}{K(x_k, x_k)} \right) \mathbf{v}(x) \right\|^2$$

and so by (2.10),

$$p_i(x) \leq \frac{1}{i} \min_{i+1 \leq k \leq n} \left(K(x, x) - \frac{|K(x, x_k)|^2}{K(x_k, x_k)} \right), \quad i < n. \quad (\text{F.1})$$

Here the right hand side is an unnormalized density, since it is a continuous function of $x \in S$ where S is compact.

The proof of the following lemma uses (F.1) to derive an explicit upper bound in the specific setting of Section 3, i.e. in the stationary case, when $S = [-1/2, 1/2]^d$, and when the eigenfunctions are Fourier basis functions as in (4.1). Let $x = (x(1), \dots, x(d)) \in \mathbb{R}^d$ and $y = (y(1), \dots, y(d)) \in \mathbb{R}^d$, and suppose that

$$\{\phi_1, \dots, \phi_n\} = \{\varphi_{j_1, \dots, j_d} : j_1 \in J_1(n_1), \dots, j_d \in J_d(n_d)\}$$

where $\varphi_{j_1, \dots, j_d}(x) = \exp\left(2\pi i \sum_{k=1}^d j_k x(k)\right)$ and for $q = 1, \dots, d$, $J_q(n_q)$ denotes some finite subset of \mathbb{Z} with n_q elements, such that $n = \prod_{q=1}^d n_q$. Then the projection kernel (2.10) becomes

$$K(x, y) = \prod_{q=1}^d \sum_{j_q \in J_q(n_q)} e^{2\pi i j_q(x(q) - y(q))}. \quad (\text{F.2})$$

Moreover, for any $r \in \mathbb{N}$, denote $S_q(r) = \sum_{j_q \in J_q(n_q)} j_q^r$, and for any number a , define $a_+ = \max(a, 0)$.

Lemma F.1. *Let K be the projection kernel (F.2). For step $i = n - 1, \dots, 1$ of Algorithm 1, given the $n - i$ previous points $x_k = (x_k(1), \dots, x_k(d))$, $k = i + 1, \dots, n$, we have*

$$p_i(x) \leq \frac{n}{i} \left(1 - \max_{i+1 \leq k \leq n} \prod_{q=1}^d \left(1 - \frac{2\pi}{n_q} |x(q) - x_k(q)| \sqrt{n_q S_q(2) - S_q^2(1)} \right)_+ \right). \quad (\text{F.3})$$

Proof. For $x, y \in \mathbb{R}$, let $K_q(x, y) = \sum_{j_q \in J_q(n_q)} e^{2\pi i j_q(x - y)}$. An analytic expansion of $|K_q(x, y)|^2$ leads to

$$|K_q(x, y)|^2 = \sum_{p=0}^{\infty} (-1)^p (x - y)^{2p} (2\pi)^{2p} \sum_{l=0}^{2p} \frac{(-1)^l}{l!(2p - l)!} S_q(2p - l) S_q(l).$$

Note that

$$\sum_{l=0}^{2p} \frac{(-1)^l}{l!(2p-l)!} S_q(2p-l) S_q(l) = \frac{1}{(2p)!} \sum_{(i,j) \in J_q^2(n_q)} (j-i)^{2p} \geq 0.$$

Therefore, the function $x \rightarrow |K_q(x, y)|^2$ can be expanded into an alternate series. In particular, for any $x, y \in \mathbb{R}$, since $S_q(0) = n_q$,

$$|K_q(x, y)|^2 \geq n_q^2 - 4\pi^2(x-y)^2(n_q S_q(2) - S_q^2(1)).$$

This lower bound is a concave function of $|x-y|$ when

$$|x-y| \leq \frac{n_q}{2\pi \sqrt{n_q S_q(2) - S_q^2(1)}}$$

and so

$$\frac{|K_q(x, y)|^2}{K_q(y, y)} = \frac{|K_q(x, y)|^2}{n_q} \geq \left(n_q - 2\pi|x-y| \sqrt{n_q S_q(2) - S_q^2(1)} \right)_+.$$

Combining this with (F.1), we obtain (F.3). \square

The upper bound in (F.3) provides an unnormalized instrumental density close to p_i . When $d = 1$, this instrumental density is a stepwise linear function, and hence it is very easy to make simulations under the instrumental density. When $d = 2$, the instrumental density provided by (F.3) is a stepwise polynomial function. One strategy is then to provide a further upper-bound making rejection sampling feasible. In our experience this is not so hard for the DPP models we have considered, but since it depends much on the points x_{i+1}, \dots, x_n and the particular model, it seems not easy to state a general result.

G Proof of Theorem 2.9 and related remarks

Theorem 2.9 was first verified in Macchi (1975). Note that the right hand side in (2.14) is not depending on the ordering of the events. Equation (2.14) follows from a longer but in principle straightforward calculation, using (2.8), (E.3), and the fact that if Y follows the homogeneous Poisson process on S with unit intensity, then

$$\rho^{(n)}(x_1, \dots, x_n) = \text{Ef}(Y \cup \{x_1, \dots, x_n\}).$$

See Shirai and Takahashi (2003) and McCullagh and Møller (2006).

Remark G.1. It is possible to express \tilde{C} and D in terms of C without any direct reference to the spectral representations (2.5) and (2.13): Let

$$C_S^1(x, y) = C_S(x, y), \quad C_S^k(x, y) = \int_S C_S^{k-1}(x, z) C_S(z, y) dz, \quad x, y \in S, \quad k = 2, 3, \dots \quad (\text{G.1})$$

Then

$$D = \sum_{k=1}^{\infty} \text{tr}_S(C_S^k)/k \quad (\text{G.2})$$

and

$$\tilde{C}(x, y) = \sum_{k=1}^{\infty} C_S^k(x, y), \quad x, y \in S. \quad (\text{G.3})$$

Also, as noticed in Macchi (1975), \tilde{C} is the unique solution to the integral equation

$$\tilde{C}(x, y) - \int_S \tilde{C}(x, z)C(z, y) dz = C(x, y), \quad x, y \in S.$$

Remark G.2. The density (2.14) is hereditary in the sense that $f(\{x_1, \dots, x_n\}) > 0$ whenever $f(\{x_1, \dots, x_{n+1}\}) > 0$. This allows us to define the Papangelou conditional intensity for all finite point configurations $\mathbf{x} = \{x_1, \dots, x_n\} \subset S$ and points $u \in S \setminus \mathbf{x}$ by

$$\lambda(u; \mathbf{x}) = f(\mathbf{x} \cup \{u\})/f(\mathbf{x}) = \det[\tilde{C}](x_1, \dots, x_n, u)/\det[\tilde{C}](x_1, \dots, x_n)$$

(taking $0/0 = 0$). Georgii and Yoo (2005) use this to study the link to Gibbs point processes, and establish the following result of statistical interest: for any finite point configurations $\mathbf{x} \subset S$ and $\mathbf{y} \subset S$,

$$\lambda(u; \mathbf{x}) \geq \lambda(u; \mathbf{y}) \quad \text{whenever } \mathbf{x} \subset \mathbf{y} \quad (\text{G.4})$$

and for any point $u \in S \setminus \mathbf{x}$,

$$\lambda(u; \mathbf{x}) \leq \tilde{C}(u, u) \quad (\text{G.5})$$

(Theorem 3.1 in Georgii and Yoo (2005)). The monotonicity property (G.4) is once again confirming the repulsiveness of a DPP, and (G.5) means that X_S is locally stable.

Hence X_S can be coupled with a Poisson process Y_S on S with intensity function given by $\tilde{C}(u, u)$, $u \in S$, such that $X_S \subseteq Y_S$ (see Kendall and Møller (2000) and Møller and Waagepetersen (2004)). This coupling is such that X_S is obtained by a dependent thinning of Y_S as detailed in the abovementioned references. By considering a sequence $S_1 \subset S_2 \subset \dots$ of compact sets such that $\mathbb{R}^d = \cup_n S_n$ (e.g. a sequence of increasing balls whose diameters converge to infinity), and a corresponding sequence of processes $X_n \sim \text{DPP}(C; S_n)$ which are coupled with a Poisson process Y on \mathbb{R}^d with intensity function given by $\tilde{C}(u, u)$, $u \in \mathbb{R}^d$, such that $X_1 \subseteq X_2 \subset \dots \subseteq Y$, we obtain that $\cup_n X_n \subseteq Y$ follows $\text{DPP}(C)$. In other words, X can be realized as a dependent thinning of the Poisson process Y .

Imposing certain conditions concerning a finite range assumption on an extended version of \tilde{C} to \mathbb{R}^d and requiring C to be small enough, it is possible to extend the Papangelou conditional intensity for X_S to a global Papangelou conditional intensity for X and hence to derive the reduced Palm distribution of X (for details, see Proposition 3.9 in Georgii and Yoo (2005)). Unfortunately, these conditions are rather restrictive, in particular when $d \geq 2$.

H Proof of Proposition 3.1

For any compact set $S \subset \mathbb{R}^d$, define the integral operator $T_S : L^2(S) \rightarrow L^2(S)$ by

$$T_S(h)(x) = \int_S C(x, y)h(y) dy, \quad h \in L^2(S), \quad x \in S. \quad (\text{H.1})$$

The $\{\lambda_k\}$'s and $\{\phi_k\}$'s involved in (2.5) correspond to the eigenvalues and eigenfunctions of T_S , i.e. for all k ,

$$T_S(\phi_k) = \lambda_k \phi_k. \quad (\text{H.2})$$

For $h \in L^2(S)$, define $h_S \in L^2(\mathbb{R}^d)$ by $h_S(x) = h(x)$ if $x \in S$ and $h_S(x) = 0$ otherwise. From (3.1), the integral operator T_S in (H.1) becomes the convolution operator given by

$$T_S(h)(x) = C_0 \star h_S(x) = \int_S C_0(x - y)h(y) dy, \quad x \in S.$$

Recall that the spectrum of T_S consists of all $\lambda \in \mathbb{C}$ such that the operator $T_S - \lambda I_S$ is not invertible or it is invertible and unbounded (with respect to the usual operator norm), where I_S denotes the identity operator on $L^2(S)$.

Consider the multiplicative operator Q_φ on $L^2(\mathbb{R}^d)$ associated to φ , i.e. $Q_\varphi(h)(x) = \varphi(x)h(x)$ for $h \in L^2(\mathbb{R}^d)$. Its restriction to $L^2(S)$ is given by $Q_{\varphi,S}(h) = Q_{\varphi_S}(h_S)$ for $h \in L^2(S)$. Note that $T_S(h) = \mathcal{F}^{-1}Q_\varphi\mathcal{F}(h_S)$ for $h \in L^2(S)$. Since the Fourier operator is a unitary operator (as $\mathcal{F}\mathcal{F}^{-1} = \mathcal{F}^{-1}\mathcal{F} = I$ where I denotes the identity operator on $L^2(\mathbb{R}^d)$), the spectrum of T_S is equal to the spectrum of Q_{φ_S} , which in turn is equal to $\text{ess-im}(\varphi_S)$ (the essential image of φ_S), see (12) in Section 8.4.3 in Birman and Solomjak (1987). In our case, $\text{ess-im}(\varphi_S)$ is the closure of $\varphi(S)$. Consequently, (C2) is equivalent to $\varphi \leq 1$.

I Proof of Corollary 3.3

Assume (i) in Corollary 3.3. Then $0 \leq \varphi \leq 1$ implies that $\int |\varphi(x)|^2 dx \leq \int |\varphi(x)| dx < \infty$, i.e. $\varphi \in L^2(\mathbb{R}^d)$, and so by Parseval's identity $C_0 \in L^2(\mathbb{R}^d)$. Further, $C_0 = \mathcal{F}^{-1}(\varphi)$ with $\varphi \in L^1(\mathbb{R}^d)$, so C_0 is continuous. By Bochner's theorem, the continuity of C_0 and the non-negativity of φ imply that C_0 is positive-definite, and so (C1) follows from (5.1). Moreover, (C2) holds by Proposition 5.1. Hence (i) implies (ii).

Conversely, assume (ii). Combining Bochner's theorem and the fact that C_0 is continuous and $C_0 \in L^2(\mathbb{R}^d)$, we deduce that there exists $\varphi \in L^1(\mathbb{R}^d)$ such that $C_0 = \mathcal{F}^{-1}(\varphi)$ (see also page 104 in Yaglom (1987)). By (C1), we have that $\varphi \geq 0$. The fact that $\varphi \leq 1$ follows from Proposition 5.1. Hence (ii) implies (i).

J Fourier approximation of the Whittle-Matérn covariance function

Consider the Whittle-Matérn covariance function C_0 given by (3.10). This appendix begins with some preliminary results on K_ν (the Bessel function of the second kind)

which appears in (3.10). Next the quality of the Fourier approximation of C_0 given in Section 4 is discussed.

There are several equivalent ways to define K_ν . By Equation 8.432 in Gradshteyn and Ryzhik (2007), for all $x > 0$ and all $\nu > 0$,

$$K_\nu(x) = \frac{\sqrt{\pi}}{\Gamma(\nu + \frac{1}{2})} \left(\frac{x}{2}\right)^\nu \int_1^\infty e^{-xt} (t^2 - 1)^{\nu - \frac{1}{2}} dt. \quad (\text{J.1})$$

As $x \rightarrow 0$, then $x^\nu K_\nu(x) \rightarrow 2^{\nu-1} \Gamma(\nu)$. Hence by (3.10), $C_0(0) = \rho$.

The following lemma provides an upper bound and gives an idea of the decay rate for K_ν . The inequality reduces to an equality for $\nu = 1/2$. Moreover, according to various plots (omitted in this article), the bound seems sharp when $\nu > 1/2$. We denote $\gamma = \Gamma(1 + 2\nu)^{-1/2\nu}$.

Lemma J.1. *For all $x > 0$,*

$$K_\nu(x) \leq 2^{\nu-1} \Gamma(\nu) x^{-\nu} (1 - (1 - e^{-\gamma x})^{2\nu}) \quad \text{if } \nu \geq 1/2 \quad (\text{J.2})$$

and

$$K_\nu(x) \leq K_{1/2}(x) = \sqrt{\pi/(2x)} e^{-x} \quad \text{if } \nu \leq 1/2. \quad (\text{J.3})$$

Proof. When $\nu \geq 1/2$, from (J.1),

$$K_\nu(x) \leq \frac{\sqrt{\pi}}{\Gamma(\nu + \frac{1}{2})} \left(\frac{x}{2}\right)^\nu \int_1^\infty e^{-xt} t^{2\nu-1} dt = \frac{2^{-\nu} \sqrt{\pi}}{\Gamma(\nu + \frac{1}{2})} x^{-\nu} \Gamma(2\nu, x)$$

where $\Gamma(2\nu, \cdot)$ denotes the incomplete Gamma function with parameter 2ν :

$$\Gamma(2\nu, x) = \int_x^\infty t^{2\nu-1} e^{-t} dt.$$

From Alzer (1997) we deduce that

$$\Gamma(2\nu, x) \leq \left(1 - (1 - e^{-\gamma x})^{2\nu}\right) \Gamma(1 + 2\nu) / (2\nu)$$

whenever $x > 0$, $\nu \geq 1/2$, and $0 \leq \gamma \leq \Gamma(1 + 2\nu)^{-1/2\nu}$. Hence (J.2) follows by using the relations $\Gamma(2\nu + 1) = 2\nu \Gamma(2\nu)$ and $\Gamma(\nu) \Gamma(\nu + 1/2) = 2^{1-2\nu} \sqrt{\pi} \Gamma(2\nu)$.

When $\nu < 1/2$, using (J.1) and the fact that $t^2 - 1 > 2t - 2$ when $t > 1$, we obtain

$$K_\nu(x) \leq \sqrt{\frac{\pi}{2}} \frac{1}{\Gamma(\nu + \frac{1}{2})} x^\nu \int_1^\infty e^{-xt} (t - 1)^{\nu - \frac{1}{2}} dt.$$

Finally, making the change of variables $u = x(t - 1)$, we obtain (J.3). \square

For the Whittle-Matérn model, the following Proposition J.2 provides an error bound for the approximation (4.4) of $C_0(u)$ by $C_{\text{app},0}(u)$ when $u \in [-1/2, 1/2]^d$. We let

$$\beta = \left(\alpha \sqrt{d} (\Gamma(1 + 2\nu)^{1/2\nu} \vee 1)\right)^{-1}$$

$$c(\rho, \nu, \alpha, d) = \begin{cases} (4\alpha)^{1-2\nu} \rho^2 \pi d / \Gamma(\nu)^2 & \text{if } \nu \leq \frac{1}{2} \\ 4\nu^2 \rho^2 d & \text{if } \nu \geq \frac{1}{2} \end{cases} \quad (\text{J.4})$$

$$\epsilon(\nu, \alpha, 1) = \frac{e^{-\beta}}{\beta} + \frac{2e^{-\beta}}{1 - e^{-\beta}} \left(\frac{e^{-\beta}}{\beta} + \frac{1}{1 - e^{-\beta}} - 1 \right) \quad (\text{J.5})$$

and for $d \geq 2$,

$$\epsilon(\nu, \alpha, d) = e^{-\beta} \left(\frac{1}{\beta} + \frac{2}{(1 - e^{-\beta})^2} - \frac{1}{2} \right) \left(\frac{1}{\beta} + \frac{2e^{-\beta}}{1 - e^{-\beta}} \left(\frac{1}{\beta} + \frac{1}{1 - e^{-\beta}} \right) \right)^{d-1}. \quad (\text{J.6})$$

Proposition J.2. *Let C_0 be the Whittle-Matérn covariance function given by (3.10) and let $C_{\text{app},0}$ be the approximation (4.4) of C_0 on $[-1/2, 1/2]^d$. If $0 \leq \rho \leq \rho_{\max}$ where ρ_{\max} is given by (3.11), then*

$$\int_{[-1/2, 1/2]^d} |C_0(x) - C_{\text{app},0}(x)|^2 dx \leq c(\rho, \nu, \alpha, d) \epsilon(\nu, \alpha, d). \quad (\text{J.7})$$

Proof. We have

$$\begin{aligned} \int_{[-1/2, 1/2]^d} |C_0(x) - C_{\text{app},0}(x)|^2 dx &= \int_{[-1/2, 1/2]^d} \left| \sum_{k \in \mathbb{Z}^d} (\alpha_k - \varphi(k)) e^{2\pi i x \cdot k} \right|^2 dx \\ &= \sum_{k \in \mathbb{Z}^d} (\alpha_k - \varphi(k))^2 \end{aligned}$$

with

$$\varphi(k) - \alpha_k = \int_{\mathbb{R}^d \setminus [-1/2, 1/2]^d} C_0(y) e^{-2\pi i k \cdot y} dy.$$

Defining $h(y) = C_0(y)(1 - \mathbb{1}_{[-1/2, 1/2]^d}(y))$, we have $\varphi(k) - \alpha_k = \mathcal{F}(h)(k)$ and

$$\sum_{k \in \mathbb{Z}^d} (\alpha_k - \varphi(k))^2 = \sum_{k \in \mathbb{Z}^d} (\mathcal{F}(h)(k))^2 = \sum_{k \in \mathbb{Z}^d} \mathcal{F}(h \star h)(k).$$

The Poisson summation formula on a lattice (see Stein and Weiss (1971), Chapter VII, Corollary 2.6) gives

$$\sum_{k \in \mathbb{Z}^d} \mathcal{F}(h \star h)(k) = \sum_{k \in \mathbb{Z}^d} h \star h(k).$$

When $\nu \geq 1/2$, we have $1 - (1 - e^{-\gamma x})^{2\nu} \leq 2\nu e^{-\gamma x}$ for all $x > 0$, so from (3.10) and (J.2),

$$h \star h(x) = \int_{\mathcal{D}} C_0(y) C_0(x - y) dy \leq 4\rho^2 \nu^2 \int_{\mathcal{D}} e^{-\frac{\gamma}{\alpha}(\|y\| + \|x - y\|)} dy, \quad x \in \mathbb{R}^d, \quad (\text{J.8})$$

where $\mathcal{D} = \mathcal{D}(x) = \{y \in \mathbb{R}^d : \|x - y\|_{\infty} > 1/2, \|y\|_{\infty} > 1/2\}$ and $\|\cdot\|_{\infty}$ denotes the uniform norm.

Suppose that $\nu \geq 1/2$. When $d = 1$, the latest integral in (J.8) can be computed easily to get

$$\int_{|y| > \frac{1}{2}, |x - y| > \frac{1}{2}} e^{-\frac{\gamma}{\alpha}(|y| + |x - y|)} dy = e^{-\frac{\gamma}{\alpha}|x|} \left(\frac{\alpha}{\gamma} e^{-\frac{\gamma}{\alpha}} + |x| - 1 \right)$$

if $|x| \geq 1$, and the value of the integral at $x = 0$ is $\frac{\alpha}{\gamma}e^{-\frac{\gamma}{\alpha}}$. Thereby, when $d = 1$,

$$\sum_{k \in \mathbb{Z}^d} (\alpha_k - \varphi(k))^2 = \sum_{k \in \mathbb{Z}^d} h \star h(k) \leq 4\rho^2 \nu^2 \left[\frac{\alpha}{\gamma} e^{-\frac{\gamma}{\alpha}} + 2 \sum_{k=1}^{\infty} e^{-\frac{\gamma}{\alpha}k} \left(\frac{\alpha}{\gamma} e^{-\frac{\gamma}{\alpha}} + k - 1 \right) \right]$$

and (J.7), which involves the terms (J.4) (for $\nu \geq 1/2$) and (J.5), follows from the expansion

$$\sum_{k=1}^{\infty} (a+k)q^k = \frac{q}{1-q} \left(a + \frac{1}{1-q} \right) \quad \text{for any } a \in \mathbb{R} \text{ and } |q| < 1.$$

When $d \geq 2$, the integral in (J.8) is more difficult to compute and we therefore establish an upper bound as follows. Since $\|y\| \geq (|y_1| + \dots + |y_d|)/\sqrt{d}$,

$$\begin{aligned} h \star h(x) &\leq 4\rho^2 \nu^2 \int_{\mathcal{D}} \prod_{j=1}^d e^{-\frac{\gamma}{\alpha\sqrt{d}}(|y_j| + |x_j - y_j|)} dy_j \\ &\leq 4\rho^2 \nu^2 d \int_{|y_1 - x_1| > \frac{1}{2}} e^{-\frac{\gamma}{\alpha\sqrt{d}}(|y_1| + |x_1 - y_1|)} dy_1 \prod_{j=2}^d \int_{\mathbb{R}} e^{-\frac{\gamma}{\alpha\sqrt{d}}(|y_j| + |x_j - y_j|)} dy_j. \end{aligned}$$

These integrals are computable: for any $\beta > 0$,

$$\int_{|y_1 - x_1| > \frac{1}{2}} e^{-\beta(|y_1| + |x_1 - y_1|)} dy = \begin{cases} \frac{e^{-\beta}}{\beta} \cosh(\beta x_1) & \text{if } |x_1| \leq \frac{1}{2}, \\ e^{-\beta|x_1|} \left(\frac{1 - e^{-\beta}}{2\beta} + |x_1| - \frac{1}{2} \right) & \text{if } |x_1| \geq \frac{1}{2} \end{cases}$$

and

$$\int_{\mathbb{R}} e^{-\beta(|y| + |x_j - y|)} dy = e^{-\beta|x_j|} \left(|x_j| + \frac{1}{\beta} \right).$$

Therefore, when $d \geq 2$, setting $\beta = \gamma/(\alpha\sqrt{d})$,

$$\begin{aligned} &\sum_{k \in \mathbb{Z}^d} (\alpha_k - \varphi(k))^2 \\ &\leq 4\rho^2 \nu^2 d \left(\frac{e^{-\beta}}{\beta} + 2 \sum_{k=1}^{\infty} e^{-\beta k} \left(\frac{1 - e^{-\beta}}{2\beta} + k - \frac{1}{2} \right) \right) \left(\sum_{k \in \mathbb{Z}} e^{-\beta|k|} \left(|k| + \frac{1}{\beta} \right) \right)^{d-1} \end{aligned}$$

and the bound (J.7), which involves the term (J.6), follows after a straightforward calculation.

Suppose that $\nu \leq 1/2$. From (J.3) we deduce

$$h \star h(x) \leq \frac{\rho^2}{\Gamma(\nu)^2} 2^{2-2\nu} \frac{\pi}{2} \int_{\mathcal{D}} \|y/\alpha\|^{\nu-\frac{1}{2}} \|(x-y)/\alpha\|^{\nu-\frac{1}{2}} e^{-\frac{1}{\alpha}(\|y\| + \|x-y\|)} dy.$$

If $\|x\|_{\infty} > 1/2$, then $\|x\| > 1/2$, and so $\|x\|^{\nu-1/2} < 2^{1/2-\nu}$ and

$$h \star h(x) \leq \frac{\rho^2}{\Gamma(\nu)^2} 2^{2-4\nu} \alpha^{1-2\nu} \pi \int_{\mathcal{D}} e^{-\frac{1}{\alpha}(\|y\| + \|x-y\|)} dy.$$

The latter integral may be bounded similarly as the one in (J.8), and thereby (J.7), which involves the term (J.4), follows. \square

Note that the inequality (J.7) reduces to an equality in the particular case $d = 1$ and $\nu = 1/2$. Finally, the plots in Figure 14 confirm that for reasonable values of ρ , ν , and α satisfying (3.11), the error bound (J.7) is small.

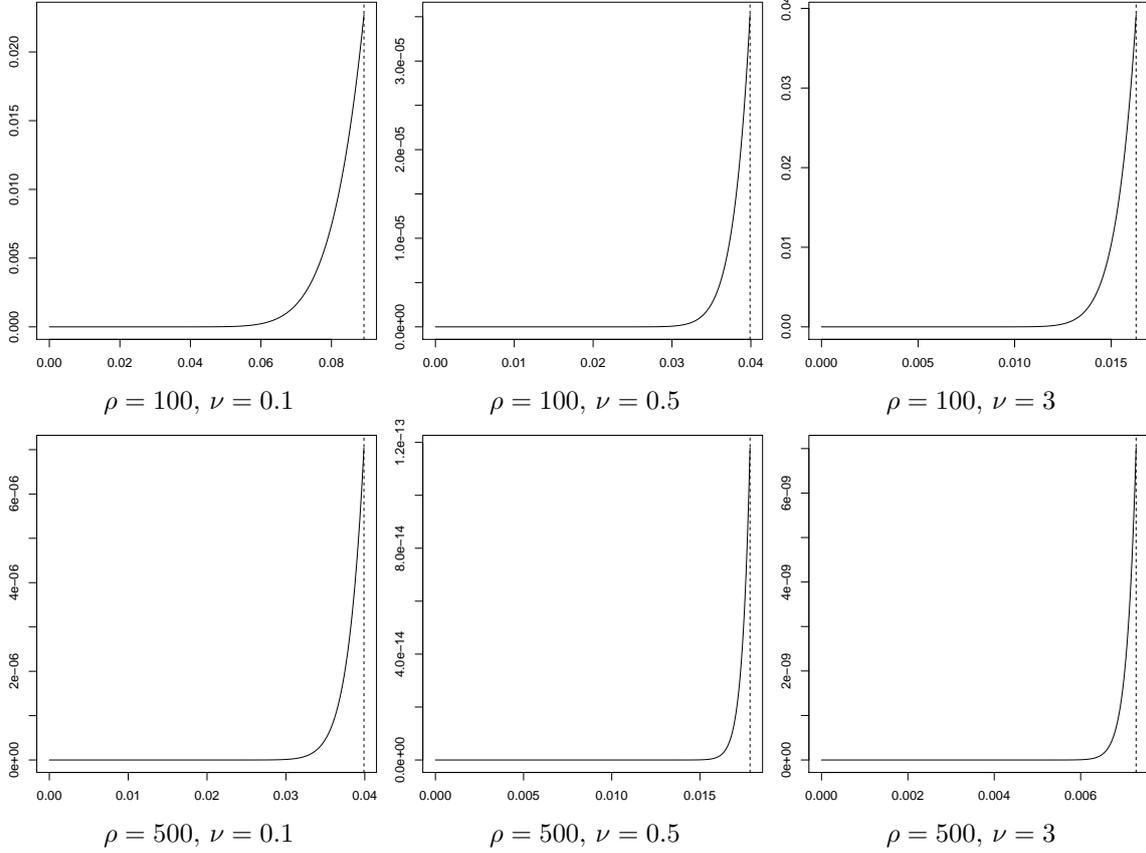


Figure 14: Error-bound (J.7) in terms of α for different values of ρ and ν , when $d = 2$. The dotted line represents the maximal possible value of α following from (3.11).

K Alternative approximation of the density

Let $S \subset \mathbb{R}^d$ be compact. In this appendix, in addition to Assumption 3.2, we assume the slightly stronger condition that the spectral density φ is strictly less than 1. This ensures that all eigenvalues λ_k are strictly less than 1 for all index k so that the density f in Theorem 2.9 is well-defined. Recall that f is given in terms of \tilde{C} and D , cf. (2.14). Below we introduce computationally convenient approximations of \tilde{C} and D which can be used with (2.14) to obtain an approximation of f .

K.1 Convolution approximation of f

We start by showing that $\tilde{C}_{\text{app},0}$ given by

$$\tilde{C}_{\text{app},0}(u) = \sum_{k=1}^{\infty} C_0^{*k}(u), \quad u \in \mathbb{R}^d, \quad (\text{K.1})$$

is well-defined, where

$$C_0^{\star 1}(u) = C_0(u), \quad C_0^{\star k}(u) = \int C_0^{\star(k-1)}(x)C_0(u-x) \, dx, \quad u \in \mathbb{R}^d, \quad k = 2, 3, \dots \quad (\text{K.2})$$

Since $0 \leq \varphi < 1$ and $\varphi \in L^1(\mathbb{R}^d)$, for all $p \in [1, \infty]$, we have $\varphi \in L^p(\mathbb{R}^d)$. Define $\tilde{\varphi} = \varphi/(1 - \varphi)$. For any $u \in \mathbb{R}^d$, $\varphi(u) = \lim_{n \rightarrow \infty} \tilde{\varphi}_n(u)$, where $\tilde{\varphi}_n(u) = \sum_{k=1}^n \varphi(u)^k$. We see that $\tilde{\varphi} \in L^1(\mathbb{R}^d)$ since

$$\|\tilde{\varphi}\|_1 = \int \tilde{\varphi}(u) \, du = \sum_{k=1}^{\infty} \int \varphi(u)^k \, du \leq \sum_{k=1}^{\infty} \|\varphi\|_{\infty}^{k-1} \int \varphi(u) \, du = \frac{\|\varphi\|_1}{1 - \|\varphi\|_{\infty}} < \infty$$

using the monotone convergence theorem to swap summation and integration to obtain the second identity. Therefore $\mathcal{F}^{-1}\tilde{\varphi}$ is well-defined. Using the dominated convergence theorem and similar arguments as above, we see that $(\mathcal{F}^{-1}\tilde{\varphi})(u)$ is equal to the right hand side of (K.1).

For $x, y \in S$, we define $\tilde{C}_{\text{app}}(x, y) = \tilde{C}_{\text{app},0}(x - y)$ and use the approximation $\tilde{C}(x, y) \approx \tilde{C}_{\text{app}}(x, y)$. The expansion (K.1) corresponds to (G.3) with $C_S^k(x, y)$ substituted by $C_0^{\star k}(x - y)$.

Using the same substitution in (G.2) leads us to approximate D by

$$D_{\text{app}} = |S| \sum_{k=1}^{\infty} C_0^{\star k}(0)/k. \quad (\text{K.3})$$

Since $C_0^{\star k}(0) = \int \varphi(u)^k \, du$, we obtain an alternative expression for D_{app} by applying the monotone convergence theorem,

$$D_{\text{app}} = |S| \int -\log(1 - \varphi(u)) \, du = |S| \int \log(1 + \tilde{\varphi}(u)) \, du.$$

Then the convolution approximation of f is defined by

$$f_{\text{app}}(\{x_1, \dots, x_n\}) = \exp(|S| - D_{\text{app}}) \det[\tilde{C}_{\text{app}}](x_1, \dots, x_n). \quad (\text{K.4})$$

As mentioned above, the approximations \tilde{C}_{app} and D_{app} involve approximating $C_S^k(x, y)$ by $C_0^{\star k}(u)$, where $u = x - y$. In fact the approximations provide upper bounds, since $C_S^k(x, y) \leq C_0^{\star k}(u)$ for all x, y and k . Heuristically, when approximating $C_S^k(x, y)$ by $C_0^{\star k}(u)$, we expect that the relative error increases as k grows, since the approximation is applied iteratively, cf. (G.1) and (K.2). However, the final approximations $\tilde{C} \approx \tilde{C}_{\text{app}}$ and $D \approx D_{\text{app}}$ involve sums of $C_S^k(x, y)$ and $C_0^{\star k}(u)$, and the terms with a large relative error may only have a small effect if $C_0^{\star k}(u)$ tends to zero sufficiently fast for $k \rightarrow \infty$. Since $C_0^{\star k}$ is a covariance function, we have $C_0^{\star k}(u) \leq C_0^{\star k}(0)$ for all $k = 1, 2, \dots$. Consequently, we expect that the accuracy of approximating f by f_{app} depends on how fast $C_0^{\star k}(0)$ tends to zero. This is further discussed in the examples below.

K.2 Examples

To use the density approximation f_{app} in practice we truncate the sums in (K.1) and (K.3), i.e.

$$\tilde{C}_{\text{app},0}(u) \approx \sum_{k=1}^N C_0^{*k}(u) \quad \text{and} \quad D_{\text{app}} \approx |S| \sum_{k=1}^N C_0^{*k}(0)/k$$

where N is a positive integer. Furthermore, we need closed form expressions for $C_0^{*k}(u)$. For the normal variance mixture models presented in Section 3.3, we have $C_0^{*k}(u) = (\rho/\rho_{\max})^k h^{*k}(u)$, and so it suffices to find closed form expressions for h^{*k} . For the Gaussian model,

$$h^{*k}(u) = (k\pi\alpha^2)^{-d/2} \exp(-\|u/\alpha\|^2/k), \quad u \in \mathbb{R}^d,$$

while for the Whittle-Matérn model,

$$h^{*k}(u) = \frac{\|u/\alpha\|^{\nu'} K_{\nu'}(\|u/\alpha\|)}{2^{\nu'-1} (\sqrt{\pi}\alpha)^d \Gamma(\nu' + d/2)}, \quad u \in \mathbb{R}^d,$$

where $\nu' = k(\nu + d/2) - d/2$. We have no closed form expression for the Cauchy model.

For both the Gaussian and the Whittle-Matérn covariance function, $h^{*k}(0)$ decays as $k^{-d/2}$ when $k \rightarrow \infty$, and therefore the rate of convergence of $\tilde{C}_{\text{app},0}(0) = \sum_{k=1}^{\infty} (\rho/\rho_{\max})^k h^{*k}(0)$ depends crucially on d and ρ . For $d < 3$, the series only converges if $\rho < \rho_{\max}$, and the series converges slowly when ρ is close to ρ_{\max} .

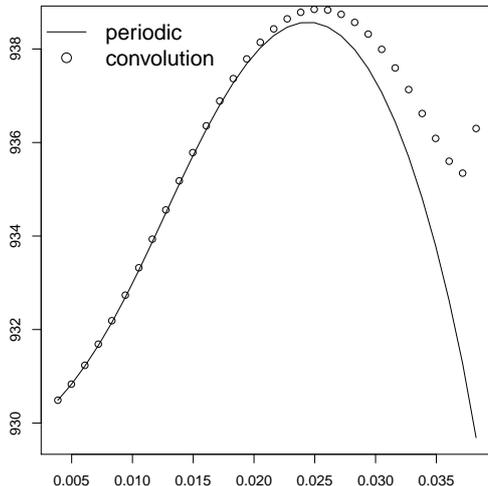


Figure 15: Comparison of using the convolution and periodic density approximations to approximate the log-likelihood of the Gaussian model as a function of α based on a simulated dataset in the unit square with $\rho = 200$ and $\alpha = 0.02$.

Based on a simulated point pattern in the unit square, Figure 15 compares the approximations obtained using the convolution and periodic density approximations to approximate the log-likelihood for the Gaussian model with $\rho = 200$ and

$\alpha = 0.02$. The simulated point pattern has $\hat{\rho} = 213$ points. In the likelihood calculations, $\rho = \hat{\rho}$ is fixed such that the only varying parameter is $\alpha \in (0, \alpha_{\max})$, where $\alpha_{\max} = 1/\sqrt{\pi\hat{\rho}} = 0.39$. For both approximations, the truncation N was increased until almost no change appeared in the approximations. In this example, $N = 256$ for the convolution approximation and $N = 512$ for the periodic approximation. As in the simulation study in Section 5.3, the periodic approximation is giving effectively unbiased estimates. However, similar simulation studies (not reported here) using the convolution approximation yielded estimates of α which were positively biased, which is in agreement with Figure 15. In particular using the convolution approximation we get a large proportion of estimates with $\hat{\alpha} = \alpha_{\max}$. For smaller values of α , $\alpha < \alpha_{\max}/2$ say, the two approximations are very similar, and in this case $\rho/\rho_{\max} < 1/2$, so the convolution approximation converges rapidly, and in this range of α -values, a truncation of $N = 10$ is sufficient to obtain stable results. This is computationally much faster than using the periodic approximation with $N = 512$, and therefore the convolution approximation is appealing when ρ/ρ_{\max} is small.

References

- Affandi, R., A. Kulesza, and E. Fox (2012). Markov determinantal point processes. In K. M. N. de Freitas (Ed.), *Proceedings of the Twenty-Eight Conference on Uncertainty in Artificial Intelligence (UAI-12)*, Corvallis, Oregon, pp. 26–35. AUAI Press.
- Alzer, H. (1997). On some inequalities for the incomplete Gamma function. *Mathematics of Computation* 66, 771–778.
- Baddeley, A., J. Møller, and R. Waagepetersen (2000). Non- and semi-parametric estimation of interaction in inhomogeneous point patterns. *Statistica Neerlandica* 54, 329–350.
- Baddeley, A. and R. Turner (2000). Practical maximum pseudolikelihood for spatial point patterns. *Australian and New Zealand Journal of Statistics* 42, 283–322.
- Baddeley, A. and R. Turner (2005). Spatstat: an R package for analyzing spatial point patterns. *Journal of Statistical Software* 12(6), 1–42. URL: www.jstatsoft.org, ISSN: 1548-7660.
- Barndorff-Nielsen, O. (1977). Exponentially decreasing distributions for the logarithm of particle size. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences* 353, 401–419.
- Barndorff-Nielsen, O. (1978). Hyperbolic distributions and distributions on hyperbolae. *Scandinavian Journal of Statistics* 5, 151–157.
- Besag, J. (1977a). Some methods of statistical analysis for spatial data. *Bulletin of the International Statistical Institute* 47, 77–92.
- Besag, J. E. (1977b). Contribution to the discussion of Dr Ripley’s paper. *Journal of Royal Statistical Society. Series B. Statistical Methodology* 39, 193–195.

- Birman, M. and Z. Solomjak (1987). *Spectral theory of Self-Adjoint Operator in Hilbert Space*. Dordrecht: D. Reidel Publishing Company.
- Crick, F. H. C. and P. A. Lawrence (1975). Compartments and polyclones in insect development. *Science* 189, 340–347.
- De Iaco, S., M. Palma, and D. Posa (2003). Covariance functions and models for complex-valued random fields. *Stochastic Environmental Research and Risk Assessment* 17, 145–156.
- Diggle, P. (2003). *Statistical Analysis of Spatial Point Patterns* (Second ed.). London: Hodder Arnold.
- Diggle, P. and R. Gratton (1984). Monte Carlo methods of inference for implicit statistical models (with discussion). *Journal of Royal Statistical Society. Series B. Statistical Methodology* 46, 193–227.
- Gelfand, A. E., P. J. Diggle, P. Guttorp, and M. Fuentes (2010). *Handbook of Spatial Statistics*. CRC Press, Boca Raton.
- Georgii, H.-O. and H. J. Yoo (2005). Conditional intensity and Gibbsianness of determinantal point processes. *Journal of Statistical Physics* 118, 617–666.
- Glass, L. and W. R. Tobler (1971). Uniform distribution of objects in a homogeneous field: Cities on a plain. *Nature* 233, 67–68.
- Gneiting, T. (1997). Normal scale mixtures and dual probability densities. *Journal of Statistical Computation and Simulation* 59, 375–384.
- Gneiting, T. (2002). Compactly supported correlation functions. *Journal of Multivariate Analysis* 83, 493 – 508.
- Goovaerts, P. (1997). *Geostatistics for Natural Resources Evaluation*. New York: Oxford University.
- Gradshteyn, I. and I. Ryzhik (2007). *Table of Integrals, Series, and Products* (7th ed.). San Diego: Academic Press.
- Harkness, R. and V. Isham (1983). A bivariate spatial point pattern of ants nests. *Applied Statistics* 32, 293–303.
- Hough, J. B., M. Krishnapur, Y. Peres, and B. Viràg (2006). Determinantal processes and independence. *Probability Surveys* 3, 206–229.
- Hough, J. B., M. Krishnapur, Y. Peres, and B. Viràg (2009). *Zeros of Gaussian Analytic Functions and Determinantal Point Processes*. Providence: American Mathematical Society.
- Huang, F. and Y. Ogata (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., A. Penttinen, H. Stoyan, and D. Stoyan (2008). *Statistical Analysis and Modelling of Spatial Point Patterns*. John Wiley and Sons, Chichester.
- Jensen, J. L. and J. Møller (1991). Pseudolikelihood for exponential family models of spatial point processes. *Annals of Applied Probability* 1, 445–461.
- Kelly, F. P. and B. D. Ripley (1976). A note on Strauss’ model for clustering. *Biometrika* 63, 357–360.
- Kendall, W. S. and J. Møller (2000). Perfect simulation using dominating processes on ordered spaces, with application to locally stable point processes. *Advances in Applied Probability* 32, 844–865.
- Kulesza, A. and B. Taskar (2012). Determinantal point processes for machine learning. *Foundations and Trends in Machine Learning* 5, 123–286.
- Kuna, T., J. Lebowitz, and E. Speer (2007). Realizability of point processes. *Journal of Statistical Physics* 129, 417–439.
- Leonardi, E. and G. L. Torrisi (2013). Large deviations of the interference in the ginibre network model. *preprint (arxiv:1304.2234)*, 1–35.
- Lieshout, M. N. M. v. (2000). *Markov Point Processes and Their Applications*. Imperial College Press, London.
- Lindgren, F., J. Lindström, and H. Rue (2011). An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. *Journal of Royal Statistical Society. Series B. Statistical Methodology* 73, 423–498.
- Macchi, O. (1975). The coincidence approach to stochastic point processes. *Advances in Applied Probability* 7, 83–122.
- McCullagh, P. and J. Møller (2006). The permanent process. *Advances in Applied Probability* 38, 873–888.
- Mecke, K. R. and D. Stoyan (2005). Morphological characterization of point patterns. *Biometrical Journal* 47, 473–488.
- Miyoshi, N. and T. Shirai (2013). A cellular network model with ginibre configured base stations. Technical report, Department of Mathematical and Computing Sciences Tokyo Institute of Technology, series B: Applied Mathematical Science.
- Møller, J. and R. P. Waagepetersen (2004). *Statistical Inference and Simulation for Spatial Point Processes*. Boca Raton: Chapman and Hall/CRC.
- Møller, J. and R. P. Waagepetersen (2007). Modern spatial point process modelling and inference (with discussion). *Scandinavian Journal of Statistics* 34, 643–711.
- Nelder, J. A. and R. Mead (1965). A simplex method for function minimization. *Computer Journal* 7, 308–313.

- Pommerening, A. (2002). Approaches to quantifying forest structures. *Forestry* 75, 305–324.
- Prokešová, M. and E. B. V. Jensen (2013). Asymptotic Palm likelihood theory for stationary point processes. *Annals of the Institute of Statistical Mathematics* 65, 387–412.
- R Development Core Team (2011). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0.
- Riesz, F. and B. Sz.-Nagy (1990). *Functional Analysis*. New York: Dover Publications.
- Ripley, B. D. (1976). The second-order analysis of stationary point processes. *Journal of Applied Probability* 13, 255–266.
- Ripley, B. D. (1977). Modelling spatial patterns (with discussion). *Journal of Royal Statistical Society. Series B. Statistical Methodology* 39, 172–212.
- Ripley, B. D. (1988). *Statistical Inference for Spatial Processes*. Cambridge University Press.
- Ripley, B. D. and F. P. Kelly (1977). Markov point processes. *Journal of the London Mathematical Society* 15, 188–192.
- Scardicchio, A., C. Zachary, and S. Torquato (2009). Statistical properties of determinantal point processes in high-dimensional Euclidean spaces. *Physical Review E* 79(4), Article 041108.
- Schoenberg, F. (2005). Consistent parametric estimation of the intensity of a spatial-temporal point process. *Journal of Statistical Planning and Inference* 128, 79–93.
- Shirai, T. and Y. Takahashi (2003). Random point fields associated with certain Fredholm determinants. I. Fermion, Poisson and boson point processes. *Journal of Functional Analysis* 2, 414–463.
- Soshnikov, A. (2000). Determinantal random point fields. *Russian Mathematical Surveys* 55, 923–975.
- Stein, E. and G. Weiss (1971). *Introduction to Fourier Analysis on Euclidean Spaces*. Princeton: Princeton University Press.
- Stoyan, D., W. S. Kendall, and J. Mecke (1995). *Stochastic Geometry and Its Applications* (Second ed.). Chichester: Wiley.
- Strauss, D. J. (1975). A model for clustering. *Biometrika* 63, 467–475.
- Wu, Z. (1995). Compactly supported positive definite radial functions. *Advances in Computational Mathematics* 4, 283–292.
- Yaglom, A. M. (1987). *Correlation Theory of Stationary and Related Random Functions*. New York: Springer-Verlag.