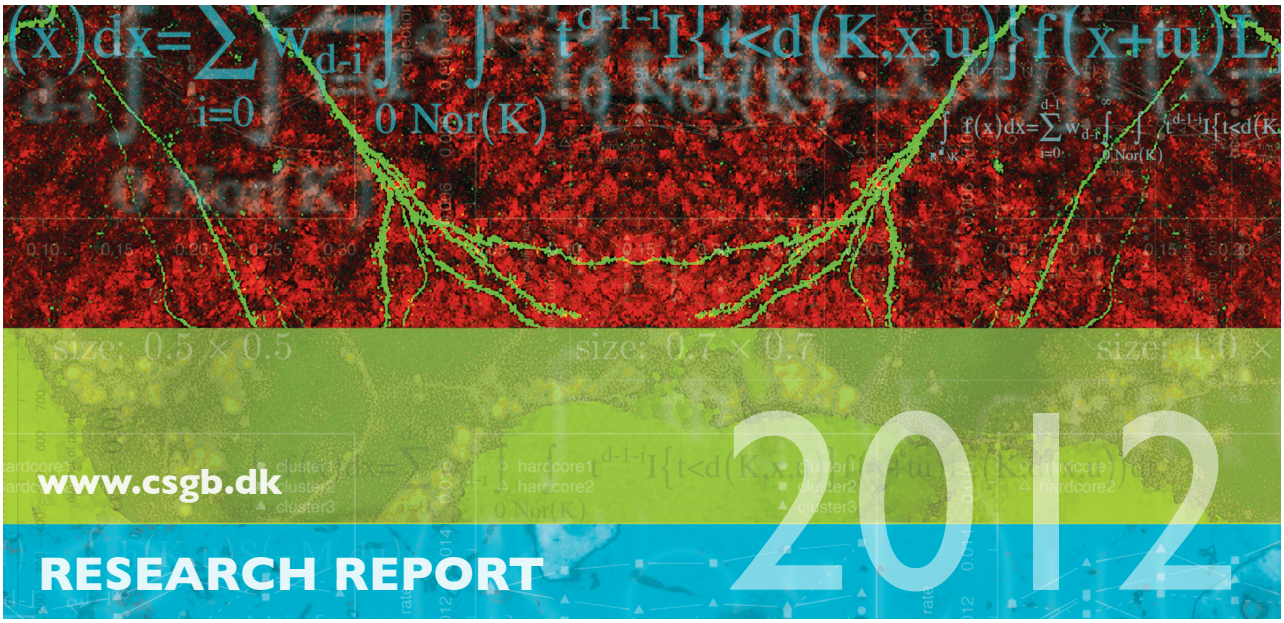
CENTRE FOR **STOCHASTIC GEOMETRY**
AND ADVANCED **BIOIMAGING**



Frédéric Lavancier, Jesper Møller and Ege Rubak

Statistical aspects of determinantal point processes

No. 04, May 2012

Statistical aspects of determinantal point processes

Frédéric Lavancier[1], Jesper Møller[2] and Ege Rubak*[2]

[1] Laboratoire de Mathématiques Jean Leray, University of Nantes, France,
Frederic.Lavancier@univ-nantes.fr
[2]Department of Mathematical Sciences, Aalborg University, jm@math.aau.dk,
rubak@math.aau.dk

Abstract

The statistical aspects of determinantal point processes (DPPs) seem largely unexplored. We review the appealing properties of DDPs, demonstrate that they are useful models for repulsiveness, detail a simulation procedure, and provide freely available software for simulation and statistical inference. We pay special attention to stationary DPPs, where we give a simple condition ensuring their existence, construct parametric models, describe how they can be well approximated so that the likelihood can be evaluated and realizations can be simulated, and discuss how statistical inference is conducted using the likelihood or moment properties.

Keywords: maximum likelihood based inference, point process density, product densities, simulation, repulsiveness, spectral approach.

1 Introduction

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, where the two most important cases are the d-dimensional Euclidean space $\mathbb{R}^d$ and a discrete state space. Hough et al. (2009) provides an excellent survey on the probabilistic aspects of DPPs.

The focus in the present paper is on statistical aspects for DPPs defined on a Borel set $B \subseteq \mathbb{R}^d$, with $d = 2$ and $B = \mathbb{R}^2$ in most of our examples, and with

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

its distribution specified by a continuous complex covariance function C defined on $B \times B$ and which is properly scaled (these regularity conditions on C are imposed to ensure existence of the process as discussed in Section 2.2). DPPs possess a number of appealing probabilistic properties:

(a) By the very definition, all orders of moments of a DPP defined on B are described by certain determinants of matrices with entries given by C (Section 2.1).

(b) The restriction of such a determinantal point process to a Borel subset A of B is also a DPP with its distribution specified by the restriction of C to $A \times A$.

(c) A one-to-one smooth transformation or an independent thinning of a DPP is also a DPP (Section 2.1).

(d) A DPP can easily be simulated, since it is a mixture of 'determinantal projection point processes' (Section 3).

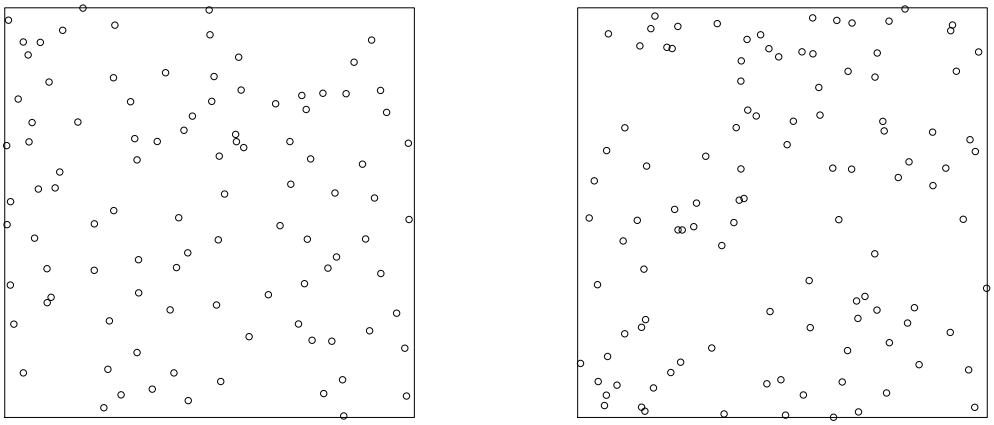
(e) A DPP restricted to a compact set has a density (with respect to a Poisson process) which is proportional to the determinant of a matrix with entries given by a complex covariance function $\tilde{C}$, where $\tilde{C}$ is obtained by a simple transformation of the eigenvalues in a spectral representation of C, and where the normalizing constant of the density has a closed form expression (Section 4).

From a statistical perspective, due to (a)–(e), modelling and estimation for DPPs become tractable. Moreover, we demonstrate that DPPs are useful models for repulsive interaction. The usual class of point processes used for modeling repulsiveness is the class of Gibbs point processes (see Møller and Waagepetersen (2004, 2007) and the references therein). For a general Gibbs point process, the moments are not expressible in closed form, the density involves an intractable normalizing constant, and rather elaborate Markov chain Monte Carlo methods are needed for simulations and approximate likelihood inference.

Figure 1 shows realizations in the unit square of two stationary DPPs with intensity $\rho = 100$. They are defined by a Gaussian covariance function with scale parameter (a) $\alpha = 0.05$ and (b) $\alpha = 0.01$ (see Section 5.2 for details). Compared to realizations of a homogeneous Poisson process, the point patterns in Figure 1 look more regular, and the regularity becomes more pronounced as α increases.

Generalizations of DPPs to weighted DPPs, which also are models for repulsion, and to the closely related permanental and weighted permanental point processes, which are models for attraction, 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. Section 5.3 discusses a useful approximation of C using a Fourier basis approach; this applies as well for weighted DPPs and weighted permanental point processes.

The paper is organized as follows. Section 2 defines DPPs and studies existence and probabilistic properties of them. Section 3 describes a general simulation procedure for DPPs. Section 4 discusses density expressions for DPPs. Section 5 studies



(a) (b)

Figure 1: Realizations of two Gaussian DPP models as described in Section 5.2 with intensity $\rho = 100$ and scale (a) $\alpha = 0.05$ and (b) $\alpha = 0.01$.

stationary DPPs and how to approximate them. Section 6 discusses statistical inference for parametric models of DPPs. Appendices A-D contain proofs of some of the results in the paper and provide supplementary methods and examples to the ones presented in the main text.

The statistical analyzes 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.

2 Basics

Section 2.1 discusses the definition of DPPs, while Section 2.2 discusses their existence.

2.1 Definition

Let $B \subseteq \mathbb{R}^d$ be a Borel set. Consider a simple locally finite spatial point process X on B, i.e. we can view X as a random locally finite subset 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. The following basic notions are needed before defining when X is a DPP.

Recall that for an integer $n > 0$, X has n'th order product density function $\rho^{(n)} : B^n \to [0, \infty)$ if this function is locally integrable (with respect to Lebesgue measure restricted to B) and for any Borel function $h : B^n \to [0, \infty)$,

$$\mathrm{E} \sum_{x_1,\ldots,x_n \in X}^{\neq} h(x_1, \ldots, x_n) = \int_B \cdots \int_B \rho^{(n)}(x_1, \ldots, x_n) h(x_1, \ldots, x_n) \, \mathrm{d}x_1 \cdots \mathrm{d}x_n \quad (2.1)$$

where $\neq$ over the summation sign means that $x_1, \ldots, x_n$ are pairwise distinct events. See e.g. Stoyan et al. (1995). Intuitively, for any pairwise distinct points $x_1, \ldots, x_n \in B$, $\rho^{(n)}(x_1, \ldots, x_n)\,\mathrm{d}x_1 \cdots \mathrm{d}x_n$ is the probability that for each $i = 1, \cdots, n$, X has a point in an infinitesimally small region around x_i of volume $\mathrm{d}x_i$. Clearly, $\rho^{(n)}$ is only uniquely defined up to a Lebesgue nullset. We shall henceforth require that $\rho^{(n)}(x_1, \ldots, x_n) = 0$ if $x_i = x_j$ for some $i \neq j$. This convention becomes consistent with Definition 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. In fact, for disjoint bounded Borel sets $A_1, A_2 \subset \mathbb{R}^d$, if $N(A)$ denotes the number of events falling in A, then the covariance between $N(A_1)$ and $N(A_2)$ is the integral over $A_1 \times A_2$ of the covariance function given by $c(x, y) = \rho(x)\rho(y)(g(x, y) - 1)$ for $x \neq y$. For a Poisson point process with an intensity function ρ, and for $x \neq y$, we have $c(x, y) = 0$, while $g(x, y) = 1$ if $\rho(x) > 0$ and $\rho(y) > 0$. In spatial statistics and stochastic geometry, g is more commonly used than c, and we shall also pay attention to g.

Let $\mathbb{C}$ denote the complex plane. For a complex number $z = z_1 + \mathrm{i}z_2$ (where $z_1, z_2 \in \mathbb{R}$ and $\mathrm{i} = \sqrt{-1}$), we denote $\overline{z} = z_1 - \mathrm{i}z_2$ the complex conjugate and $|z| = \sqrt{z_1^2 + z_2^2}$ the modulus.

For any function $C : B \times B \to \mathbb{C}$, let $[C](x_1, \ldots, x_n)$ be the $n \times n$ matrix with (i, j)'th entry $C(x_i, x_j)$. 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*

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

Then X is called a determinantal point process (DPP) *with kernel C, and we write $X \sim DPP_B(C)$.*

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

Remark 2.2. The focus in this paper is mostly on the case $B = \mathbb{R}^d$ and sometimes on the case $B = S$ where S is compact.

For $X \sim \mathrm{DPP}_B(C)$ and any Borel set $A \subseteq B$, define $X_A = X \cap A$ and denote its distribution by $\mathrm{DPP}_B(C; A)$. We also write $\mathrm{DPP}_A(C)$ for the distribution of the DPP on A with kernel given by the restriction of C to $A \times A$. Then property (b) in Section 1 follows directly from Definition 2.1, i.e. $\mathrm{DPP}_A(C) = \mathrm{DPP}_B(C; A)$. Further, when $B = \mathbb{R}^d$, we write $\mathrm{DPP}(C)$ for $\mathrm{DPP}_{\mathbb{R}^d}(C)$, and $\mathrm{DPP}(C; A)$ for $\mathrm{DPP}_{\mathbb{R}^d}(C; A)$.

Before discussing conditions on C ensuring the existence of DPPs, we notice the following properties.

Suppose $X \sim \mathrm{DPP}(C)$. 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

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

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 $\rho^{(n)}(x_1, \ldots, x_n)$ 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 (repulsiveness is also discussed in Remark 4.3).

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 $C(y,y) = 0$. When C is a covariance function, R is its corresponding correlation function and

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

The following propositions concern smooth transformations and independent thinning of DPPs.

Proposition 2.3. *Let $U \subseteq \mathbb{R}^d$ be an open set and $T : \mathbb{R}^d \to U$ a continuous differentiable bijection such that its inverse T^{-1} has a non-zero Jacobian determinant $J_{T^{-1}}(x)$ for all $x \in U$. If $X_1 \sim DPP(C_1)$ and $X_2 = T(X_1)$, then $X_2 \sim DPP_U(C_2)$ with*

$$C_2(x,y) = |J_{T^{-1}}(x)|^{1/2} C_1(T^{-1}(x), T^{-1}(y)) |J_{T^{-1}}(y)|^{1/2}. \tag{2.5}$$

Proof. Follows immediately from (2.1) and (2.2). □

Proposition 2.4. *If $X_1 \sim DPP(C_1)$ and X_2 is obtained as an independent thinning of X_1 with retention probabilities $p(x)$, $x \in \mathbb{R}^d$, then $X_2 \sim DPP(C_2)$ 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 $\mathrm{P}(U(x) = 1) = p(x)$ 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 $X_2 \sim \mathrm{DPP}(C_2)$. □

2.2 Existence

Existence of a DPP on $\mathbb{R}^d$ is ensured by the following assumptions on C, where $S \subset \mathbb{R}^d$ denotes a generic compact set.

Let $C : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{C}$ be Hermitian, i.e. $C(x,y) = \overline{C(y,x)}$ for all $x, y \in \mathbb{R}^d$. For convenience, assume that C is continuous. Denote $L^2(S)$ the space of square-integrable functions $h : S \to \mathbb{C}$, and define the integral operator $T_S : L^2(S) \to L^2(S)$ by

$$T_S(h)(x) = \int_S C(x,y) h(y) \, \mathrm{d}y, \quad x \in S. \tag{2.6}$$

By Mercer's theorem (see e.g. Section 98 in Riesz and Sz.-Nagy (1990)), for any compact set $S \subset \mathbb{R}^d$, C restricted to $S \times S$ has a spectral representation,

$$C(x,y) = \sum_{k=1}^{\infty} \lambda_k \phi_k(x) \overline{\phi_k(y)}, \quad (x,y) \in S \times S, \tag{2.7}$$

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.

$$T_S(\phi_k) = \lambda_k \phi_k, \quad \int_S \phi_k(x)\overline{\phi_l(x)}\,\mathrm{d}x = \begin{cases} 1 & \text{if } k = l, \\ 0 & \text{if } k \neq l, \end{cases} \tag{2.8}$$

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

When we need to stress that the eigenvalue λ_k depends on S, we write λ_k^S. We say that C (or T_S) is of local trace class if $\mathrm{tr}_S(C) = \int_S C(x,x)\,\mathrm{d}x$ is finite, i.e.

$$\mathrm{tr}_S(C) = \sum_{k=1}^{\infty} |\lambda_k^S| < \infty \quad \text{for all compact } S \subset \mathbb{R}^d. \tag{2.9}$$

Finally, we introduce the following conditions (C1) and (C2), recalling that C is a complex covariance function if and only if it is Hermitian and non-negative definite:

(C1) C is a continuous complex covariance function;
(C2) $\lambda_k^S \leq 1$ for all compact $S \subset \mathbb{R}^d$ and all k.

Theorem 2.5. *Under (C1), existence of DPP(C) is equivalent to (C2).*

Proof. A slightly different result where the eigenvalues are strictly less than one was first given in Theorem 12 of Macchi (1975). We apply Theorem 4.5.5 in Hough et al. (2009), where $C : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{C}$ is Hermitian, locally square integrable, of local trace class, and, as (2.7) may not hold on a Lebesgue nullset, that C is simply given by (2.7); then existence of DPP(C) is equivalent to that the spectrum of T_S is contained in $[0,1]$ for all compact $S \subset \mathbb{R}^d$ (i.e. $0 \leq \lambda_k^S \leq 1$, $k = 1, 2, \ldots$). 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 $\sum_{k=1}^{\infty} |\lambda_k^S| = \sum_{k=1}^{\infty} \lambda_k^S = \int_S C(x,x)\,\mathrm{d}x$ is finite. Thereby Theorem 2.5 follows. □

Usually, for statistical models of covariance functions, (C1) is satisfied, and so (C2) becomes the essential condition. As discussed in Section 5.1, (C2) simplifies in the stationary case of X.

Assumption 2.6. *In the remainder of this paper, $X \sim$ DPP(C) with C satisfying the conditions (C1) and (C2).*

Remark 2.7. 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 (2.7). We therefore confine ourselves to the Hermitian case of C.

Theorem 4.5.5 in Hough et al. (2009) deals with a more general setting for the existence of DPPs, where C satisfies the technical conditions given in our proof of Theorem 2.5. Then C is not assumed to be continuous, while non-negative definiteness of C becomes a necessary condition for existence of DPP(C), and so C has to be a covariance function. However, we find that (C1) is a natural condition for several reasons. First, statisticians are used to deal with covariance functions. Second, as seen in the proof of Theorem 2.5, the situation simplifies when C is assumed to be continuous. Third, 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. All in all, we therefore confine ourselves to the case of (C1).

Though the Poisson process is determinantal from Definition 2.1 (taking $C(x, y) = 0$ if $x \neq y$), it is neither covered by our approach nor by that in Hough et al. (2009). It actually corresponds to a limiting case of integral operators, where the limit is the multiplication operator Q given by $Q(h)(x) = C(x, x)h(x)$.

Remark 2.8. In continuation of Remark 2.2, 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 3-4 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 5-6 consider stationary DPPs.

Remark 2.9. Recall that for any simple locally finite spatial point process Y on $\mathbb{R}^d$ with intensity function ρ, there exist unique so-called reduced Palm distributions $\mathrm{P}_x^!$ for Lebesgue almost all $x \in \mathbb{R}^d$ with $\rho(x) > 0$, and the reduced Palm distributions are determined by that

$$\mathrm{E} \sum_{x \in Y} h(x, Y \setminus \{x\}) = \int \int \rho(x) h(x, \boldsymbol{x}) \, \mathrm{dP}_x^!(\boldsymbol{x}) \, \mathrm{d}x \tag{2.10}$$

for any non-negative Borel function h, where $\boldsymbol{x}$ 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, $\mathrm{P}_x^!$ 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, $n = 1, 2, \ldots$, then for Lebesgue almost all $x \in \mathbb{R}^d$ with $\rho(x) > 0$, $\mathrm{P}_x^!$ has n'th order product density function

$$\rho_x^{(n)}(x_1, \ldots, x_n) = \rho^{(n+1)}(x, x_1, \ldots, x_n)/\rho(x) \tag{2.11}$$

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

For our DPP X, using (2.2) it can be shown that for all $x \in \mathbb{R}^d$ with $C(x, x) > 0$, we can take $\mathrm{P}_x^! = \mathrm{DPP}(C_x^!)$ 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, (2.11) holds whenever $C(x, x) > 0$.

Let $S \subset \mathbb{R}^d$ be compact. A special simple case of X_S occurs when all non-zero eigenvalues λ_k^S are one, or equivalently, T_S is a projection. Then we call the restriction of C to $S \times S$ a projection kernel. McCullagh and Møller (2006) calls then X_S a special DPP, but in the present paper we use a more commonly used terminology (e.g. Hough et al. (2006) and Hough et al. (2009)).

Definition 2.10. *When C restricted to $S \times S$ is a projection kernel, X_S is called a* determinantal projection point process*.*

3 Simulation

A general algorithm for simulating a finite DPP is provided in Hough et al. (2006), including a proof of its validity in a very general setup. While it is the only known general simulation procedure for DPPs, there are special cases 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 general simulation algorithm in the specific case where we want to simulate $X_S \sim \mathrm{DPP}(C; S)$ with $S \subset \mathbb{R}^d$ compact. Our implementation of the general algorithm becomes more efficient than the one in Scardicchio et al. (2009), and our exposition uses mainly linear algebra and is less technical than the exposition in Hough et al. (2006).

Consider the spectral representation (2.7) of C restricted to $S \times S$. The general simulation algorithm comes from the following result proved in Theorem 7 in Hough et al. (2006); see also Theorem 4.5.3 in Hough et al. (2009).

Theorem 3.1. *For $k = 1, 2, \ldots$, let B_k be independent Bernoulli variables with mean λ_k. Define the random projection kernel $K : S \times S \to \mathbb{C}$ by*

$$K(x, y) = \sum_{k=1}^{\infty} B_k \phi_k(x) \overline{\phi_k(y)}. \tag{3.1}$$

Then

$$DPP_S(K) \sim DPP(C; S) \tag{3.2}$$

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

$$N(S) \sim \sum_{k=1}^{\infty} B_k, \quad \mathrm{E}[N(S)] = \sum_{k=1}^{\infty} \lambda_k, \quad \mathrm{Var}[N(S)] = \sum_{k=1}^{\infty} \lambda_k(1 - \lambda_k). \tag{3.3}$$

The first result in (3.3) follows from (3.2) and Theorem 3.2 below (or from Lemma 4.4.1 in Hough et al. (2009)), and the first result immediately implies the two other results.

In the following two sections, a two step simulation procedure based on Theorem 3.1 is described in detail.

3.1 Simulation of Bernoulli variables

We start by describing how the independent Bernoulli variables $B_1, B_2, \ldots$ can be simulated.

Recall that $\mathrm{P}(B_k = 1) = 1 - \mathrm{P}(B_k = 0) = \lambda_k$, $k = 1, 2, \ldots$. Define $B_0 = \lambda_0 = 1$. With probability one, $\sum B_k < \infty$, since $\sum \lambda_k < \infty$. 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 $B_0, \ldots, B_{m-1}$ 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 $k = 1, \ldots, m-1$ (if $m = 0$ we do nothing), and third set $B_m = 1$ and $B_k = 0$ for $k = m+1, m+2, \ldots$. Simulation of these Bernoulli variables is of course easily done. For simulation of M, we use the inversion method described below.

For $m = 0, 1, 2, \ldots$, 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 $m \geq m'$, computation of the p_m's may be done via the recursion

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

The calculation of $p_{m'}$ may involve numerical methods.

Let F denote the distribution function of M and introduce

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

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

3.2 Simulation of determinantal projection point process

Suppose we have generated a realization of the Bernoulli variables B_k as described in Section 3.1 and we now want to generate a realization from $\mathrm{DPP}_S(K)$ with K given by (3.1).

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 (3.3), n can be considered as a realization of the count $N(S)$). If $n = 0$, then $K = 0$ and a realization from $\mathrm{DPP}_S(K)$ is simply equal to the empty point configuration. Assume that $n > 0$ and without loss of generality that

$$K(x, y) = \sum_{k=1}^{n} \phi_k(x)\overline{\phi_k(y)} = \boldsymbol{v}(y)^* \boldsymbol{v}(x) \tag{3.4}$$

where $\boldsymbol{v}(x) = (\phi_1(x), \ldots, \phi_n(x))^T$, and where ^{T} and * denote the transpose and conjugate transpose of a vector or a matrix. For n-dimensional complex column

vectors such as $\boldsymbol{v}(x)$ and $\boldsymbol{v}(y)$, we consider their usual inner product $\langle \boldsymbol{v}(x), \boldsymbol{v}(y) \rangle = \boldsymbol{v}(y)^* \boldsymbol{v}(x)$.

Algorithm 1 Simulation of determinantal projection point process

sample X_n from the distribution with density $p_n(x) = \|\boldsymbol{v}(x)\|^2/n$, $x \in S$
set $\boldsymbol{e}_1 = \boldsymbol{v}(X_n)/\|\boldsymbol{v}(X_n)\|$
for $i = (n-1)$ to 1 **do**
  sample X_i from the distribution with density

$$p_i(x) = \frac{1}{i}\left[\|\boldsymbol{v}(x)\|^2 - \sum_{j=1}^{n-i} |\boldsymbol{e}_j^* \boldsymbol{v}(x)|^2\right], \quad x \in S \tag{3.5}$$

  set $\boldsymbol{w}_i = \boldsymbol{v}(X_i) - \sum_{j=1}^{n-i} \left(\boldsymbol{e}_j^* \boldsymbol{v}(X_i)\right) \boldsymbol{e}_j$, $\boldsymbol{e}_{n-i+1} = \boldsymbol{w}_i/\|\boldsymbol{w}_i\|$
end for
return $\{X_1, \ldots, X_n\}$

Theorem 3.2. *If $n > 0$ and $K(x,y) = \sum_{k=1}^n \phi_k(x)\overline{\phi_k(y)}$ for all $x, y \in S$, then $\{X_1, \ldots, X_n\}$ generated by Algorithm 1 is distributed as $DPP_S(K)$.*

Proof. See Appendix A. □

Remark 3.3. Let $n > 0$, and define $H_n = \{\boldsymbol{0}\}$ and for $i = n-1, \ldots, 1$,

$$H_i = \mathrm{span}_{\mathbb{C}}\{\boldsymbol{v}(X_n), \ldots, \boldsymbol{v}(X_{i+1})\} = \left\{ \sum_{j=i+1}^{n} \alpha_j \boldsymbol{v}(X_j) : \alpha_j \in \mathbb{C} \right\}. \tag{3.6}$$

With probability one, $\boldsymbol{v}(X_n), \ldots, \boldsymbol{v}(X_i)$ are linearly independent, cf. Appendix A. Thus, almost surely, H_i is a subspace of $\mathbb{C}^n$ of dimension $n-i$. For $i = n-1, \ldots, 1$, by the Gram-Schmidt procedure employed in Algorithm 1, $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_{n-i}$ is an orthonormal basis of H_i. Further, for $i = n, \ldots, 1$, $ip_i(x)$ is the square norm of the orthogonal projection of $\boldsymbol{v}(x)$ onto $H_i^\perp$ (the orthogonal complement to H_i). Moreover, as verified in Appendix A, with probability one, $p_i(x)$ becomes a density, where for $i < n$ we are conditioning on $(X_n, \ldots, X_{i+1})$.

Remark 3.4. According to the previous remark,

$$ip_i(x) = \|P_i \boldsymbol{v}(x)\|^2 \tag{3.7}$$

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 $i < n$,

$$P_i = \prod_{k=n}^{i+1} \left(I_n - \frac{\boldsymbol{v}(x_k)\boldsymbol{v}(x_k)^*}{K(x_k, x_k)} \right). \tag{3.8}$$

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 3.5. To implement Algorithm 1 we need a way to sample from the densities p_i, $i = n, \ldots, 1$. This may simply be done by rejection sampling with a uniform instrumental density and acceptance probability $p_i(x)/\sup_{y \in S} p_i(y)$. Note that for x such that $\boldsymbol{v}(x) \in H_i^{\perp}$, $p_i(x) = \|\boldsymbol{v}(x)\|^2/i$. 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 B discusses other choices of the instrumental distribution.

4 Densities

This section briefly discusses some useful density expressions for $X_S \sim \mathrm{DPP}(C; S)$ when $S \subset \mathbb{R}^d$ is compact. Recall that the eigenvalues $\lambda_k = \lambda_k^S$ are 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 3.1, or just condition on $K(x, y)$ for all $x, y \in S$, to obtain the conditional density. Note that the trace $\mathrm{tr}_S(K) = \int_S K(x, x)\,\mathrm{d}x = \sum_{k=1}^{\infty} B_k$ is almost surely finite. Conditional on K, when $\mathrm{tr}_S(K) = n > 0$, the ordered n-tuple of events of the determinantal projection point process X_S has density

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

as verified in (A.2). Moreover, by Algorithm 1 and Theorem 3.2,

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

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.

Assume that $\lambda_k < 1$ for all $k = 1, 2, \ldots$, which means that no B_k is almost surely one. Then the density of X_S exists and is specified in Theorem 4.1 below, where the following considerations and notation are used. If $P(N(S) = n) > 0$, then $P(N(S) = m) > 0$ for $m = 0, \ldots, n$, cf. (3.3). Thus

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

is strictly positive, and we can define

$$D = -\log \mathrm{P}(N(S) = 0) = -\sum_{k=1}^{\infty} \log(1 - \lambda_k). \tag{4.1}$$

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

$$\tilde{C}(x, y) = \sum_{k=1}^{\infty} \tilde{\lambda}_k \phi_k(x) \overline{\phi_k(y)} \tag{4.2}$$

where

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

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

Theorem 4.1. *Assuming $\lambda_k < 1$, $k = 1, 2, \ldots$, then X_S is absolutely continuous with respect to the homogeneous Poisson process on S with unit intensity, and has density*

$$f(\{x_1, \ldots, x_n\}) = \exp(|S| - D) \det[\tilde{C}](x_1, \ldots, x_n) \tag{4.3}$$

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

Proof. This was first verified in Macchi (1975). Note that the right hand side in (4.3) is not depending on the ordering of the events. Equation (4.3) follows from a longer but in principle straightforward calculation, using (3.2), (A.3), and the fact that if Y follows the homogeneous Poisson process on S with unit intensity, then

$$\rho^{(n)}(x_1, \ldots, x_n) = \mathrm{E} f(Y \cup \{x_1, \ldots, x_n\}).$$

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

Remark 4.2. It is possible to express $\tilde{C}$ and D in terms of C without any direct reference to the spectral representations (2.7) and (4.2): Let

$$C_S^1(x, y) = C_S(x, y), \quad C_S^k(x, y) = \int_S C_S^{k-1}(x, z) C_S(z, y) \, \mathrm{d}z, \quad x, y \in S, \ k = 2, 3, \ldots. \tag{4.4}$$

Then

$$D = \sum_{k=1}^{\infty} \mathrm{tr}_S(C_S^k)/k \tag{4.5}$$

and

$$\tilde{C}(x, y) = \sum_{k=1}^{\infty} C_S^k(x, y), \quad x, y \in S. \tag{4.6}$$

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

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

Section 5.3 and Appendix D discuss efficient ways of approximating $\tilde{C}$ and D when X is stationary.

Remark 4.3. The density (4.3) is hereditary in the sense that $f(\{x_1, \ldots, x_n\}) > 0$ whenever $f(\{x_1, \ldots, x_{n+1}\}) > 0$. This allows us to define the Papangelou conditional intensity for all finite point configurations $\boldsymbol{x} = \{x_1, \ldots, x_n\} \subset S$ and points $u \in S \backslash \boldsymbol{x}$ by

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

(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 $\boldsymbol{x} \subset S$ and $\boldsymbol{y} \subset S$,

$$\lambda(u; \boldsymbol{x}) \geq \lambda(u; \boldsymbol{y}) \quad \text{whenever } \boldsymbol{x} \subset \boldsymbol{y} \tag{4.7}$$

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

$$\lambda(u; \boldsymbol{x}) \leq \tilde{C}(u, u) \tag{4.8}$$

(Theorem 3.1 in Georgii and Yoo (2005)).

The monotonicity property (4.7) is once again confirming the repulsiveness of a DPP.

Equation (4.8) means that X_S is locally stable and hence that 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 $S_1 \subset S_2 \subset \ldots$ 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 $X_n \sim \mathrm{DPP}(C; S_n)$ 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 $X_1 \subseteq X_2 \subset \ldots \subseteq Y$, we obtain that $\cup_n X_n \subseteq Y$ follows $\mathrm{DPP}(C)$. 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 $d \geq 2$.

5 Stationary models

Suppose $X \sim \mathrm{DPP}(C)$ is stationary, i.e. its distribution is invariant under translations, or equivalently, C is of the form

$$C(x, y) = C_0(x - y), \quad x, y \in \mathbb{R}^d. \tag{5.1}$$

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 2.3–2.4, as inhomogeneous DPPs can be obtained by transforming or thinning X, stationarity is not a very restrictive assumption. For example, by (2.5), 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

$$C_{\mathrm{trans}}(x, y) = |J_{T^{-1}}(x)|^{1/2} C_0(T^{-1}(x) - T^{-1}(y)) |J_{T^{-1}}(y)|^{1/2}. \tag{5.2}$$

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 $\mathbb{R}^d$. 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

$$|R_0(r)| = \sqrt{1 - g_0(r)} \tag{5.3}$$

we can introduce the '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 (5.13).

Isotropy is also a natural simplification, since in the stationary case, any anisotropic covariance function can be obtained from some isotropic covariance function using some rotation followed by some rescaling, see e.g. Goovaerts (1997). Examples of stationary isotropic covariance functions are studied in Section 5.2. However, the following Section 5.1 does not involve an assumption of isotropy, and the approximation of C_0 studied in Section 5.3 is only approximately isotropic when C_0 is isotropic.

5.1 A simple spectral condition for existence

The following Proposition 5.1 concerns the meaning of (C2) in terms of the spectral density for C_0. We start by recalling what the spectral density is.

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

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

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

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

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

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

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 $\mathcal{F} : L^2(\mathbb{R}^d) \to L^2(\mathbb{R}^d)$ and $\mathcal{F}^{-1} : L^2(\mathbb{R}^d) \to L^2(\mathbb{R}^d)$. 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

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

If F is differentiable, then the derivative $\varphi(x) = \mathrm{d}F(x)/\mathrm{d}x$ is called the spectral density, and 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.

Proposition 5.1. *Under (C1) and* (5.1)*, if* $C_0 \in L^2(\mathbb{R}^d)$*, then (C2) is equivalent to that*

$$\varphi \leq 1. \tag{5.4}$$

Proof. Consider any compact set $S \subset \mathbb{R}^d$. 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 (5.1), the integral operator T_S associated to C on $L^2(S)$ (see (2.6)) becomes the convolution operator given by

$$T_S(h)(x) = C_0 \star h_S(x) = \int_S C_0(x-y)h(y)\,\mathrm{d}y, \quad x \in S.$$

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 ess-im(φ_S) (the essential image of φ_S), see (12) in Section 8.4.3 in Birman and Solomjak (1987). In our case, ess-im(φ_S) is the closure of $\varphi(S)$. Consequently, (C2) is equivalent to $\varphi \leq 1$.

□

Assumption 5.2. *Henceforth, in addition to (C1), we assume that* $C_0 \in L^2(\mathbb{R}^d)$ *and that* (5.4) *holds.*

The following corollary becomes useful in Section 5.4 where we discuss a spectral approach for constructing stationary DPPs.

Corollary 5.3. *Under* (5.1) *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 $C_0 \in L^2(\mathbb{R}^d)$.

Proof. Assume (i). Then $0 \leq \varphi \leq 1$ implies that $\int |\varphi(x)|^2\,\mathrm{d}x \leq \int |\varphi(x)|\,\mathrm{d}x < \infty$, i.e. $\varphi \in L^2(\mathbb{R}^d)$, and so by Parsevals 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). □

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 (5.4). As exemplified in Section 5.2, $\rho_{\max}$ 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 4.1, we may require that $\varphi < 1$.

5.2 Examples of covariance models

Numerous examples of stationary isotropic covariance functions exist (see e.g. Gelfand et al., 2010), while examples of stationary anisotropic covariance functions are discussed in De laco 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 5.4 discusses further examples based on a spectral approach.

It may be appealing to construct isotropic covariance functions $C_0(x)$, where the range

$$\delta = \sup\{\|x\| : C_0(x) \neq 0\} \tag{5.5}$$

is finite. Examples of such finite range covariance functions are given in Wu (1995) and Gneiting (2002). By Definition 2.1, if $A, B \subset \mathbb{R}^d$ are separated by a distance larger than δ, then X_A and X_B are independent DPPs. Let $d = 2$ and consider the circular covariance function with finite range $\delta > 0$ and given by

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

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 (5.4), a DPP with kernel C_0 exists if $0 \leq \rho \leq \rho_{\max}$, where $\rho_{\max} = 4/(\pi\delta^2)$. Therefore, we require

$$\rho\delta^2 \leq 4/\pi. \tag{5.6}$$

In this paper we only use the circular covariance function to understand well the quality of our approximations in Section 5.3.

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 $\mathrm{E}(W^{-d/2}) < \infty$, where Z and W are independent. Then $Y = \sqrt{W}Z$ follows a normal-variance mixture distribution, with density

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

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

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

which is positive, showing that C_0 is a stationary isotropic covariance function. Note that φ is given by the Laplace transform of W. By (5.4), a stationary DPP with kernel C_0 exists if $0 \le \rho \le \rho_{\max}$, where ρ is the intensity and

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

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

$$C_0(x) = \rho \exp\left(-\|x/\alpha\|^2\right), \quad x \in \mathbb{R}^d, \tag{5.7}$$

and

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

Hence

$$\rho_{\max} = (\sqrt{\pi}\alpha)^{-d} \tag{5.8}$$

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 C). Hence

$$C_0(x) = \rho\, \frac{2^{1-\nu}}{\Gamma(\nu)} \|x/\alpha\|^\nu K_\nu(\|x/\alpha\|), \quad x \in \mathbb{R}^d, \tag{5.9}$$

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

$$\rho_{\max} = \frac{\Gamma(\nu)}{\Gamma(\nu + d/2)(2\sqrt{\pi}\alpha)^d} \tag{5.10}$$

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

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

is the density of a multivariate t-distribution, and

$$C_0(x) = \frac{\rho}{\left(1 + \|x/\alpha\|^2\right)^{\nu+d/2}}, \quad x \in \mathbb{R}^d, \tag{5.11}$$

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

$$\rho_{\max} = \frac{\Gamma(\nu + d/2)}{\Gamma(\nu)(\sqrt{\pi}\alpha)^d} \tag{5.12}$$

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

For later use, notice that the Gaussian covariance function (5.7) with $\alpha = 1/\sqrt{\pi\rho}$ is the limit of both

(i) the Whittle-Matérn covariance function (5.9) with $\alpha = 1/\sqrt{4\pi\nu\rho}$, and

(ii) the Cauchy covariance function (5.11) with $\alpha = \sqrt{\nu/(\pi\rho)}$

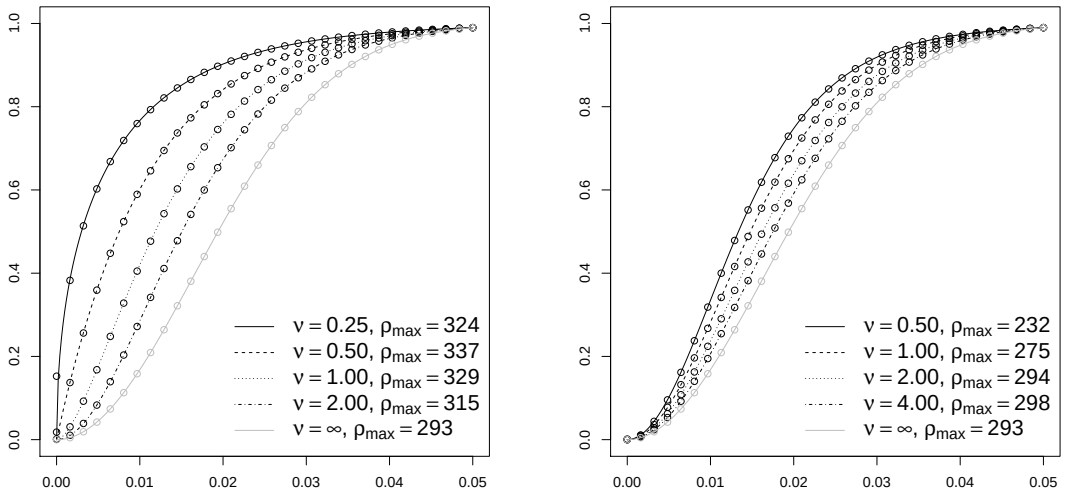
as $\nu \to \infty$.

We refer to a DPP model with kernel (5.7), (5.9), or (5.11) as the Gaussian, Whittle-Matérn, or Cauchy model, respectively. In all three models, α is a scale parameter of C_0. For the Whittle-Matérn and Cauchy models, ν is a shape parameter of C_0. The isotropic pair correlation functions are

for the Gaussian model:	$g_0(r) = 1 - \exp\left(-2(r/\alpha)^2\right), \quad r \geq 0;$
for the Whittle-Matérn model:	$g_0(r) = 1 - \left[2^{1-\nu}(r/\alpha)^\nu K_\nu(r/\alpha)/\Gamma(\nu)\right]^2, \quad r \geq 0;$
for the Cauchy model:	$g_0(r) = 1 - \left[1 + (r/\alpha)^2\right]^{-2\nu-d}, \quad r \geq 0.$

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 (5.3) 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 thereby

$$r_0 = \alpha\sqrt{-\log(0.1)}, \quad r_0 = \alpha\sqrt{8\nu}, \quad r_0 = \alpha\sqrt{0.1^{-1/(\nu+1)} - 1}, \tag{5.13}$$



(a) (b)

Figure 2: 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, and as a reference 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 $\rho_{\max}$ 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 5.3.

respectively. So r_0 depends linearly on α. Notice, when ν is fixed, the upper bound $\rho_{\max}$ on the intensity decreases as r_0 increases, since $\rho_{\max}$ is proportional to r_0^{-d}, cf. (5.8), (5.10), and (5.12). There is a similar trade-off between how large the intensity and the range of the circular covariance function can be, cf. (5.6).

Figure 2 shows examples of the isotropic pair correlation functions with a fixed range of correlation. In particular the Whittle-Matérn models are seen to constitute a quite flexible model class with several different shapes of pair correlation functions. From the figure it is also evident, that the value of $\rho_{\max}$ is of the same order of magnitude for all these models indicating that the range of interaction has a major effect on the maximal permissible intensity of the model.

For Ripley's K-function (Ripley, 1976, 1977),

$$K(r) = 2\pi \int_0^r t g_0(t)\,\mathrm{d}t, \quad r \geq 0, \tag{5.14}$$

we have

$$\text{for the Gaussian model:} \quad K(r) = \pi r^2 - \frac{\pi\alpha^2}{2}\left(1 - \exp\left(\frac{-2r^2}{\alpha^2}\right)\right);$$
$$\text{for the Cauchy model:} \quad K(r) = \pi r^2 - \frac{\pi\alpha^3}{2\nu + d - 1}\left(\frac{1}{\alpha} - \left(\frac{\alpha}{r^2+\alpha^2}\right)^{2\nu+d-1}\right);$$

while for the Whittle-Matérn model the integral in (5.14) has to be evaluated by numerical methods.

Recall that $\rho K(r)$ is the conditional expectation of the number of further points of X in a ball of radius r centered at x given that X has a point at x. For two DPPs $\mathrm{DPP}(C_1)$ and $\mathrm{DPP}(C_2)$ with common intensity ρ and corresponding K-functions K_1 and K_2, we say that $\mathrm{DPP}(C_1)$ exhibits stronger repulsiveness than $\mathrm{DPP}(C_2)$ 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

$$K_1 \leq K_2 \quad \text{if and only if} \quad g_{10} \leq g_{20}. \tag{5.15}$$

In this sense, within each class of the Gaussian, Whittle-Matérn, and Cauchy models, 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. (5.8), (5.10), and (5.12). For fixed ρ and ν, α has an upper limit $\alpha_{\max}$ given by

$$\alpha_{\max} = 1/\sqrt{\pi\rho}, \quad \alpha_{\max} = 1/\sqrt{4\pi\nu\rho}, \quad \alpha_{\max} = \sqrt{\nu/(\pi\rho)} \tag{5.16}$$

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) above. Moreover, we consider the variance stabilizing transformation of the K-function, $L(r) = \sqrt{K(r)/\pi}$ (Besag, 1977), and Figure 3 shows $L(r) - r$ for seven different models, which clearly illustrates the dependence between the degree of repulsiveness and ν.

5.3 Approximation

This section describes approximations of the distribution, kernel and density of the stationary DPP X restricted to a unit box $[-1/2, 1/2]^d$. Furthermore, it is explained

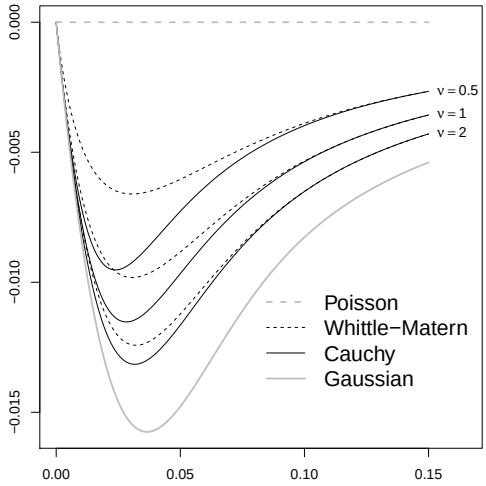


Figure 3: Plots of $L(r) - r$ vs. r for the Whittle-Matérn, Cauchy, and Gaussian model with $\alpha = \alpha_{\max}$ and $\rho = 100$. For the Whittle-Matérn and Cauchy models, $\nu \in \{0.5, 1, 2\}$. The horizontal line at zero is $L(r) - r$ for a stationary Poisson process.

how X can be approximately simulated on any rectangular set and how the density of X on such a set can be approximated. Throughout this section, $S = [-1/2, 1/2]^d$, $S/2 = [-1/4, 1/4]^d$, and $2S = [-1, 1]^d$.

Approximation of the kernel C

Consider the orthonormal Fourier basis of $L^2(S)$ given by

$$\phi_k(x) = \mathrm{e}^{2\pi \mathrm{i} k \cdot x}, \quad k \in \mathbb{Z}^d, \ x \in S, \tag{5.17}$$

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

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

where

$$\alpha_k = \int_S C_0(t) \mathrm{e}^{-2\pi \mathrm{i} k \cdot t} \, \mathrm{d}t. \tag{5.18}$$

Substituting the finite integral in (5.18) by the infinite integral

$$\varphi(k) = \int C_0(t) \mathrm{e}^{-2\pi \mathrm{i} k \cdot t} \, \mathrm{d}t \tag{5.19}$$

leads us to approximate C_0 on S by

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

So we consider the approximation

$$C_0(u) \approx C_{\mathrm{app},0}(u), \quad u \in S. \tag{5.20}$$

Since $x - y \in S$ if $x, y \in S/2$, this leads to the following approximation of the kernel C on $S/2 \times S/2$

$$C(x, y) \approx C_{\text{app}}(x, y), \quad x, y \in S/2, \tag{5.21}$$

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

Comparing (5.18) and (5.19) we see that the error of the approximations (5.20) and (5.21) 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 (5.5)), $C_0(t) = 0$ 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 (5.20) and (5.21) are then exact. For instance, considering the circular covariance function and the existence condition (5.6), we have $\delta < 1/2$ if $\rho > 16/\pi$, which indeed is not a restrictive requirement in practice.

Appendix C 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 5.2, and show that the error is small provided the intensity ρ is not too small. Furthermore, Figure 2 in Section 5.2 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

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

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 (5.20), we use the approximation $C_0^{\text{per}}(u) \approx C_{\text{app},0}^{\text{per}}(u)$, $u \in 2S$, where $C_{\text{app},0}^{\text{per}}(u) = \sum_{k \in \mathbb{Z}^d} \varphi(k) \mathrm{e}^{2\pi \mathrm{i} k \cdot u}$, $u \in 2S$, 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

$$C_{\text{app}}^{\text{per}}(x, y) = \sum_{k \in \mathbb{Z}^d} \varphi(k) \phi_k(x) \overline{\phi_k(y)}, \quad x, y \in S. \tag{5.22}$$

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

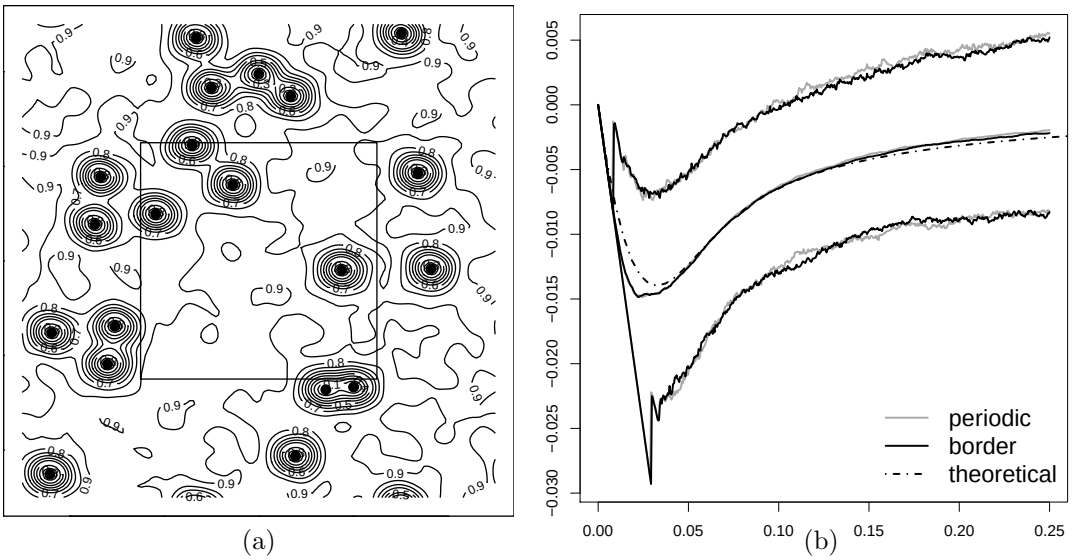
The border method

Since $\varphi \leq 1$, the DPP $X^{\text{per}} \sim \text{DPP}_S(C_{\text{app}}^{\text{per}})$ is well-defined. 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 (5.22) is of the form required for the simulation algorithm of Section 3. Recall that for finite range covariance functions with $\delta < 1/2$ (see (5.5)), 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. Then we define the affine transformation $T(x) = Ax + b$ such that $T(R) = S/2$. Then $Y = T(X)$ is a stationary DPP, with kernel given by (5.2) and spectral density $\varphi_Y(x) = \varphi(A^T x)$. Let Y^{per} be the DPP on S with kernel (5.22) 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.

The periodic method

From our practical experience it appears that $\text{DPP}_S(C^{\text{per}}_{\text{app}})$ 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 $C^{\text{per}}_{\text{app},0}$ mimics the influence of points outside S. To illustrate this, Figure 4(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 3.5) when X^{per} is simulated by the algorithm in Section 3. The qualitative behavior of the process in S and in the interior region $S/2$ are similar in the sense that there are regions at the border where the acceptance probability is low. For the process on $S/2$ this is due to the influence of points outside $S/2$ whereas for the process on S this influence is created artificially by points at the opposite border.



(a) (b)

Figure 4: (a) Acceptance probability for a uniformly distributed proposal at an intermediate step of the simulation algorithm of Section 3. The process being simulated is X^{per} on $S = [-1/2, 1/2]^d$ and the interior box corresponds to 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.

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).

For a Gaussian covariance function, Figure 4(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 4(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 4(b) were generated in approximately three minutes on a laptop with a dual core processor.

Approximation of the density f

First, consider the density f for X_S as specified in Theorem 4.1 (so we assume $\varphi < 1$). We use the approximation $f \approx f^{\text{per}}$, where f^{per} denotes the density of X^{per}. Letting

$$\tilde{\varphi}(u) = \varphi(u)/(1 - \varphi(u)), \quad u \in S, \tag{5.23}$$

$$\tilde{C}_{\text{app}}^{\text{per}}(x, y) = \tilde{C}_{\text{app},0}^{\text{per}}(x - y) = \sum_{k \in \mathbb{Z}^d} \tilde{\varphi}(k) \mathrm{e}^{2\pi \mathrm{i} k \cdot (x-y)}, \quad x, y \in S, \tag{5.24}$$

and

$$D_{\text{app}}^{\text{per}} = \sum_{k \in \mathbb{Z}^d} \log\left(1 + \tilde{\varphi}(k)\right) \tag{5.25}$$

we have

$$f^{\text{per}}(\{x_1, \ldots, x_n\}) = \exp(|S| - D_{\text{app}}^{\text{per}}) \det[\tilde{C}_{\text{app}}^{\text{per}}](x_1, \ldots, x_n), \quad \{x_1, \ldots, x_n\} \subset S. \tag{5.26}$$

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 $\tilde{C}^{\mathrm{per}}_{\mathrm{app},0}$. The details of the truncation and use of FFT are given in Section 6.1

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^{per}_Y denotes the approximate density of Y as specified by the right hand side of (5.26), we can approximate the density of X_R by

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

We call f^{per} the periodic approximation of f. The simulation study in Section 6 shows that likelihood inference based on f^{per} works well in practice for the examples in this paper. Appendix D introduces a convolution approximation of the density which in some cases may be computationally faster to evaluate. However, as discussed in Appendix D, this approximation appears to be poor in some situations and in general we prefer the periodic approximation.

Remark 5.4. It would have been desirable if we could compute explicitly the spectral representation (2.7) for a given parametric family of the covariance function 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 for instance 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 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 (5.22) 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 Section 5.2.

5.4 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 $\varphi : \mathbb{R}^d \to [0, 1]$, which becomes the spectral density, cf. Corollary 5.3. In fact knowledge about φ is all we need for the approximate simulation procedure and density approximation in Section 5.3. 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.

Quantifying repulsiveness

In Section 5.2, we said that $\mathrm{DPP}(C_1)$ exhibits stronger repulsiveness than $\mathrm{DPP}(C_2)$ if their intensities agree and their corresponding K-functions satisfy $K_1 \leq K_2$. Considering Figure 3, we cannot always use this concept when comparing a Whittle-Matérn model with a Cauchy model.

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

$$\mu = \rho \int [1 - g(x)] \, \mathrm{d}x$$

as a rough measure for repulsiveness provided the integral exists. Denote o the origin of $\mathbb{R}^d$ and note that the function $x \to \rho g(o, x) = \rho g(x)$ is the intensity function for the reduced Palm distribution $P_o^!$, cf. (2.11). Therefore, μ is the limit as $r \to \infty$ 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,

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

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 $\mathrm{DPP}(C_1)$ is more repulsive than $\mathrm{DPP}(C_2)$ if $\rho_1 = \rho_2$ and $\mu_1 \geq \mu_2$. In the isotropic case, this is in agreement with our former concept: if $\rho_1 = \rho_2$, then $K_1 \leq K_2$ implies that $\mu_1 \geq \mu_2$, cf. (5.15).

Suppose we are interested in a stationary DPP with intensity ρ and a maximal value of μ. Since $0 \leq \varphi(x)^2 \leq \varphi(x) \leq 1$, we have $\mu = 1$ if and only if $\int \varphi(x)^2 \, \mathrm{d}x = \int \varphi(x) \, \mathrm{d}x = \rho$. 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

$$\varphi(x) = \begin{cases} 1 & \text{if } \|x\| \leq r \\ 0 & \text{otherwise} \end{cases} \tag{5.27}$$

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

$$C_0(x) = \sin(\pi \rho x)/(\pi x) \quad \text{if } d = 1. \tag{5.28}$$

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

$$C_0(x) = \sqrt{\rho} J_1(2\sqrt{\pi \rho}\|x\|)/\|x\| \quad \text{if } d = 2. \tag{5.29}$$

A general class of spectral densities

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 than the Gaussian model which appears as a special case. In particular, as shown after (5.36) below, the extreme case (5.27) is a limiting case of this class.

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

$$c = \int_{\mathbb{R}^d} f(\|x\|) \, \mathrm{d}x = \frac{d\pi^{d/2}}{\Gamma(d/2 + 1)} \int_0^\infty r^{d-1} f(r) \, \mathrm{d}r. \tag{5.30}$$

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

$$\varphi(x) = \rho f(\|x\|)/c, \quad x \in \mathbb{R}^d, \tag{5.31}$$

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

$$\rho \leq \rho_{\max} = c/\sup f. \tag{5.32}$$

Below we give an example of a parametric model class for such functions f, where the integral in (5.30) and the supremum in (5.32) 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 (5.30) and (5.31),

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

and

$$\varphi(x) = \rho\frac{\Gamma(d/2+1)\nu\alpha^d}{d\pi^{d/2}\Gamma(\gamma + \frac{d-1}{\nu})}\|\alpha x\|^{\gamma\nu-1}\exp(-\|\alpha x\|^\nu). \tag{5.33}$$

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

$$\rho_{\max} = \frac{c}{f((\gamma - 1/\nu)^{1/\nu})} = \frac{d\pi^{d/2}\alpha^{-d}\Gamma(\gamma + \frac{d-1}{\nu})\exp(\gamma - 1/\nu)}{\Gamma(d/2+1)\nu(\gamma - 1/\nu)^{\gamma-1/\nu}} \quad \text{if } \gamma\nu \geq 1. \tag{5.34}$$

We call a DPP model with a spectral density of the form (5.33) a generalized gamma model. For $\gamma\nu > 1$, the spectral density (5.33) attains its maximum at a non-zero value, which makes it fundamentally different from the other models considered in this paper where the maximum is attained at zero.

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

$$\varphi(x) = \rho\frac{\Gamma(d/2+1)\nu\alpha^d}{d\pi^{d/2}\Gamma(d/\nu)}\exp(-\|\alpha x\|^\nu). \tag{5.35}$$

We call a DPP model with a spectral density of the form (5.35) a power exponential spectral model. For $\nu = 2$, this is the Gaussian model of Section 5.2.

For the power exponential spectral model, $\alpha_{\max}$ is given in terms of ρ and ν by $\alpha_{\max}^d = \Gamma(d/\nu + 1)r^{-d}$, where r is defined in (5.27). For the choice $\alpha = \alpha_{\max}$ in (5.35), the spectral density of the power exponential spectral model becomes

$$\varphi(x) = \exp(-\|\Gamma(d/\nu + 1)^{1/d}x/r\|^\nu). \tag{5.36}$$

This function tends to the indicator function (5.27) as ν tends to ∞. Thus the power exponential spectral model contains a 'most repulsive possible stationary DPP' as a limiting case.

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) shows the spectral densities for these models, and it clearly illustrates how the spectral density approaches an indicator function as $\nu \to \infty$.

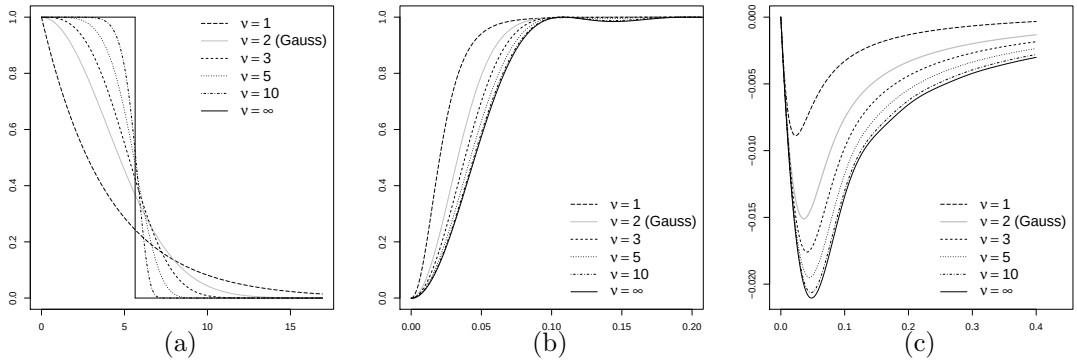


Figure 5: (a) Isotropic spectral densities, (b) approximate isotropic pair correlation functions, and (c) approximate $L(r) - r$ functions for power exponential spectral models with $\rho = 100$, $\nu = 1, 2, 3, 5, 10, \infty$ and $\alpha = \alpha_{\max}$ the maximal permissible value determined by (5.34).

Figure 5(b) shows the pair correlation functions g. Since we are not aware of a close form expression for $C_0 = \mathcal{F}^{-1}(\varphi)$ when φ is given by (5.36), we approximate C_0 by the periodic method, leading to approximating pair correlation functions in Figure 5(b). The figure shows that the repulsiveness of the process increases as ν increases. Notice the slightly oscillating nature of g for large values of ν, which at first may appear to be an artifact of models (5.28) and (5.29), but in fact such behaviour is expected for very repulsive processes (consider e.g. a stationary point process which is so repulsive that all of its realisations are uniform translations of a rectangular lattice, then g is periodic).

Figure 5(c) shows the corresponding approximations of $L(r) - r$ (analogously to Figure 3 in Section 5.2). The figure confirms once again that the repulsiveness of the process increases as ν increases.

6 Inference

In this section, we discuss how to estimate parameters of stationary DPP models and to a certain extent how to do model comparison and model checking. Section 6.1 focuses on maximum likelihood based inference, while Section 6.2 discusses alternative ways of performing inference. In Section 6.3, the approaches of Sections 6.1 and 6.2 are compared in a simulation study. Finally, in Section 6.4, a parametric determinantal point process model is fitted to a real data set.

We assume that X is a stationary DPP and that the kernel C is given by one of the parametric covariance models described in Sections 5.2 and 5.4. Further, we let $\boldsymbol{x} = \{x_1, \ldots, x_n\}$ denote a realization of X_S, where S is a bounded rectangular region and we refer to $\boldsymbol{x}$ as the data.

The covariance function is assumed to be parametrized by the intensity ρ and an additional parameter θ for the corresponding correlation function. Irrespective of the estimation procedure used, we always estimate ρ using the unbiased non-parametric estimate $\hat{\rho} = n/|S|$. This is a computationally simple estimate and it reduces the parameter dimension for the subsequent estimation procedure. The estimate $\hat{\rho}$

introduces a bound on the parameter space, since the remaining parameters have to satisfy the restriction $\rho_{\max}(\theta) \geq \hat{\rho}$, cf. Section 5.2.

6.1 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 5.3, where we use truncation and FFT to evaluate f^{per}. This is described in more detail in the following.

Let $\mathbb{Z}_N = \{-N, -N+1, \ldots, N-1, N\}$ and define truncated versions of (5.24) and (5.25) by

$$D_N = \sum_{k \in \mathbb{Z}_N^d} \log(1 + \tilde{\varphi}(k)) \tag{6.1}$$

and

$$\tilde{C}_N(u) = \sum_{k \in \mathbb{Z}_N^d} \tilde{\varphi}(k) \mathrm{e}^{2\pi \mathrm{i} k \cdot u}, \quad u \in \mathbb{R}^d. \tag{6.2}$$

Note that D_N and $\tilde{C}_N$ depend only on θ through $\tilde{\varphi}$ given by (5.23). For a given N (the choice of N is discussed below), the approximate maximum likelihood estimate (MLE) is the value of θ which maximizes the approximate log-likelihood

$$\ell_N(\theta) = \log \det[\tilde{C}_N](x_1, \ldots, x_n) - D_N$$

where $[\tilde{C}_N](x_1, \ldots, x_n)$ is the $n \times n$ matrix with (i,j)'th element $\tilde{C}_N(x_i - x_j)$. 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 (6.1) for large values of N, the evaluation of (6.2) is more problematic since it needs to be carried out for every pair of points in $\boldsymbol{x}$. For moderate N (few hundreds) direct calculation of (6.2) can be used, but 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.

6.2 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 (2010), and the references therein. Here we only briefly recall how the minimum contrast estimate (MCE) (Diggle and Gratton, 1984) is calculated.

Given a value of θ, let $s(r;\theta)$, $r \geq 0$, denote a functional summary statistic for which we have a closed form expression. 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 $\boldsymbol{x}$. The MCE based on the functional summary statistic s is the value of θ which minimizes

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

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 6.1, 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.

6.3 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 $\alpha_{\max}$ is given by (5.16). 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 6 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.

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)

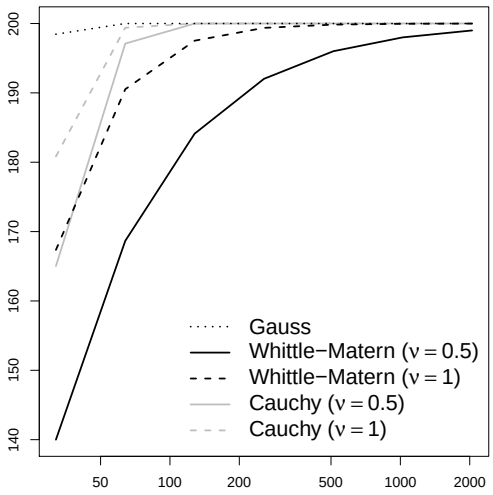


Figure 6: S_N as a function of N.

6.4 Real data example

Figure 7 shows a plot of the Norwegian spruces dataset available in `spatstat`. In the following we fit a DPP model to this dataset. The aim is not to conduct a detailed analysis of this specific dataset, but rather to illustrate that it is practically feasible to fit a DPP model to a real dataset. This dataset has previously been modelled by a five parameter multiscale process in Møller and Waagepetersen (2004), using elaborate Markov chain Monte Carlo MLE methods. Below we fit a more parsimonious DPP model by using the much simpler and faster methods from Section 6.

The non-parametric estimate of the intensity is $\hat{\rho} = 0.063$. First, we fit a Gaussian model to the data. The estimate of ρ implies $\alpha_{\max} = 2.248$. Both the MLE and the MCE of α is $\hat{\alpha} = \alpha_{\max}$ (the MCE based on the K-function and the MCE based on g are identical in this case). Second, we fit a Whittle-Matérn model and a Cauchy model to the data. For these models all the estimation procedures yield

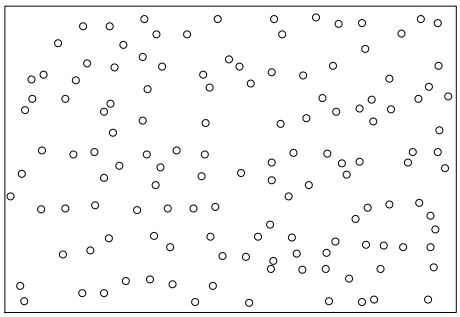


Figure 7: Locations of 134 pine trees in a 56 m by 38 m region.

estimates $\hat{\theta} = (\hat{\alpha}, \hat{\nu})$ with large values of $\hat{\nu}$ (not reported here), and the models are practically indistinguishable from the fitted Gaussian model, which is the limiting model as $\nu \to \infty$, cf. Section 5.2. The fitted Gaussian model has both the smallest value of $D(\hat{\theta})$ and the largest value of $\ell_N(\hat{\theta})$, which leads us to prefer the Gaussian model. Figure 8 is used to assess the goodness of fit for the Gaussian 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% 97.5% pointwise quantiles (gray lines) for these summary statistics based on 4000 simulations of the fitted Gaussian model (for definitions of F and G, see e.g. Møller and Waagepetersen (2004)). Figure 8 indicates a lack of fit of the Gaussian model applied to the Norwegian spruces, and a more repulsive model may be appropriate.

We therefore fit the power exponential spectral model of Section 5.4, where we have truncated the parameter space for ν to $0 < \nu \leq 10$, since models with larger values of ν are almost indistinguishable from the model with $\nu = 10$. The approximate MLE is $\hat{\theta} = (\hat{\alpha}, \hat{\nu}) = (6.36, 10)$. As $\hat{\alpha}$ is close to the maximal permissible value $\alpha_{\max} = 6.77$ for $\nu = 10$, the fitted model is close to a 'most repulsive possible stationary DPP', that is, the jinc-like function (5.29). This model is judged to be a good fit based on Figure 8 where the simulation based quantiles (black lines) cover the non-parametric estimates based on the dataset for all the summary statistics.

Acknowledgments

We are grateful to Philippe Carmona, Morten Nielsen and Rasmus Waagepetersen for helpful comments. Supported by the Danish Natural Science Research Council, grant 09-072331, "Point process modelling and statistical inference", and by the Centre for Stochastic Geometry and Advanced Bioimaging, funded by a grant from the Villum Foundation.

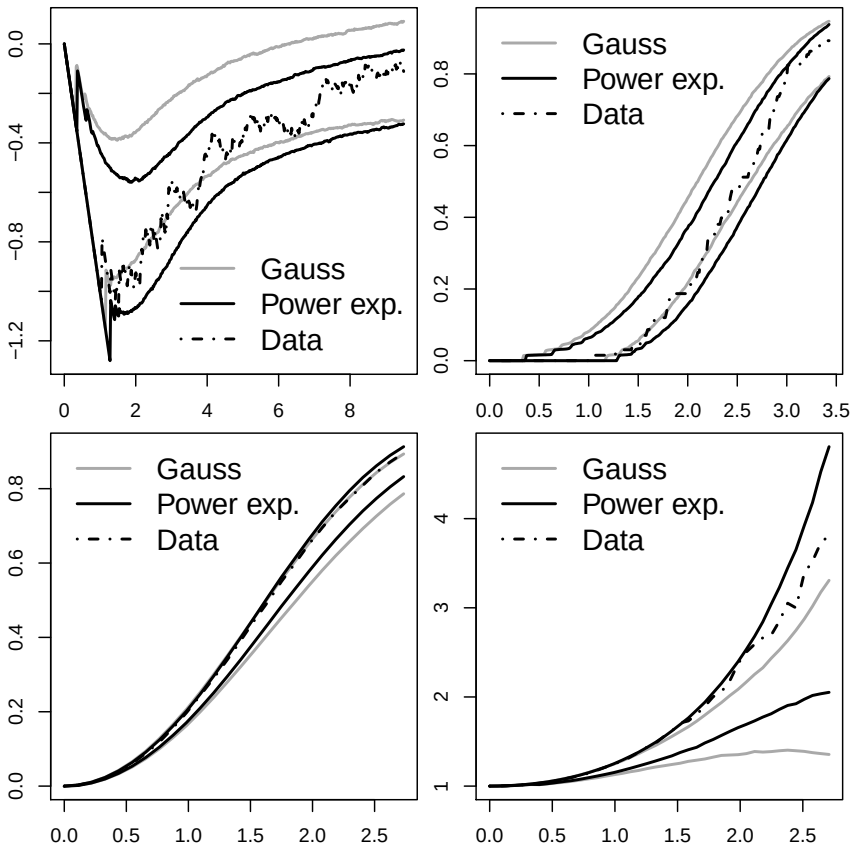


Figure 8: Clockwise from top left: Non-parametric estimate of $L(r) - r$, $G(r)$, $F(r)$, and $J(r)$ for the Norwegian spruces dataset. All four plots contain simulation based 2.5% and 97.5% pointwise quantiles for both the Gaussian model and the power exponentialspectral model fitted via MLE. For both models the quantiles are based on 4000 realizations.

Appendices

A Proof of Theorem 3.2

Let the situation be as in Theorem 3.2. In the sequel, for ease of presentation, we ignore null sets.

We start by proving by induction that for $i = n, \dots, 1$, (3.5) is a probability density and $\boldsymbol{v}(X_n), \dots, \boldsymbol{v}(X_i)$ are linearly independent (considering complex scalars).

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

$$\int_S p_n(x)\,\mathrm{d}x = \frac{1}{n}\int_S \|\boldsymbol{v}(x)\|^2\,\mathrm{d}x = \frac{1}{n}\int_S \sum_{k=1}^{n} |\phi_k(x)|^2\,\mathrm{d}x = 1.$$

Hence p_n is a probability density. Clearly, $p_n(x) = 0$ whenever $\boldsymbol{v}(x) = \boldsymbol{0}$, so as X_n is generated from p_n, $\boldsymbol{v}(X_n) \neq \boldsymbol{0}$ (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 (3.6) has dimension $n - i$. Let P_i be the matrix of the orthogonal projection from $\mathbb{C}^n$ onto $H_i^\perp$. By (3.7), for all $x \in S$, $p_i(x) = \|P_i\boldsymbol{v}(x)\|^2/i \geq 0$ and

$$p_i(x) = 0 \quad \text{whenever } \boldsymbol{v}(x) \in H_i. \tag{A.1}$$

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

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

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

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

Thus p_i is a probability density. Finally, it follows immediately from (A.1) and the induction hypothesis that $\boldsymbol{v}(X_n), \dots, \boldsymbol{v}(X_{i+1}), \boldsymbol{v}(X_i)$ 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

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

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

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

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

$$p(x_1, \ldots, x_n) = \frac{1}{n!} \det[K](x_1, \ldots, x_n). \tag{A.2}$$

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

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

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

This completes the proof of Theorem 3.2.

B Close upper bounds on the conditional distributions of Algorithm 1

In Remark 3.5 we discussed rejection sampling from the densities p_i, $i = n, \ldots, 1$, 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 3.4 that $ip_i(x)$ is the norm of a vector obtained after $n - i$ successive orthogonal projections of $\boldsymbol{v}(x)$. These projections commute, so that $ip_i(x)$ is lower than the norm of any projection of $\boldsymbol{v}(x)$ of a lower order. By (3.7) and (3.8), if $i + 1 \leq k \leq n$, then

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

and so by (3.4),

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

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 (B.1) to derive an explicit upper bound in the specific setting of Section 5, i.e. in the stationary case, when $S = [-1/2, 1/2]^d$,

and when the eigenfunctions are Fourier basis functions as in (5.17). Let $x = (x(1), \ldots, x(d)) \in \mathbb{R}^d$ and $y = (y(1), \ldots, y(d)) \in \mathbb{R}^d$, and suppose that

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

where $\varphi_{j_1,\ldots,j_d}(x) = \exp\left(2\pi \mathrm{i} \sum_{k=1}^d j_k x(k)\right)$ and for $q = 1, \ldots, d$, $J_q(n_q)$ denotes some finite subset of $\mathbb{Z}$ with n_q elements, such that $n = \prod_{q=1}^d n_q$. Then the projection kernel (3.4) becomes

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

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

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

Proof. For $x, y \in \mathbb{R}$, let $K_q(x, y) = \sum_{j_q \in J_q(n_q)} \mathrm{e}^{2\pi \mathrm{i} j_q (x-y)}$. 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

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

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

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

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

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

and so

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

Combining this with (B.1), we obtain (B.3). □

The upper bound in (B.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 (B.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.

C Fourier approximation of the Whittle-Matérn covariance function

Consider the Whittle-Matérn covariance function C_0 given by (5.9). This appendix begins with some preliminary results on K_ν (the Bessel function of the second kind) which appears in (5.9). Next the quality of the Fourier approximation of C_0 given in Section 5.3 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 $\nu > 0$,

$$K_\nu(x) = \frac{\sqrt{\pi}}{\Gamma(\nu + \frac{1}{2})} \left(\frac{x}{2}\right)^\nu \int_1^\infty \mathrm{e}^{-xt}(t^2 - 1)^{\nu - \frac{1}{2}} \,\mathrm{d}t. \tag{C.1}$$

As $x \to 0$, then $x^\nu K_\nu(x) \to 2^{\nu-1}\Gamma(\nu)$. Hence by (5.9), $C_0(0) = \rho$.

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 C.1. *For all $x > 0$,*

$$K_\nu(x) \le 2^{\nu-1}\Gamma(\nu)x^{-\nu}\left(1 - (1 - \mathrm{e}^{-\gamma x})^{2\nu}\right) \quad \textit{if } \nu \ge 1/2 \tag{C.2}$$

and

$$K_\nu(x) \le K_{1/2}(x) = \sqrt{\pi/(2x)}\,\mathrm{e}^{-x} \quad \textit{if } \nu \le 1/2. \tag{C.3}$$

Proof. When $\nu \ge 1/2$, from (C.1),

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

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

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

From Alzer (1997) we deduce that

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

whenever $x > 0$, $\nu \geq 1/2$, and $0 \leq \gamma \leq \Gamma(1+2\nu)^{-1/2\nu}$. Hence (C.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 (C.1) and the fact that $t^2 - 1 > 2t - 2$ when $t > 1$, we obtain

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

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

For the Whittle-Matérn model, the following Proposition C.2 provides an error bound for the approximation (5.20) of $C_0(u)$ by $C_{\mathrm{app},0}(u)$ when $u \in [-1/2, 1/2]^d$. We let

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

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

$$\epsilon(\nu,\alpha,1) = \frac{\mathrm{e}^{-\beta}}{\beta} + \frac{2\mathrm{e}^{-\beta}}{1-\mathrm{e}^{-\beta}}\left(\frac{\mathrm{e}^{-\beta}}{\beta} + \frac{1}{1-\mathrm{e}^{-\beta}} - 1\right) \tag{C.5}$$

and for $d \geq 2$,

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

Proposition C.2. *Let C_0 be the Whittle-Matérn covariance function given by* (5.9) *and let $C_{\mathrm{app},0}$ be the approximation* (5.20) *of C_0 on $[-1/2,1/2]^d$. If $0 \leq \rho \leq \rho_{\max}$ where $\rho_{\max}$ is given by* (5.10)*, then*

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

Proof. We have

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

with

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

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

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

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 $1-(1-\mathrm{e}^{-\gamma x})^{2\nu} \leq 2\nu \mathrm{e}^{-\gamma x}$ for all $x>0$, so from (5.9) and (C.2),

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

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 (C.8) can be computed easily to get

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

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

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

and (C.7), which involves the terms (C.4) (for $\nu\geq 1/2$) and (C.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 (C.8) is more difficult to compute and we therefore establish an upper bound as follows. Since $\|y\| \geq (|y_1|+\cdots+|y_d|)/\sqrt{d}$,

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

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

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

and

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

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

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

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

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

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

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

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

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

□

Note that the inequality (C.7) reduces to an equality in the particular case $d = 1$ and $\nu = 1/2$. Finally, the plots in Figure 9 confirm that for reasonable values of ρ, ν, and α satisfying (5.10), the error bound (C.7) is small.

D Alternative approximation of the density

Let $S \subset \mathbb{R}^d$ be a compact set. In this appendix, in addition to Assumption 5.2, 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 4.1 is well defined. Recall that f is given in terms of $\tilde{C}$ and D, cf. (4.3). Below we introduce computationally convenient approximations of $\tilde{C}$ and D which can be used with (4.3) to obtain an approximation of f.

D.1 Convolution approximation of f

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

$$\tilde{C}_{\mathrm{app},0}(u) = \sum_{k=1}^{\infty} C_0^{\star k}(u), \quad u \in \mathbb{R}^d, \tag{D.1}$$

is well-defined, where

$$C_0^{\star 1}(u) = C_0(u), \quad C_0^{\star k}(u) = \int C_0^{\star(k-1)}(x) C_0(u-x) \, \mathrm{d}x, \quad u \in \mathbb{R}^d, \ k = 2, 3, \ldots. \tag{D.2}$$

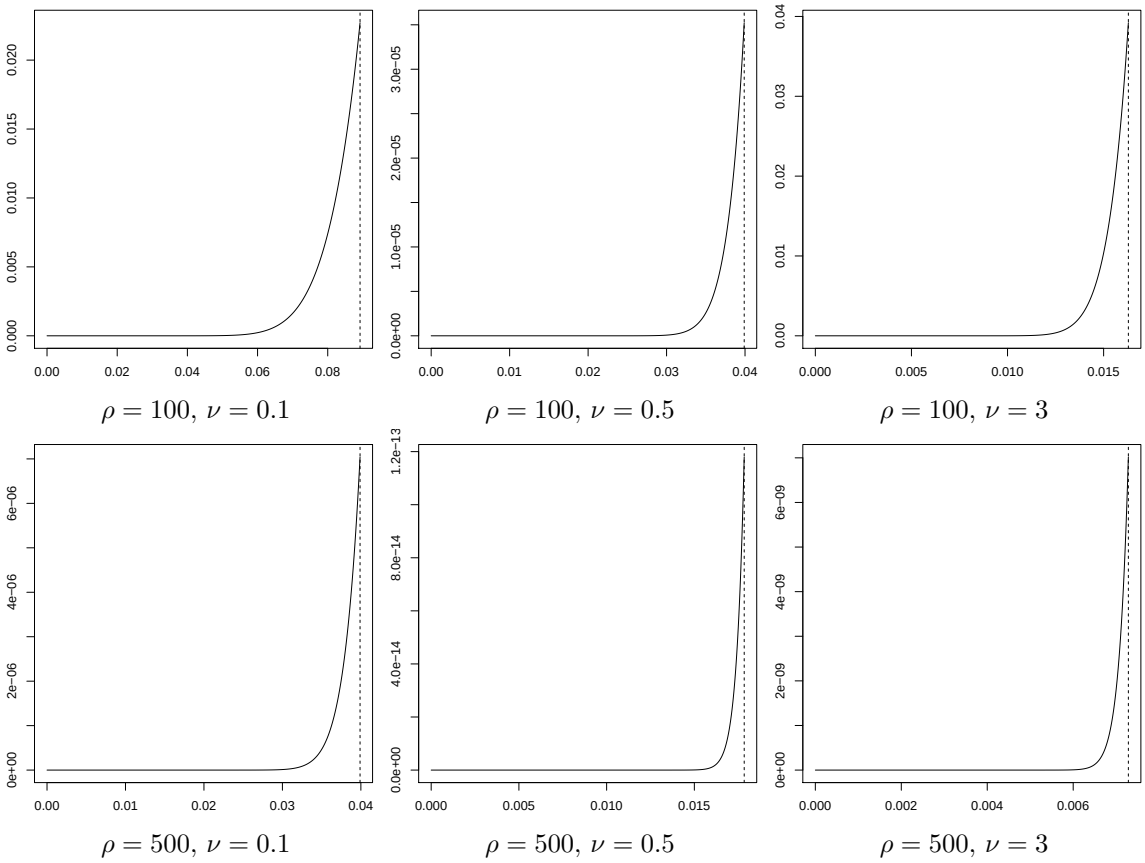


Figure 9: Error-bound (C.7) in terms of α for different values of ρ and ν, when $d = 2$. The dotted line represents the maximal possible value of α following from (5.10).

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 $u \in \mathbb{R}^d$, $\varphi(u) = \lim_{n\to\infty} \tilde{\varphi}_n(u)$, where $\tilde{\varphi}_n(u) = \sum_{k=1}^n \varphi(u)^k$. We see that $\tilde{\varphi} \in L^1(\mathbb{R}^d)$ since

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

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 (D.1).

For $x, y \in S$, we define $\tilde{C}_{\mathrm{app}}(x,y) = \tilde{C}_{\mathrm{app},0}(x-y)$ and use the approximation $\tilde{C}(x,y) \approx \tilde{C}_{\mathrm{app}}(x,y)$. The expansion (D.1) corresponds to (4.6) with $C_S^k(x,y)$ substituted by $C_0^{\star k}(x-y)$.

Using the same substitution in (4.5) leads us to approximate D by

$$D_{\mathrm{app}} = |S| \sum_{k=1}^{\infty} C_0^{\star k}(0)/k. \tag{D.3}$$

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

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

Then the convolution approximation of f is defined by

$$f_{\mathrm{app}}(\{x_1, \ldots, x_n\}) = \exp(|S| - D_{\mathrm{app}}) \det[\tilde{C}_{\mathrm{app}}](x_1, \ldots, x_n). \tag{D.4}$$

As mentioned above, the approximations $\tilde{C}_{\mathrm{app}}$ 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. (4.4) and (D.2). However, the final approximations $\tilde{C} \approx \tilde{C}_{\mathrm{app}}$ and $D \approx D_{\mathrm{app}}$ 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 $k \to \infty$. Since $C_0^{\star k}$ is a covariance function, we have $C_0^{\star k}(u) \leq C_0^{\star k}(0)$ for all $k = 1, 2, \ldots$. 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.

D.2 Examples

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

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

where N is a positive integer. Furthermore, we need closed form expressions for $C_0^{\star k}(u)$. For the normal variance mixture models presented in Section 5.2, we have $C_0^{\star k}(u) = (\rho/\rho_{\max})^k h^{\star k}(u)$, and so it suffices to find closed form expressions for $h^{\star k}$. For the Gaussian model,

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

while for the Whittle-Matérn model,

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

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, $h^{\star k}(0)$ decays as $k^{-d/2}$ when $k \to \infty$, and therefore the rate of convergence of $\tilde{C}_{\text{app},0}(0) = \sum_{k=1}^{\infty} (\rho/\rho_{\max})^k h^{\star k}(0)$ 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 $\rho_{\max}$.

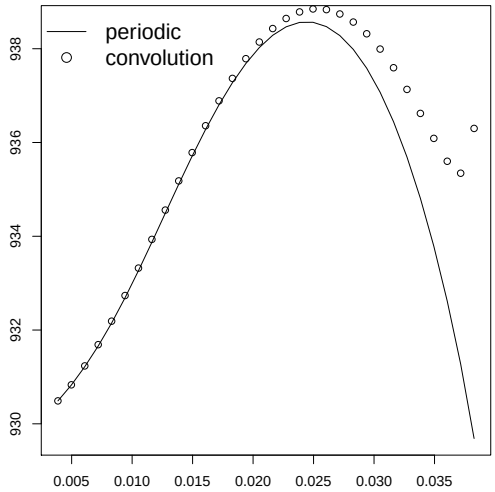


Figure 10: Comparison of using the convolution and periodic density approximations to approximate the log-likelihood of the Gaussian model as a function of α based on a simulated dataset in the unit square with $\rho = 200$ and $\alpha = 0.02$.

Based on a simulated point pattern in the unit square, Figure 10 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 6.3, 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 10. 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 $\rho/\rho_{\max}$ is small.

References

Alzer, H. (1997). On some inequalities for the incomplete Gamma function. *Mathematics of Computation 66*, 771–778.

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. E. (1977). 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*. Dortrecht: D. Reidel Publishing Company.

De laco, 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.

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.

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.

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.

Kuna, T., J. Lebowitz, and E. Speer (2007). Realizability of point processes. *Journal of Statistical Physics 129*, 417–439.

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 permanental process. *Advances in Applied Probability 38*, 873–888.

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.

Prokešová, M. and E. B. V. Jensen (2010). Asymptotic Palm likelihood theory for stationary point processes. Technical Report No. 11, Center for Stochastic Geometry and Advanced Bioimaging. Conditionally accepted by Annals of the Institute of Statistical Mathematics.

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.

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.

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.

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.