ORIGINAL ARTICLE

Cox processes driven by transformed Gaussian processes on linear networks

Jesper Møller[1] | **Jakob G. Rasmussen**[1]

[1]Department of Mathematical Sciences, Aalborg University, Skjernvej 4A, DK-9220 Aalborg Ø, Denmark

Correspondence
Jesper Møller, Department of Mathematical Sciences, Aalborg University, Skjernvej 4A, DK-9220 Aalborg Ø, Denmark
Email: jm@math.aau.dk

In this paper, (some more text will be added here)

Keywords – geodesic distance, interrupted point process, isotropic covariance function, log Gaussian Cox process, moment properties and statistical procedures, permanental point process, resistance metric, simulation.

1 | INTRODUCTION

In Sir David R. Cox's highly influential paper 'Some Statistical Methods Connected with Series of Events' (Cox, 1955) he invented doubly stochastic Poisson processes obtained by a generalization of Poisson processes where the intensity function Λ that varies over space or time is a stochastic process. These Cox models play nowadays an important role when analysing point patterns in Euclidean spaces or on spheres (see Møller and Waagepetersen, 2004; Lawrence et al., 2016, and the references therein). In particular, Cox processes driven by a transformed Gaussian process (GP) Y or independent copies $Y_1, \ldots, Y_h$ of Y play a major role: A Log Gaussian Cox process (LGCP) has $\Lambda(u) = \exp(Y(u))$ (Møller et al., 1998; Cuevas-Pacheco and Møller, 2018); LGCPs constitute the most widely used subclass of Cox processes. An interrupted Cox process (ICP) is obtained by an independent thinning of a Poisson process, where the retention probability of a point u is given by $\exp(-\sum_{i=1}^{h} Y_i(u)^2)$ (Stoyan, 1979; Lavancier and Møller, 2016). Moreover, a permanental point process (PPP) is obtained when $\Lambda(u) = \sum_{i=1}^{h} Y_i(u)^2$ (Macchi, 1975; McCullagh and Møller, 2006).

In recent years there has been an increasing interest in analysing point patterns on a linear network L, that is, L is a connect set given by a finite union of bounded, closed, line segments in $\mathbb{R}^k$ (the real coordinate space of dimension k) which can only overlap at their endpoints, see Ang et al. (2012), Baddeley et al. (2015), the references therein as well as further references given later in the present paper. The purpose of the present paper is to use the isotropic covariance function models developed in Anderes et al. (2020) as well as new models developed in this paper, thereby constructing models for GPs and hence LGCPs, ICPs, and PPPs on linear networks. Also we construct new simulation algorithms, consider statistical procedures and applications, and discuss whether we should use the geodesic or resistance metric.

The paper consists of two parts, Sections 2-3 on our setting for isotropic covariance functions and related GPs, and Sections 4-7 on point processes, in particular Cox processes, including the cases of LGCPs, ICPs, and PPPs on

linear networks and how these models can be used for fitting real data. More specifically: Section 2.1 considers a general metric space (S, d) in order later to compare covariance functions and point processes for the case $S = L$ with the cases where $S = \mathbb{R}^k$ or $S = \mathbb{S}^k$ (the k-dimensional unit sphere). When $S = L$ we have in mind that d is either the geodesic metric d_G or the resistance metric d_R. Section 2.2 is a summary of results for d_R, including a useful expression for the metric. Furthermore, when $S = L$, Section 3.1 studies isotropic covariance functions of the form $c(u, v) = c_0(d(u, v))$ and provides a less technical summary of results from Anderes et al. (2020) and examples of isotropic covariances function not appearing in that paper. Simulation algorithms for GPs on L with an isotropic covariance function are developed in Section 3.2, where the case with L a tree is particular tractable. Our setting for point processes is given in Section 4.1, and Section 4.2 introduces first and higher order intensity functions which become useful when we later study Cox process models. In particular, we focus on the pair correlation function and the related K-function defined in Section 4.2, where we stress the importance of considering an isotropic pair correlation function $g(u, v) = g_0(d(u, v))$, partly because inference procedures have mainly been developed for this case and partly since for LGCPs, ICPs, and PPPs isotropy of g becomes equivalent to isotropy of the covariance function for the underlying GPs. As discussed in Section 4.3, g, K, and other functional characteristics for point processes become useful for statistical inference. Section 5.1 treats Cox processes in our general setting, since moment and many other properties of Cox processes driven by transformed Gaussian processes do not depend on which space we consider – the dependence of the metric space (S, d) occurs when we want to specify well-defined parametric models of covariance functions, simulate the underlying GPs, and use statistical inference procedures, which for the case $S = L$ are the topics of Sections 3 and 6. Moreover, Section 5.2 studies the properties of LGCPs, ICPs, and PPPs models, and in Section 6 we demonstrate how these models may be fitted to real data. Finally, Section 7 summaries our findings and discuss some open problems.

(Some more text will be added here)

2 | METRIC SPACES

2.1 | Setting

The state space for the points of the point processes considered in this paper is a metric space (S, d) equipped with a reference measure ν. We focus on the case where S is a linear network and compare with the cases where $S = \mathbb{R}^k$ is the k-dimensional Euclidean space or $S = \mathbb{S}^k = \{u \in \mathbb{R}^{k+1} \mid \|u\| = 1\}$ is the k-dimensional unit sphere, with d and ν specified as follows.

(a) The case $S = \mathbb{R}^k$: $d(u, v) = \|u - v\|$ is given by the usual Euclidean distance and $\nu(A) = \int_A \mathrm{d}u$ is Lebesgue measure.

(b) The case $S = \mathbb{S}^k$: $d(u, v) = d_G(u, v) = \arccos(u \cdot v)$ is the geodesic (orthodromic, great-circle, or shortest path) metric and $\nu = \nu_k$ is k-dimensional Hausdorff/surface measure (see e.g. Dai and Xu, 2013, Chapter 1). Here, $u \cdot v$ is the usual inner product of u and v.

(c) The case $S = L = \cup_{i=1}^m L_i$ of a linear network: We assume $m < \infty$, each $L_i \subset \mathbb{R}^k$ is a closed line segment of length $l_i \in (0, \infty)$, $L_i \cap L_j$ is either empty or an endpoint of both L_i and L_j whenever $i \neq j$, and L is a path-connected set. Furthermore, d is a 'natural' metric and $\nu(A) = \int_A \mathrm{d}_L(u)$ is arc length measure. We let $|L| = \sum_{i=1}^m l_i = \int_L \mathrm{d}_L(u)$ denote the length of the linear network.

The remainder of this section are remarks to the case (c).

For infinite and/or disconnected linear networks, definitions and results may be applied separately to each connected component of the network if we consider point and Gaussian processes which are mutually independent on the connected components.

Our definition in (c) may be extended to the more abstract case of a graph with Euclidean edges, i.e., graphs with edges viewed as line or curve segments allowing us to consider points on the graph which are either vertices or points on an edge (Anderes et al., 2020). In the present paper we avoid this generalization partly for ease of presentation and partly since statistical methods have so far only been developed for the case (c).

For each line segment L_i, there are two possible arc length parametrisations. We assume one is chosen and given by $u_i(t) = (1 - t/l_i)\, a_i + (t/l_i)\, b_i$, $t \in [0, l_i]$, where a_i and b_i are the endpoints of L_i. The definitions and results in this paper will not depend on this choice, including when calculating arc length measure restricted to L_i: For Borel sets $A \subseteq L_i$, $\int_A \mathrm{d}_L(u) = \int_0^{l_i} 1(u_i(t) \in A)\, \mathrm{d}t$, where $1(\cdot)$ denotes the indicator function.

Let V denote the set of endpoints of $L_1, \ldots, L_m$ and consider the graph with vertex set V and edge set E given by $L_1, \ldots, L_m$, i.e., two distinct vertices $u, v \in V$ form an edge if and only if $\{u, v\} = \{a_j, b_j\}$ for some $j \in \{1, \ldots, m\}$, in which case we write $u \sim v$. We have mainly two cases of natural metrics in mind, namely when d is the geodesic metric $d_{\mathcal{G}}$ or the resistance metric $d_{\mathcal{R}}$ as developed in Anderes et al. (2020) when extending the original definition in Klein and M. Randić (1993) for V to L: Let $u, v \in L$. Then $d_{\mathcal{G}}(u, v) = \min \int_{p_{uv}} \mathrm{d}_L(w)$ where the minimum is over all paths $p_{uv} \subseteq L$ connecting u and v. When defining $d_{\mathcal{R}}(u, v)$, without loss of generality assume that $u, v \in V$, since if e.g. $u \in L_j \setminus V$, we may split L_j into the two line segments with endpoints $\{a_j, v\}$ and $\{v, b_j\}$, and then consider a new graph with vertex set $V \cup \{u\}$ and edge set $E \cup \{\{a_j, v\}, \{v, b_j\}\}$ (this follows from (B) in Theorem 1 below). Viewing the graph (V, E) as an electrical network with resistor l_i at edge L_i, $i = 1, \ldots, m$, then $d_{\mathcal{R}}(u, v)$ is the effective resistance between u and v as obtained by Kirkhoff's laws. In Section 2.2 below we provide the detailed definition of $d_{\mathcal{R}}$. Indeed there are other interesting metric when $S = L$, including the least-cost metric (Rakshit et al., 2017), but to the best of our knowledge parametric models for covariance functions (which is one topic of our main interest) have so far only been developed when $d = d_{\mathcal{G}}$, $d = d_{\mathcal{R}}$, or d is given by the usual Euclidean distance; the latter case is usually not a natural metric on a linear network.

2.2 | The resistance metric

This section defines the resistance metric $d_{\mathcal{R}}$ for a graph with Euclidean edges (Anderes et al., 2020) in the special case of a linear network $L = \cup_{i=1}^m L_i$ as given in case (c) in Section 2.1. The section also summarises some properties of $d_{\mathcal{R}}$ and compares with $d_{\mathcal{G}}$.

Consider the graph $G = (V, E)$ and its relation $\sim$ as defined above. Since the resistance metric on L is an extension of the classic (effective) resistance metric d_V defined on V, we start by recalling what d_V is: Let $u_0 \in V$ be an arbitrarily chosen vertex called the origin. For any $u, v \in V$, define the so-called conductance function by

$$\mathrm{con}(u, v) = \begin{cases} 1/\|u - v\| & \text{if } u \sim v \\ 0 & \text{otherwise} \end{cases}$$

and define a matrix Δ with rows and columns indexed by V so that its entry (u, v) is given by

$$\Delta(u, v) = \begin{cases} 1 + c(u) & \text{if } u = v = u_0 \\ c(u) & \text{if } u = v \neq u_0 \\ -\mathrm{con}(u, v) & \text{otherwise} \end{cases}$$

where $c(u) = \sum_{w \in V: w \sim u} \mathrm{con}(u, w)$ is the sum of the conductances associated to the edges incident to vertex u. In fact Δ is symmetric and strictly positive definite, and it is similar to the 'Laplacian matrix' from electrical network theory (when viewing G as an electrical network over the nodes with resistors given by the length of each line segment, see e.g. Kigami, 2003; Jorgensen and Pearse, 2010) except that Δ has the additional 1 added at entry (u_0, u_0) which makes Δ invertible. Let B_0 be a Gaussian vector indexed by V and having mean zero and covariance matrix $\Sigma = \Delta^{-1}$. Then the resistance metric on V is the variogram

$$d_V(u, v) = \mathbb{V}\mathrm{ar}(B_0(u) - B_0(v)) = \Sigma(u, u) + \Sigma(v, v) - 2\Sigma(u, v) \quad \text{for } u, v \in V. \tag{1}$$

It can be shown that d_V does not depend on the choice of origin u_0, cf. Theorem 1 below. If V is large, we may expect Δ to be sparse and hence there are quick methods for obtaining $\Sigma = \Delta^{-1}$ (see e.g. Rue and Held, 2005).

Now, extend B_0 by linear interpolation to a mean zero Gaussian process (GP) Z_0 on L so that

$$Z_0(u) = \frac{\|u - b_i\|}{l_i} B_0(a_i) + \frac{\|u - a_i\|}{l_i} B_0(b_i) \quad \text{for } u \in L_i.$$

For $i = 1, \ldots, m$, define a mean zero Brownian bridge B_i on L_i so that

$$\mathbb{C}\mathrm{ov}(B_i(u), B_i(v)) = \min\{\|u - a_i\| \|v - b_i\|, \|v - a_i\| \|u - b_i\|\}/l_i \quad \text{for } u, v \in L_i,$$

and define

$$Z_i(u) = \begin{cases} B_i(u) & \text{for } u \in L_i \\ 0 & \text{for } u \in L \setminus L_i. \end{cases}$$

Finally, the resistance metric on L is defined by

$$d_{\mathcal{R}}(u, v) = \sum_{i=0}^{m} \mathbb{V}\mathrm{ar}(Z_i(u) - Z_i(v)) \quad \text{for } u, v \in L. \tag{2}$$

Note that $d_{\mathcal{R}}(u, v) = d_V(u, v)$ if $u, v \in V$.

For the following theorem, which follows from Anderes et al. (2020, Propositions 2-4), we use a terminology as follows. A closed line segment in $\mathbb{R}^k$ with endpoints a and b is denoted $[a, b] = \{at + b(1 - t) \mid 0 \le t \le 1\}$. A path is a subset of L of the form $[u, v_1] \cup [v_1, v_2] \cdots \cup [v_{i-1}, v_i] \cup [v_i, v]$ where $u, v \in L$, $v_1, \ldots, v_i \in V$ are vertices, $i \ge 0$ is an integer, and we interpret $[v_1, v_2] \cdots \cup [v_{i-1}, v_i]$ as the empty set if $i = 0$. If all vertices in G are of order two, we say that L is a loop (since L is isomorphic to a circle). If there is no loop, we say that L is a tree.

Theorem 1 *We have the following for the resistance and geodesic metrics on a linear network L.*

(A) *The definition 2 of $d_{\mathcal{R}}$ does not depend on the choice of origin $u_0 \in V$.*

(B) *Both $d_{\mathcal{R}}$ and $d_{\mathcal{G}}$ are metrics on L, and their definitions are invariant to splitting a line segment L_i into two line segments.*

(C) *For every $u, v \in L$, $d_{\mathcal{G}}(u, v) \ge d_{\mathcal{R}}(u, v)$, with equality if and only if there is only one path connecting u and v. In particular, $d_{\mathcal{G}} = d_{\mathcal{R}}$ if and only if L is a tree.*

(D) *If G is a loop, then $d_{\mathcal{R}}(u, v) = d_{\mathcal{G}}(u, v) - d_{\mathcal{G}}(u, v)^2 / \sum_{i=1}^{m} l_i$.*

The following proposition shows how $d_{\mathcal{R}}(u,v)$ can be easily calculated.

Proposition 1 *For any $u \in L_j$ and $v \in L_i$, let*

$$s = \|u - a_j\|, \quad t = \|v - a_i\|, \quad A_i = d_V(a_i, b_i)/l_i^2 - 1/l_i.$$

Then $A_i \le 0$ with equality if and only if L_i is the only path connecting a_i and b_i, and $d_{\mathcal{R}}(u,v)$ satisfies the following.

(A) *If $i = j$ then*

$$d_{\mathcal{R}}(u,v) = \begin{cases} A_i(t-s)^2 + t - s & \text{if } t \ge s, \\ A_i(s-t)^2 + s - t & \text{if } t \le s, \end{cases} \tag{3}$$

so $d_{\mathcal{R}}(u,v)$ considered as a function of t is linear (the case $A_i = 0$) or quadratic (the case $A_i < 0$) on each of the intervals $[0,s]$ and $[s,l_i]$, continuous on $[0,l_i]$, and differentiable on $[0,l_i] \setminus \{s\}$.

(B) *If $i \ne j$ then*

$$d_{\mathcal{R}}(u,v) = A_i t^2 + B_{ij}(s)t + C_{ij}(s) \tag{4}$$

where

$$B_{ij}(s) = 1 - \frac{2}{l_i}\left[\Sigma(a_i,a_i) - \Sigma(a_i,b_i) - \frac{l_j - s}{l_j}\Sigma(a_j,a_i) + \frac{l_j - s}{l_j}\Sigma(a_j,b_i) - \frac{s}{l_j}\Sigma(b_j,a_i) + \frac{s}{l_j}\Sigma(b_j,b_i)\right]$$

and

$$C_{ij}(s) = \frac{(l_j - s)^2}{l_j^2}\Sigma(a_j,a_j) + \frac{s^2}{l_j^2}\Sigma(b_j,b_j) + 2\frac{s(l_j - s)}{l_j^2}\Sigma(a_j,b_j) + \Sigma(a_i,a_i) - 2\frac{l_j - s}{l_j}\Sigma(a_j,a_i) - 2\frac{s}{l_j}\Sigma(b_j,a_i) + \frac{s(l_j - s)}{l_j},$$

so $d_{\mathcal{R}}(u,v)$ is a linear or quadratic concave function of $t \in [0,l_i]$.

(C) *If $i \ne j$ then*

$$d_{\mathcal{R}}(u,v) \ge \min\{d_{\mathcal{R}}(u,a_i), d_{\mathcal{R}}(u,b_i)\}. \tag{5}$$

Proof Since $l_i = d_{\mathcal{G}}(a_i,b_i) = d_V(a_i,b_i)$, Theorem 1[C] gives that $A_i \le 0$ with equality if and only if L_i is the only path connecting a_i and b_i. From (1) and (2) we obtain (3) and (4) by a straightforward calculation, and thereby we easily see that $d_{\mathcal{R}}(u,v)$ as a function of t behaves as stated in (A) and (B). Finally, since $d_{\mathcal{R}}(u,v)$ is a concave function of $t \in [0,l_i]$ in the case $i \ne j$, the inequality (5) follows.

Some remarks are in order. It follows from (3) and (4) that once Σ has been calculated, $d_{\mathcal{R}}(u,v)$ can be quickly calculated. We have $d_{\mathcal{R}}(v,a_i) = A_i t^2 + t$, cf. (3) with $s = 0$, and $d_{\mathcal{R}}(u,a_i) = C_{ij}(s)$, cf. (4) with $t = 0$, so we can rewrite (4) as

$$d_{\mathcal{R}}(u,v) = d_{\mathcal{R}}(u,a_i) + d_{\mathcal{R}}(v,a_i) + \left[B_{ij}(s) - 1\right]t.$$

This together with the triangle inequality show that $B_{ij}(s) \le 1$. The inequality (5) becomes useful when searching for point pairs $u, v \in L$ with $d_{\mathcal{R}}(u,v) \le r$ and $v \in L_i$, since we need only to consider the cases where $d_{\mathcal{R}}(u,a_i) \le r$ or $d_{\mathcal{R}}(u,b_i) \le r$.

3 | GAUSSIAN PROCESSES AND ISOTROPIC COVARIANCE FUNCTIONS

Let (S,d) be a metric space as in Section 2.1 and recall that a function $c : S \times S \mapsto \mathbb{R}$ is positive definite (sometimes called positive semi-definite) if $\sum_{j,\ell=1}^{n} a_j a_\ell c(u_j,u_\ell) \ge 0$ for all $a_1,\ldots,a_n \in \mathbb{R}$, all pairwise distinct $u_1,\ldots,u_n \in S$, and $n = 1,2,\ldots$. Let $Y = \{Y(u) \mid u \in S\}$ be a GP, where each $Y(u)$ is a real-valued random variable. The distribution of Y is specified by that for $n = 1,2,\ldots$ and every $u_1,\ldots,u_n$, $(Y(u_1),\ldots,Y(u_n))$ follows an n-dimensional normal distribution given by the mean function $\mu(u) = \mathbb{E}Y(u)$ and the covariance function

$$c(u,v) = \mathbb{C}\mathrm{ov}(Y(u),Y(v)) = \mathbb{E}[Y(u)Y(v)] - \mu(u)\mu(v).$$

The necessary and sufficient condition for a well-defined GP is the property of the covariance function that it is symmetric and positive definite.

3.1 | Isotropic covariance functions

We are in particular interested in isotropic covariance functions c meaning that c is of the form $c(u,v) = c_0(d(u,v))$ for all $u,v \in S$. Here, with some abuse of terminology, we also call c_0 a covariance function and it is required to be positive definite, that is, $\sum_{j,\ell=1}^{n} a_j a_\ell c_0(d(u_j,u_\ell)) \ge 0$ for all $a_1,\ldots,a_n \in \mathbb{R}$, all pairwise distinct $u_1,\ldots,u_n \in S$, and $n = 1,2,\ldots$. In Sections 3.1.1 and 3.1.2 below we assume that the variance $\sigma^2 = c_0(0)$ is strictly positive and consider the correlation function $r_0(t) = c_0(t)/\sigma^2$ in connection to the cases (a)–(c) in Section 2.1. In all of our examples, r_0 will be a completely monotone function.

3.1.1 | The classical cases of Euclidean spaces and spheres

Let $S = \mathbb{R}^k$ and $d(u,v) = \|u-v\|$, cf. case (a) in Section 2.1. Then Schoenberg (1938) and Gneiting (2001) provide detailed studies of continuous isotropic covariance functions, and Table 1 shows examples of popular parametric models for continuous isotropic correlation functions which are well-defined for every dimension $k = 1,2,\ldots$. Instead let $S = \mathbb{S}^k$ and $d(u,v) = d_{\mathcal{G}}(u,v)$, cf. case (b) in Section 2.1. Then a complete characterization of continuous isotropic covariance functions is given in Schoenberg (1942); see also Gneiting (2013). In particular, r_0 defines a continuous correlation functions for any dimension $k = 1,2,\ldots$ if and only if $r_0(t) = \sum_{\ell=0}^{\infty} \beta_\ell \cos^\ell(t)$ where the sequence $\beta_0,\beta_1,\ldots$ is a probability mass function. Examples of such correlation functions are seen in Table 1. Compared to the case $S = \mathbb{R}^k$, the ranges of shape and smoothness parameters are more restrictive.

3.1.2 | The case of linear networks

Suppose $S = L$ is a linear network, cf. case (c) in Section 2.1, where we let either $d = d_{\mathcal{G}}$ or $d = d_{\mathcal{R}}$. Many of the commonly used isotropic correlation functions, including those in Table 1, are valid with respect to the resistance metric but not always with respect to the geodesic metric. The reason for this is discussed at the end of this section

Model	Correlation function $r_0(t)$	Range of shape and smoothness parameters
Powered exponential	$\exp(-t^\alpha/\phi)$	$\alpha \in (0,2]$ if $S=\mathbb{R}^k$; $\alpha \in (0,1]$ if $S=\mathbb{S}^k$ or $S=L$
Matérn	$\frac{2^{1-\alpha}}{\Gamma(\alpha)}\left(\sqrt{2\alpha}\frac{t}{\phi}\right)^\alpha K_\alpha\left(\sqrt{2\alpha}\frac{t}{\phi}\right)$	$\alpha>0$ if $S=\mathbb{R}^k$; $0<\alpha\le\frac{1}{2}$ if $S=\mathbb{S}^k$ or $S=L$
Generalized Cauchy	$(1+(\frac{t}{\phi})^\alpha)^{-\tau/\alpha}$	$\tau>0$; $\alpha\in(0,2]$ if $S=\mathbb{R}^k$; $\alpha\in(0,1]$ if $S=\mathbb{S}^k$ or $S=L$
Dagum	$1-((\frac{t}{\phi})^\tau/(1+(\frac{t}{\phi})^\tau))^{\frac{\alpha}{\tau}}$	$\tau\in(0,2]$ and $\alpha\in(0,\tau)$ if $S=\mathbb{R}^k$;
		$\tau\in(0,1]$ and $\alpha\in(0,1]$ if $S=\mathbb{S}^k$ or $S=L$

TABLE 1 Four parametric models for an isotropic correlation function $r_0(r)$. Here, Γ is the gamma function, K_ν is the modified Bessel function of the second kind, ϕ is a scale parameter, τ is a shape parameter, and α is a smoothness parameter. The correlation functions are well-defined at all scales $\phi>0$ but the range of shape and smoothness parameters depend on the model and the space S. For $S=\mathbb{R}^k$ or $S=\mathbb{S}^k$, the correlation functions are well-defined for every dimension $k=1,2,\ldots$. For $S=L$, conditions on L may be needed if distance is not measured by the resistance but the geodesic metric, see Section 3.1.2.

and is based on the following results.

Recall that a function $f:[0,\infty)\mapsto\mathbb{R}$ is completely monotonic if it is non-negative and continuous on $[0,\infty)$ and for $j=0,1,\ldots$ and all $u>0$, the j-th derivative $f^{(j)}(u)$ exists and satisfies $(-1)^j f^{(j)}(u)\ge 0$. By Bernstein's theorem, f is completely monotone if and only if it is the Laplace transform of a non-negative finite measure on $[0,\infty)$, meaning that for every $t\ge 0$,

$$f(t)=f(0)\int \exp(-st)\,\mathrm{d}F(s) \tag{6}$$

where F is a cumulative distribution function with $F(s)=0$ for $s<0$. We refer to F as the Bernstein CDF corresponding to f. The following gives a few examples of completely monotone functions obtained from using F specifying a gamma distribution $\Gamma(\tau,\phi)$ with shape parameter τ and inverse scale parameter ϕ, or an inverse gamma distribution $\Gamma^{-1}(\tau,\phi)$, or a generalized inverse Gaussian distribution. Similarly other non-negative valued distributions with a known Laplace transform can be used to produce completely monotone functions.

Example 1 *The following functions f_1,f_2,f_3 are completely monotone functions with $f_1(0)=f_2(0)=f_3(0)=1$ and they have corresponding Bernstein CDFs F_1,F_2,F_3 defined as follows. For $\tau>0$, $\phi>0$, and $t\ge 0$,*

$$f_1(t)=(1+t/\phi)^{-\tau},\quad F_1\sim\Gamma(\tau,\phi), \tag{7}$$

and

$$f_2(t)=\frac{2\phi^\tau}{\Gamma(\tau)}(t/\phi)^{\tau/2}K_\tau(2\sqrt{t\phi}),\quad F_2\sim\Gamma^{-1}(\tau,\phi). \tag{8}$$

Moreover, for $\psi>0$, $\chi>0$, $\lambda\in\mathbb{R}$, and $t\ge 0$,

$$f_3(t)=(1+2t/\psi)^{-\lambda/2}\frac{K_\lambda(\sqrt{(2t+\psi)\chi})}{K_\lambda(\sqrt{\psi\chi})} \tag{9}$$

and F_3 is the CDF for a generalized inverse Gaussian distribution with probability density function

$$\frac{(\psi/\chi)^{\lambda/2}}{2K_\lambda(\sqrt{\psi\chi})} s^{\lambda-1} \exp(-s\psi/2 - \chi/(2s)), \quad s \geq 0.$$

In the next theorem, which summaries Theorems 1 and 2 in Anderes et al. (2020), we need the following definition. We say that L is a 1-sum of $\mathcal{L}_1 = L_1 \cup \ldots \cup L_j$ and $\mathcal{L}_2 = L_{j+1} \cup \ldots \cup L_m$ if $\mathcal{L}_1$ and $\mathcal{L}_2$ are (connected) linear networks where $1 \leq j < m$, $\mathcal{L}_1 \cap \mathcal{L}_2 = \{u_0\}$ consists of a single point u_0, and

$$d(u,v) = d(u,u_0) + d(v,u_0) \quad \text{whenever } u \in \mathcal{L}_1 \text{ and } v \in \mathcal{L}_2.$$

This property is possible if $d = d_{\mathcal{G}}$ or $d = d_{\mathcal{R}}$ but unless L is a straight line segment it is impossible if d is the given by the usual Euclidean distance. Using induction we say for $n = 3, 4, \ldots$ that $L = \mathcal{L}_1 \cup \ldots \cup \mathcal{L}_n$ is a 1-sum of $\mathcal{L}_1, \ldots, \mathcal{L}_n$ if L is a 1-sum of $\mathcal{L}_1 \cup \ldots \cup \mathcal{L}_{n-1}$ and $\mathcal{L}_n$.

Theorem 2 *Let $f : [0,\infty) \mapsto \mathbb{R}$ be a completely monotone and non-constant function. Then $f(d_{\mathcal{R}}(u,v))$ is strictly positive definite over $(u,v) \in L \times L$. Moreover, if L is a 1-sum of trees and loops, then $f(d_{\mathcal{G}}(u,v))$ is strictly positive definite over $(u,v) \in L \times L$. However, if there are three distinct paths between two points on L, then there exists a constant $\phi > 0$ so that $\exp(-d_{\mathcal{G}}(u,v)/\phi)$ is not positive definite over $(u,v) \in L \times L$.*

In Table 1, for each model, r_0 is completely monotone for the ranges of the parameters (cf. the comments to Theorem 1 in Anderes et al., 2020). So by Theorem 2, if L is a 1-sum of trees and loops, for each example of r_0 in Table 1 and for r_0 given by f_1, f_2, or f_3 in (7)–(9), $r_0(d_{\mathcal{R}}(u,v))$ is a valid correlation function, but $r_0(d_{\mathcal{G}}(u,v))$ is only valid in more special cases. Note that in Table 1, the ranges of the parameters are the same when $S = \mathbb{S}^k$ or $S = L$. Finally, (7) is the special case of the generalized Cauchy function when $\alpha = 1$ in Table 1, whilst (8) and (9) are not covered by Table 1.

For example, consider the Chicago street network in Figure 1 (a standard example of a linear network, cf. Ang et al., 2012)). Then L is not a a 1-sum of trees and loops and therefore we have not used the geodesic metric for the cases of covariance functions related to Figure 1 (more details on this figure is given in Section 3.2). See also the counter examples in Anderes et al. (2020, Section 5).

3.2 | Simulation of GPs on linear networks

This section discusses how to simulate a GP $Y = \{Y(u) \mid u \in L\}$ on a linear network $L = \cup_{i=1}^m L_i$ as given in case (c) in Section 2.1. We assume without loss of generality that the mean function of Y is zero.

A straightforward algorithm applicable to any metric d and any linear network L consists in first selecting finite subsets $D_j \subset L_j$, $j = 1, \ldots, m$; then simulating Y restricted to the set D given by the union of V and $D_1, \ldots, D_m$, e.g. by using Choleski decomposition of the corresponding covariance matrix Σ_D; and finally, for u not in V or $\cup_{j=1}^m D_j$, approximating $Y(u)$ by the average of those $Y(v)$ where $v \in D$ is closest to u ('closest' with respect to the given metric d or perhaps better $d_{\mathbb{G}}$). The disadvantage of this algorithm is of course that the dimension of Σ_D can be large and hence Choleski decomposition (as well as other methods) can be slow.

If c is an exponential covariance function, L is a tree, and $d = d_{\mathcal{G}} = d_{\mathcal{R}}$, the algorithm in the Theorem 4 below is much faster to use. For $s > 0$, the exponential correlation function $r_0(t) = \exp(-ts)$ appears as two special cases in Table 1 with scale parameter $\phi = 1/s$, namely the powered exponential model with $\alpha = 1$ and the Mátern model with

$\alpha = \frac{1}{2}$. To prove Theorem 4 we first need to establish a Markov property given in the following theorem, where we denote the shortest path between $u, v \in L$ by p_{uv}.

Theorem 3 *Suppose that for $s > 0$ and $\sigma > 0$, Y is a GP on a tree L with exponential covariance function $c(u,v) = \sigma^2 \exp(-sd(u,v))$ with $d = d_{\mathcal{G}} = d_{\mathcal{R}}$. If $u, v, w \in L$ with $w \in p_{uv}$ then $Y(u)$ and $Y(v)$ are conditionally independent given $Y(w)$. Furthermore, for $w_1, \ldots, w_n \in L$, $Y(u)$ and $Y(v)$ are conditionally independent given $Y(w_1), \ldots, Y(w_n)$ provided $w_i \in p_{uv}$ for at least one w_i.*

Proof For the first statement in the theorem, note that since $w \in p_{uv}$ then $d(u,v) = d(w,v) + d(w,v)$ and therefore $c(u,v) = c(u,w)c(w,v)/\sigma^2$. Thus the covariance matrix for (u,v,w) has the form

$$\Sigma_{u,v,w} = \begin{pmatrix} \sigma^2 & c(u,w)c(w,v)/\sigma^2 & c(u,w) \\ c(u,w)c(w,v)/\sigma^2 & \sigma^2 & c(w,v) \\ c(u,w) & c(w,v) & \sigma^2 \end{pmatrix}.$$

Inverting the covariance matrix, we get that the corresponding precision matrix has 0 at entries $(1,2)$ and $(2,1)$, thus implying that $Y(u)$ and $Y(v)$ are conditionally independent given $Y(w)$.

For the second statement in the theorem, note that the case $n = 1$ is the first statement. Consider the case $n = 2$, i.e., we condition on $Y(w_1)$ and $Y(w_2)$, where $w_1, w_2 \in L$ and e.g. $w_1 \in p_{uv}$. Since L is a tree $Y(w_2)$ must be conditionally independent of either $Y(u)$ or $Y(v)$ given $Y(w_1)$. Assume without loss of generality that this is $Y(u)$. Thus, $Y(u)$ is conditionally independent of $(Y(v), Y(w_2))$ given $Y(w_1)$, which implies that $Y(u)$ and $Y(v)$ are conditionally independent given $(Y(w_1), Y(w_2))$. In a similar way we verify the case with $n \geq 3$.

Using Theorem 3, we can now prove that the simulation algorithm given in Theorem 4 below generates a GP with an exponential covariance function on a tree. Pick an arbitrary origin $u_0 \in V$ and set $G_0(u_0) = \{u_0\}$. For $j = 1, 2, \ldots$, if $u \in V \setminus \cup_{i=0}^{j-1} G_i(u_0)$ and $u \sim v$ for some $v \in G_{j-1}(u_0)$, we call u a child of j-th generation to u_0 and define $G_j(u_0) \subset V$ as the set of all children of j-th generation to u_0. Moreover, for the GP $Y = \{Y(w) \mid w \in L\}$ constructed in the following theorem, if $u \in G_{j-1}(u_0)$, $v \in G_j(u_0)$, and $u \sim v$, we define $Y(u,v) = \{Y(w) \mid w \in (u,v]\}$ where $(u,v]$ is the half-open line segment with endpoints u and v so that u is excluded and v is included.

Theorem 4 *Suppose that L is a tree, and that $s > 0$ and $\sigma > 0$. Consider the following iterative construction of random variables $Y(w)$ where $u_0 \in V$ is arbitrary.*

- *For $w = u_0$, generate $Y(w)$ from $N(0, \sigma^2)$.*
- *For $j = 1, 2, \ldots$, conditioned on all the $Y(w)$ so far generated, generate independent GPs $Y(u,v)$ for all $u \in G_{j-1}(u_0)$ and all $v \in G_j(u_0)$ with $u \sim v$, where $Y(u,v)$ depends only on $Y(u)$ and for every $w, w_1, w_2 \in (u,v]$ we have*

$$\mathbb{E}[Y(w) \mid Y(u)] = \exp(-\|w-u\|s))Y(u) \tag{10}$$

$$\mathbb{C}\mathrm{ov}[Y(w_1, w_2) \mid Y(u)] = \sigma^2 \left(\exp(-\|w_1 - w_2\|s) - \exp(-\|w_1 - u\|s - \|w_2 - u\|s)\right). \tag{11}$$

- *Output $Y = \{Y(w) \mid w = u_0$ or $w \in (u,v]$ for some $j \in \mathbb{N}$, $u \in G_{j-1}(u_0)$, $v \in G_j(u_0)$ with $u \sim v\}$.*

Then Y is a mean-zero GP on L with exponential covariance function $c(u,v) = \sigma^2 \exp(-sd(u,v))$ where $d = d_{\mathcal{G}} = d_{\mathcal{R}}$.

Proof We prove by induction that if Y is as stated at the end of theorem, then the iterative construction is correct. Clearly the distribution of $Y(u_0)$ is correct; let us denote this step by $j = 0$. For step $j \geq 1$, condition on all the $Y(w)$ so far generated, which by the induction hypothesis have been correctly generated. If we take two points w_1 and w_2 contained in different line segments between points in $G_{j-1}(u_0)$ and $G_j(u_0)$, then by Theorem 3, $Y(w_1)$ and $Y(w_2)$ are (conditionally) independent, which is in accordance to our construction. So it suffices to consider the (conditional) distribution of $(Y(w_1), Y(w_2))$ when $w_1, w_2 \in (u, v]$, $u \in G_{j-1}(u_0)$, $v \in G_j(u_0)$, and $u \sim v$. This (conditional) distribution depends only on $Y(u)$, cf. Theorem 3, and it is straightforwardly seen to be a bivariate normal distribution with mean and covariance matrix given by 10. Therefore the output of the algorithm has the correct distribution.

In practice, when using Theorem 4 for simulation, a discretization on each line segment is needed. For some integer $n_j > 0$ (which may depend on l_j) and each $u \in L_j$, $Y(u)$ is approximated by $Y(u_j(s))$ if u is closest to $u_j(s_i)$ with $s_i = i l_j / n_j$ and $i \in \{0, 1, \ldots, n_j\}$; or, if u is the midpoint between s_{i-1} and s_i, we approximate $Y(u)$ by the average $(Y(u_j(s_{i-1})) + Y(u_j(s_i)))/2$. To generate the $n_j - 1$ normal variables $Y(u_j(s_i))$, $i = 1, \ldots, n_j - 1$, we start by generating the variable $Y(u_j(s_1))$ in accordance to (10)-(11) with $w = w_1 = w_2 = s_1$. Then we can add $u_j(s_1)$ to the vertex set, whereby we split l_j into the two line segments given by this new vertex and the endpoints of l_j. Hence, if $n_j > 1$, we can repeat the procedure when generating $Y(u_j(s_2))$, and so on until all the $n_j - 1$ normal variables have been generated.

It is important to notice that Theorems 3 and 4 do not hold if L is not a tree. Indeed, a GP on the circle $\mathbb{S}^1$ (which is equivalent to a loop of length 2π) with exponential covariance function, considering four arbitrary points on $\mathbb{S}^1$, it can be shown that the GP is not Markov. On the other hand, letting $c(u, v) = a \cosh(b(d_{\mathcal{G}}(u, v) - \pi))$ for $a, b > 0$, Pitt (1971) verified that the GP on $\mathbb{S}^1$ with covariance function c is Markov, but considering two arbitrary points on a tree together with a point on the path connecting them, it can be shown that the GP on the tree with covariance function c is not Markov. Consequently, we cannot have a covariance function only depending on the geodesic distance which makes an arbitrary linear network Markov, and in general, if we want to simulate a GP with an exponential covariance function on a linear network which is not a tree, we cannot rely on Markov properties and have to use the straightforward, but slower, algorithm described in the beginning of this section.

For other covariance functions than the exponential, Theorem 4 can be used in connection to the following theorem which follows from (6) and the central limit theorem.

Theorem 5 *Suppose d is a metric on L so that $(u, v) \mapsto \exp(-s d(u, v))$ for $(u, v) \in L^2$ is a well-defined correlation function for all $s > 0$, and let Y be a mean-zero GP on L with covariance function*

$$c(u, v) = \sigma^2 \int \exp(-s d(u, v)) \, \mathrm{d}F(s) \tag{12}$$

where $\sigma > 0$ and F is a CDF with $F(s) = 0$ for $s < 0$. For an integer $n > 0$ and $i = 1, \ldots, n$, generate S_i from F and then Y_i as a mean-zero GP on L with covariance function $\sigma^2 \exp(-S_i d(u, v))$ so that $(S_1, Y_1), \ldots, (S_n, Y_n)$ are independent. Calculate $\bar{Y}_n = \sum_{i=1}^n Y_i / n$. Then $\sqrt{n}\bar{Y}_n$ is a mean-zero stochastic process on L with covariance function c. As $n \to \infty$, $\sqrt{n}\bar{Y}_n$ approximates Y in the sense that any finite dimensional distribution of $\sqrt{n}\bar{Y}_n$ converges in distribution towards the corresponding finite dimensional distribution of Y.

Theorem 5 allows simulation of any GP with a covariance function of the form (12), if a simulation algorithm for F is available and the metric d satisfies the condition in the theorem. For the case of a tree and $d = d_{\mathcal{G}} = d_{\mathcal{R}}$, this gives a fast simulation algorithm when combined with the algorithm in Theorem 4. For other cases it may be faster

simply to use the straight forward algorithm in the beginning of this section, provided of course that d satisfies the condition in the theorem (so it may work for $d = d_{\mathcal{R}}$ but not for $d = d_{\mathcal{G}}$, cf. Theorem 2).

Figure 1 shows examples of simulations of zero mean GPs defined on the Chicago street network and with various isotropic covariance functions $r(u, v) = r_0(d(u, v))$ where $d = d_{\mathcal{R}}$ is the resistance metric. In the first two plots (the top row), $r_0(t) = \exp(-st)$ is an isotropic exponential correlation function, the next two plots (the middle row) relate to Theorem 5 with the Bernstein CDF given by a gamma distribution or inverse gamma distribution (see the left panel in the bottom row), and the last plot shows the corresponding correlation functions for $t \leq 200$ feet (the side length of a square surrounding the network is a little less than 1000 feet and the maximal distance with respect to the resistance metric is about 675 feet). Since the resistance metric is used, the covariance functions for the GPs in plots 1–4 are well-defined, cf. Theorem 2. The top row shows the scaling effect of the parameter s for the exponential correlation function. Plots 2–4 are comparable, since the mean values of s all agree, and the last plot indicates that the correlation is smallest when r is fixed and rather similar when using the gamma or inverse gamma distribution. Accordingly, in plot 2 we see less smoothness than in plots 3 and 4 which show a similar degree of smoothness.

4 | POINT PROCESSES AND THEIR CHARACTERISTICS

4.1 | Setting

Consider again a general metric space (S, d), where $S = L$ is of our primary interest, cf. Section 2.1. We restrict attention to point processes whose realisations can be viewed as locally finite subsets x of S: For $x, A \subseteq S$, define $x_A = x \cap A$ and let $n(x_A)$ denote the cardinality of x_A, then the state space of a point process is $\mathcal{N} = \{x \subset S \mid n(x_A) < \infty \text{ whenever } A \subseteq S \text{ is bounded}\}$. Equip $\mathcal{N}$ with the smallest σ-algebra such that the mapping $x \mapsto n(x_A)$ is measurable whenever $A \subseteq S$ is a bounded Borel set. In this paper, by a point process is meant a random variable X with values in $\mathcal{N}$ (in the terminology of point process theory, X is a simple locally finite point process, see e.g. Daley and Vere-Jones, 2003). Hence, for any bounded Borel set $A \subseteq S$, the count $N(A) = n(X_A)$ is a random variable.

We need the following notions where ν is a reference measure on S, cf. Section 2.1. We say that X is a Poisson process with intensity function $\rho : S \mapsto [0, \infty)$ if for any bounded Borel set $A \subseteq L$, $N(A)$ is Poisson distributed with finite mean $\int_A \rho(u)\, \mathrm{d}\nu(u)$, and conditioned on $N(A)$, the points in X_A are independent and each point has a density proportional to ρ with respect to ν restricted to A. For every $u \in L$, we let X_u be the point process which follows the reduced Palm distribution of X at u, that is,

$$\mathbb{E} \sum_{u \in X} h(X \setminus \{u\}, u) = \int \rho(u) \mathbb{E} h(X_u, u)\, \nu(u)$$

for any non-negative measurable function h defined on $\mathcal{N} \times L$. Intuitively, X_u follows the distribution of $X \setminus \{u\}$ conditioned on that $u \in X$ (see e.g. Møller and Waagepetersen, 2004, Appendix C). If $\rho(u) = 0$, X_u may follow an arbitrary distribution. If X is Poisson process with intensity function ρ, then X and X_u are identically distributed whenever $\rho(u) > 0$.

Let X_1 denote the Poisson process with intensity 1. Suppose $S = \mathbb{S}^k$ or $S = L$, and X has density f with respect to X_1 (implicitly assuming the distribution of X is absolutely continuous with respect the distribution of X_1). If $\rho(u) > 0$, then X_u has a density f_u with respect to X_1 such that

$$f_u(\{u_1, \ldots, u_n\}) = f(\{u, u_1, \ldots, u_n\})/\rho(u) \quad \text{for } n = 1, 2, \ldots \text{ and pairwise distint } u_1, \ldots, u_n \in S \setminus \{u\}. \tag{13}$$

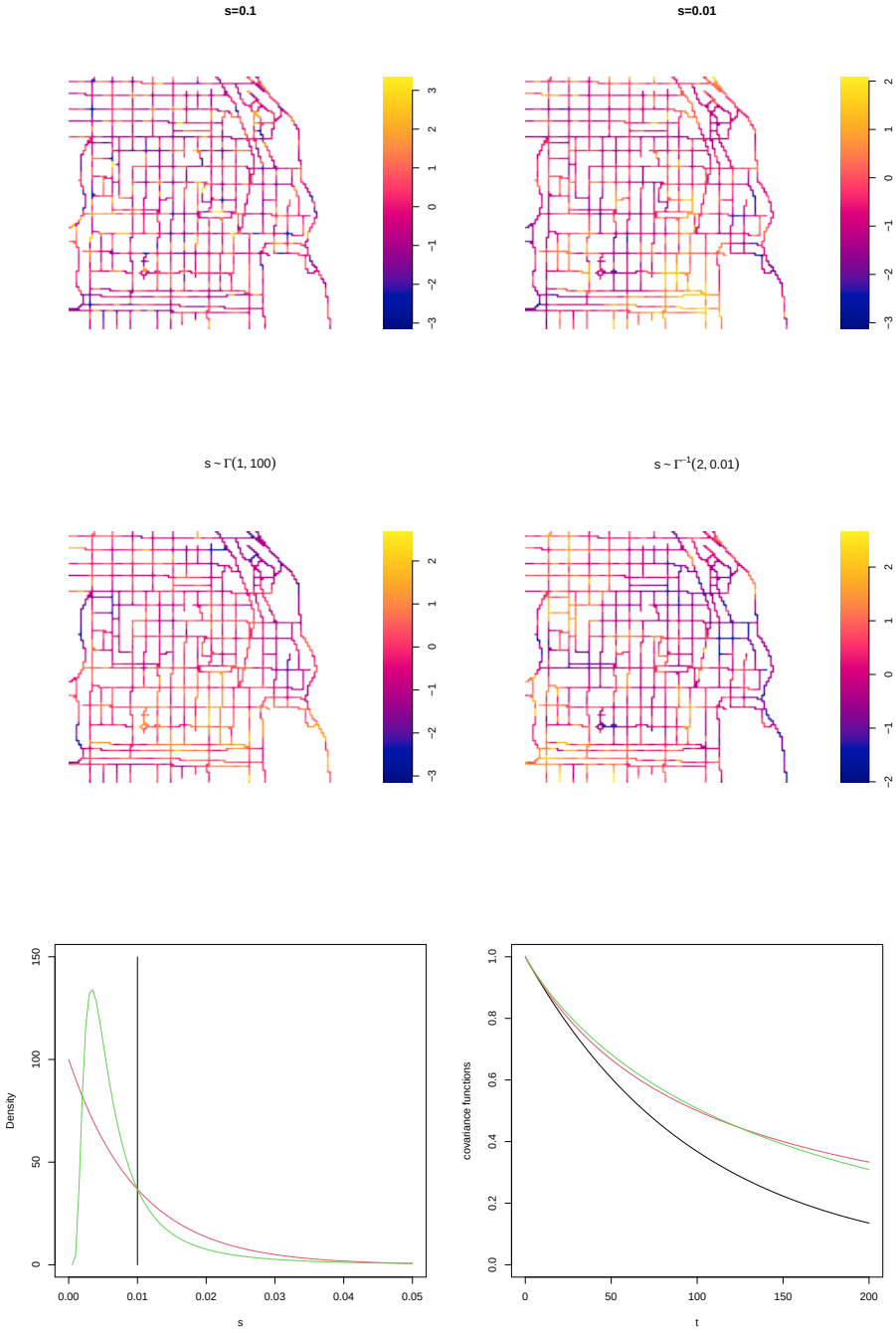

FIGURE 1 Simulation of zero mean GPs on the Chicago street network with an isotropic exponential covariance function $c(u, v) = r_0(d_R(u, v))$. Top row: When $r_0(t) = \exp(-st)$ with parameter $s = 0.1$ or $s = 0.01$. Middle row: When r_0 is a mixture of exponential correlation functions with s following a gamma distribution or inverse gamma distribution, where in both cases the mean of s is 0.01. Bottom row: For plots 2–4, the corresponding densities for F and correlation functions, where the curves in black, red, and green correspond to plots 2–4, respectively.

4.2 | Moment and invariance properties

Let S and the reference measure ν be as in one of the cases (a)–(c) in Section 2.1. The point process X has n-th order intensity function $\rho(u_1,\dots,u_n)$ (with respect to the n-fold product measure of ν) if this is a non-negative Borel function so that

$$\mathbb{E}\left[N(A_1)\cdots N(A_n)\right] = \int_{A_1}\cdots\int_{A_n}\rho(u_1,\dots,u_n)\,\mathrm{d}\nu(u_1)\cdots\mathrm{d}\nu(u_n) < \infty \tag{14}$$

for every pairwise disjoint, bounded, Borel sets $A_1,\dots,A_n \subset S$ ($\rho(u_1,\dots,u_n)$ is also called the n-th order product density for the n-th order reduced moments measure). Thus, $\rho(u_1,\dots,u_n)$ is a locally integrable function, which is almost everywhere unique on S^n (with respect to the n-fold product measure of ν). In the following, for simplicity nullsets are ignored, so the non-uniqueness of $\rho(u_1,\dots,u_n)$ is ignored. Moreover, when we write $\rho(u_1,\dots,u_n)$ it is implicitly assumed that the n-th order intensity function exists.

In particular, $\rho(u)$ is the usual intensity function. The point process is said to be (first-order) homogeneous if $\rho(u)=\rho_0$ is constant. For instance, this is the case if $S=\mathbb{R}^k$ and X is stationary, i.e., its distribution is invariant under translations in $\mathbb{R}^k$; or if $S=\mathbb{S}^k$ and X is isotropic, i.e., its distribution is invariant under rotations on $\mathbb{S}^k$. A similar example is not easy to specify if $S=L$, since there is no natural (transitive) group action on a linear network.

Instead of the second order intensity function, one usually considers the pair correlation function (pcf) given by

$$g(u,v) = \frac{\rho(u,v)}{\rho(u)\rho(v)},$$

setting $\frac{a}{0}=0$ for $a\geq 0$. For a Poisson process, $\rho(u_1,\dots,u_n)=\rho(u_1)\cdots\rho(u_n)$, so $g=1$. One often interprets $g(u,v)\leq 1$ as repulsion/inhibition/regularity and $g(u,v)\geq 1$ as attraction/clustering/aggregation, though care should be taken if u and v are 'distant apart'.

The specific models in this paper are typically attractive ($g\geq 1$) and satisfies the following stronger property: For $n=2,3,\dots$, $i=1,\dots,n-1$, and any pairwise distinct $u_1,\dots,u_n\in S$,

$$\rho(u_1,\dots,u_n) \geq \rho(u_1,\dots,u_i)\rho(u_{i+1},\dots,u_n), \tag{15}$$

or equivalently, for any pairwise disjoint, bounded, Borel sets $A_1,A_2,\dots\subset S$,

$$\mathbb{E}\left[N(A_1)\cdots N(A_n)\right] \geq \mathbb{E}\left[N(A_1)\cdots N(A_i)\right]\mathbb{E}\left[N(A_{i+1})\cdots N(A_n)\right].$$

In other words, (15) means that the counts $N(A_1), N(A_2),\dots$ are positively correlated at all orders, and for brief we shall say that X is positively correlated at all orders.

Non-parametric estimation of ρ and g are discussed in e.g. Baddeley et al. (2015) and the references therein; see also Shaw et al. (2021) when $S=\mathbb{R}^k$; Lawrence et al. (2016) when $S=\mathbb{S}^k$; and Rakshit et al. (2017) and Rakshit et al. (2019) when $S=L$. For non-parametric estimation of the pcf, kernel methods are used. Since these may be sensible to the choice of bandwidth, popular alternatives which avoid using kernel methods are given by estimators of the K-function in (16)–(17) below. On the other hand, for the specific parametric models in this paper, we have simple expressions for g but not for K, and since the K-function is an accumulated versions of g, it may be harder to interpret (we return to this at the end of Section 4.2.2).

4.2.1 | K-functions in the classical cases of Euclidean spaces and spheres

Let either $(S,d) = (\mathbb{R}^k, \|\cdot\|)$ or $(S,d) = (\mathbb{S}^k, d_{\mathcal{G}})$, cf. cases (a)–(b) in Section 2.1, and suppose $g(u,v) = g_0(d(u,v))$ is isotropic. This is satisfied if $S = \mathbb{R}^k$ and X is stationary and isotropic or if $S = \mathbb{S}^k$ and X is isotropic, but we do not assume that X is homogeneous. Following Baddeley et al. (2000), Lawrence et al. (2016), and Møller and Rubak (2016), we say that X is second-order intensity reweighted stationary (SOIRS) if $S = \mathbb{R}^k$ and second order intensity reweighted isotropic (or pseudo/correlation isotropic) if $S = \mathbb{S}^k$, and the (inhomogeneous) K-function is for an arbitrary $u \in S$ given by

$$K(t) = \int_{d(u,v)\le t} g(u,v)\,\mathrm{d}v(v). \tag{16}$$

So

$$K(t) = K_{\mathbb{R}^k}(t) = \sigma_{k-1}\int_0^t r^{k-1} g_0(r)\,\mathrm{d}r \quad \text{if } S = \mathbb{R}^k \text{ and } t \ge 0$$

where $\sigma_{k-1} = 2\pi^{k/2}/\Gamma(k/2)$ is the surface area of the $(k-1)$-dimensional unit sphere, and

$$K(t) = K_{\mathbb{S}^k}(t) = \sigma_{k-1}\int_0^t g_0(\vartheta)\sin^{k-1}\vartheta\,\mathrm{d}\vartheta \quad \text{if } S = \mathbb{S}^k \text{ and } 0 \le t \le \pi.$$

Thus, for $k = 1$, $K_{\mathbb{R}}(t) = K_{\mathbb{S}}(t)$ agree for every $t \in [0,\pi]$.

4.2.2 | K-functions in the case of linear networks

Suppose $S = L$ and $g(u,v) = g_0(\delta(u,v))$ where δ is a metric on L (we switch from the previous notation d to δ for convenience since we consider derivatives below). Then X is said to be δ-correlated, cf. Rakshit et al. (2017); if $\delta = d_{\mathcal{G}}$, X is also said to be second-order reweighted pseudostationary (Ang et al., 2012). Following Rakshit et al. (2017) and defining $R = \inf_{u\in L}\sup_{v\in L} d(u,v)$, the K-function is given by

$$K(t) = K_L(t) = \int_0^t g_0(r)\,\mathrm{d}r, \quad 0 \le t \le R. \tag{17}$$

Note that K depends only on g_0, but g depends on both g_0 and δ. We have $K_L = K_{\mathbb{R}}/2$ if L is a straight line segment which is broken into m line segments. If X is a Poisson process, then $K(t) = t$.

Non-parametric estimation of K is carefully studied in Rakshit et al. (2017) when the following technical assumption for the metric is made. Suppose δ is regular, meaning that for every $u \in L$, $\delta(u,v)$ is a continuous function of $v \in L$ and there is a finite set $N \subset L$ such that for $i = 1,\dots,m$ and all $v \in L_i \setminus N$, the Jacobian

$$J_\delta(u,v) = |(\mathrm{d}/\mathrm{d}t)\delta(u,v)|$$

exists and is non-zero where $t = \|v - a_i\|$. Both $d_{\mathcal{G}}$ and $d_{\mathcal{R}}$ are regular, where $J_{d_{\mathcal{G}}} = 1$ and a useful expression for the calculation of $J_{d_{\mathcal{R}}}$ is given in the following corollary which follows immediately from Proposition 1.

Corollary 1 *For all $u \in L_j$ and $v \in L_i$ with $u \neq v$, using a notation as in Proposition 1, we have*

$$\frac{\mathrm{d}}{\mathrm{d}t} d_R(u,v) = \begin{cases} 2A_i(t-s)+1 & \text{if } i=j,\ t>s, \\ 2A_i(t-s)-1 & \text{if } i=j,\ t<s, \\ 2A_i t + B_{ij}(s) & \text{if } i \neq j. \end{cases} \tag{18}$$

Some final remarks are in order. For $u \in L$ and $0 \le t \le R$, define

$$w_\delta(u,t) = 1 \Big/ \sum_{v \in L:\, \delta(u,v)=t} 1/J_\delta(u,v).$$

This is a weight which accounts for the geometry of the linear network when shifting from arc length measure on L to Lebesgue measure on the positive half-line (Rakshit et al., 2017, Propositions 1 and 2). For $\delta = d_{\mathcal{G}}$, we have $w_{\mathcal{G}}(u,t) = 1/\#\{v \in L \mid \delta(u,v) = t\}$, since $J_{d_{\mathcal{G}}} = 1$, and for $\delta = d_R$, once the matrix Σ from Section 2.2 has been calculated, w_{d_R} is quickly calculated from (18). It follows from (14), (17), and Rakshit et al. (2017, Equation (8)) that for any Borel set $A \subseteq L$ of positive arc length measure,

$$K(t) = \frac{1}{\nu_L(A)} \mathbb{E} \sum_{u \in X_A} \sum_{v \in X \setminus \{u\}} \frac{1(\delta(u,v) \le t) w_\delta(u, \delta(u,v))}{\rho(u)\rho(v)}. \tag{19}$$

Non-parametric estimators of K are based on omitting the expectation symbol in (19), possibly after elaborating on the right hand side in (19) in order to realize how correction factors can be included in order to adjust for edge effects, cf. Rakshit et al. (2017). In terms of Palm probabilities, for d_L-almost all $u \in L$ with $\rho(u) > 0$,

$$K(t) = \frac{1}{|L|} \mathbb{E} \sum_{v \in X_u} \frac{1(\delta(u,v) \le t) w_\delta(u, \delta(u,v))}{\rho(v)}.$$

In general the weight makes it hard to interpret this expression of K.

4.3 | Estimation and model checking

For the parametric families of Cox point process models as considered in this as well as other papers, the most common estimation methods are based on the intensity, pair correlation, or K-functions using either minimum contrast estimation, composite likelihood, or Palm likelihoods, see Møller and Waagepetersen (2007, 2017) and the references therein. (Some more text will be added here)

For model checking other functional summaries are needed when ρ, g, or K and their corresponding non-parametric estimators have been used for estimation. For the case $S = L$, Cronie et al. (2020) define analogies to the F (the empty space), G (the nearest-neighbour distribution), and J-functions in Van Lieshout (2011). The main result in Cronie et al. (2020) is that these definitions make good sense under a certain condition called intensity reweighted moment pseudostationarity (IRMPS): X is IRMPS if $\inf \rho > 0$ and δ is a regular metric on L such that for $n = 2, 3, \ldots$, any pairwise distinct $u_1, \ldots, u_n \in L$, and any $u \in L$, $g(u_1, \ldots, u_n) = \rho^{(n)}(u_1, \ldots, u_n)/[\rho(u_1) \cdots \rho(u_n)]$ is of the form

$$g(u_1, \ldots, u_n) = g_0(\delta(u, u_1), \ldots, \delta(u, u_n)) \tag{20}$$

for some function g_0. This condition is satisfied if X is a Poisson process or if L equal to a line segment and X is a LGCP having a stationary pair correlation function (see Section 5.2.1), but apart from these examples Cronie et al. (2020) do not correctly verify any other cases of models where IRMPS is satisfied, and in our opinion IRMPS is a very restrictive assumption (see again Section 5.2.1). However, Cronie et al. (2020) demonstrate for examples of planar linear networks the practical usefulness of empirical estimators of the inhomogeneous linear J-function for both a Poisson process, a simple sequential inhibition (SSI) point process, and a LGCP (as we show in Section 5.2.1, IRMPS is in general not satisfied for the LGCP, and we expect it is also not satisfied for the SSI point process where it is hard to evaluate $\rho^{(n)}$ for $n \geq 2$). Furthermore, Christensen and Møller (2020) introduce three purely empirical summary functions obtained by modifying the empirical F, G, and J-functions for inhomogeneous point patterns on a Euclidean space to linear networks. Briefly, the modification consists of replacing the Euclidean space with the linear network, introducing the shortest path distance instead of the Euclidean distance, and adapting the notion of an eroded set to linear networks. Christensen and Møller (2020) demonstrate the usefulness of the empirical summary functions when modelling spine locations on dendrite trees.

To validate a fitted model when considering one or more empirical summary functions, the most popular method uses simulations to obtain confidence regions called global envelopes and p-values for global envelope tests based on the extreme rank length as described in Myllymäki et al. (2017), Mrkvička et al. (2020), and Myllymäki and Mrkvička (2019). (Some more text will be added here)

5 | COX PROCESSES DRIVEN BY TRANSFORMED GAUSSIAN PROCESSES

5.1 | General setting

Let X be a point process on S as considered in Section 4.1 and let $\Lambda = \{\Lambda(u) \mid u \in S\}$ be a non-negative stochastic process so that with probability 1, for any bounded Borel set $A \subseteq S$, the measure $\xi(A) = \int_A \Lambda(u)\,\mathrm{d}\nu(u)$ is finite. Suppose X is a Cox process driven by Λ, that is, X conditioned on Λ is almost surely a Poisson process with intensity function Λ. Usually in applications Λ is unobserved, and so the Cox process X is indistinguishable from the inhomogeneous Poisson process $X|\Lambda$ when only one point pattern dataset is available (for a discussion of which of two models is most appropriate, see Møller and Waagepetersen, 2004, Chapter 5).

Henceforth, assume $\mathbb{E}\Lambda(u)$ is a locally integrable function with respect to ν. Then the Cox process X is well-defined and has intensity function $\rho(u) = \mathbb{E}\Lambda(u)$. Let $W \subseteq S$ be a bounded Borel set which we think of as an observation window with $\nu(W) > 0$. Then X_W is a Cox process driven by Λ restricted to W, and X_W has a density given by

$$f(\{u_1,\dots,u_n\}) = \mathbb{E}\left[\exp\left(\int_W (1-\Lambda(u))\,\mathrm{d}\nu(u)\right)\prod_{i=1}^n \Lambda(u_i)\right] \quad \text{for pairwise distint } u_1,\dots,u_n \in W \tag{21}$$

with respect to $X_1 \cap W$, where X_1 is the unit rate Poisson process, cf. Section 4.1. In particular, if $S = \mathbb{S}^k$ or $S = L$, letting $W = S$, then for $u \in S$ with $\rho(u) > 0$, X_u has density

$$f_u(\{u_1,\dots,u_n\}) = \mathbb{E}\left[\exp\left(\int_W (1-\Lambda(u))\,\mathrm{d}\nu(u)\right)\frac{\Lambda(u)}{\rho(u)}\prod_{i=1}^n \Lambda(u_i)\right] \tag{22}$$

with respect to X_1, cf. (13). In general the densities in (21) and (22) are intractable because the expected values are difficult to evaluate. Instead the following moment properties are exploited for inference: By conditioning on Λ it

follows from (14) and the moment properties of the Poisson process that

$$\rho(u_1,\dots,u_n) = \mathbb{E}\left[\Lambda(u_1)\cdots\Lambda(u_n)\right].$$

It is often useful to write Λ as

$$\Lambda(u) = \rho(u)\Lambda_0(u) \tag{23}$$

where $\Lambda_0 = \{\Lambda_0(u) \mid u \in S\}$ is a non-negative 'residual' stochastic process with $\mathbb{E}\Lambda_0(u) = 1$ whenever $\rho(u) > 0$. Then

$$\rho(u_1,\dots,u_n) = \rho(u_1)\cdots\rho(u_n)\mathbb{E}\left[\Lambda_0(u_1)\cdots\Lambda_0(u_n)\right]$$

and

$$g(u,v) = \mathbb{E}\left[\Lambda_0(u)\Lambda_0(v)\right]. \tag{24}$$

For most statistical models considered in the literature, it is only $\rho(u)$ which is allowed to depend on covariate information, whilst Λ_0 is considered to account for unobserved covariates or other effects which has not been successfully fitted by a Poisson process with intensity function ρ, see e.g. Møller and Waagepetersen (2007) and Diggle (2014). Usually, Λ_0 is assumed to satisfy additional conditions such as stationarity if $S = \mathbb{R}^k$, isotropy if $S = \mathbb{S}^k$, or homogeneity if $S = L$ as exemplified many places in the following. Furthermore, for specific models of Λ_0 it usually happens that

$$\mathbb{E}\left[\Lambda_0(u_1)\cdots\Lambda_0(u_n)\right] \geq \mathbb{E}\left[\Lambda_0(u_1)\cdots\Lambda_0(u_i)\right]\mathbb{E}\left[\Lambda_0(u_{i+1})\cdots\Lambda_0(u_n)\right] \tag{25}$$

for $n = 2, 3, \ldots, i = 2, \ldots, n$, and all pairwise distinct $u_1, \ldots, u_n \in S$, meaning that X is positively correlated at all orders, cf. (15).

5.2 | Models

Consider a GP $Y = \{Y(u) \mid u \in S\}$ with mean function μ and covariance function c, and let $Y_1, \ldots, Y_h$ be independent copies of Y. In the remainder of this paper we study the following models.

- X is a log Gaussian Cox process (LGCP) if

  $$\Lambda_0(u) = \exp(Y(u)) \tag{26}$$

  and $\mu(u) = -c(u,u)/2$ for all $u \in S$. The latter condition is required since we want $\mathbb{E}\Lambda_0(u) = 1$.
- Assume $\mu = 0$. If $\Pi(u) = \exp(-\sum_{i=1}^{h} Y_i(u)^2)$ and

  $$\Lambda_0(u) = \Pi(u)(1 + 2c(u,u))^{h/2} \tag{27}$$

  for all $u \in S$, then X is a interrupted Cox process (ICP). Since $\mathbb{E}\Pi(u) = (1 + 2c(u,u))^{-h/2}$, we have $\mathbb{E}\Lambda_0(u) = 1$.

Note that the definition of Π differs slightly from the one used in Lavancier and Møller (2016), which includes a factor $1/2$ inside the exponential function.

- Assume $\mu = 0$ and $c(u, u) = 1$ for all $u \in L$ (so c is a correlation function). If

$$\Lambda_0(u) = \frac{1}{h} \sum_{i=1}^{h} Y_i(u)^2 \tag{28}$$

then X is a permanental point process (PPP). Since the sum in (28) is $\chi^2(h)$-distributed, $\mathbb{E}\Lambda_0(u) = 1$.

In all cases, the distribution of X is completely specified by (ρ, c) and in the case of an ICP or PPP the value of h. Note that the intensity function ρ can be any non-negative locally integrable function with respect to ν.

LGCP, ICP, and PPP models are well-studied when $S = \mathbb{R}^k$ or $S = \mathbb{S}^k$, and most of their properties immediately extend to the case $S = L$ as discussed in the following.

5.2.1 | Log Gaussian Cox processes

Let X be a LGCP, cf. (26). Møller et al. (1998) studied the case $S = \mathbb{R}^k$, and Cuevas-Pacheco and Møller (2018) the case $S = \mathbb{S}^k$. As in Møller et al. (1998) and Coeurjolly et al. (2017), we obtain for the general setting of the space S the following results.

For any integer $n \geq 2$ and any pairwise distinct $u_1, \ldots, u_n \in S$,

$$\rho(u_1, \ldots, u_n) = \rho(u_1) \cdots \rho(u_n) \exp\Big(\sum_{1 \leq i < j \leq n} c(u_i, u_j) \Big). \tag{29}$$

In particular, the LGCP is determined by ρ and $g = \exp(c)$, i.e., by its first and second order moment properties. In most specific models, including those in Table 1, $c \geq 0$ or equivalently X is positively correlated at all orders, cf. (15) and (29). If $c(u, v) = c_0(d(u, v))$ is isotropic, then $\rho(u_1, \ldots, u_n)$ depends only on the inter-point distances $d(u_i, u_j)$, $1 \leq i < j \leq n$.

For any $u \in S$ with $\rho(u) > 0$, the reduced Palm distribution of X_u is a LGCP with intensity function $\rho(v|u) = \rho(v) \exp(c(u, v))$ but the pair correlation function is still $g(v, w|u) = g(v, w) = \exp(c(v, w))$ for $v, w \in S$. This follows from (21) and (22) when $S = W = \mathbb{S}^k$ or $S = W = L$; if $S = \mathbb{R}^k$, see Coeurjolly et al. (2017). See also the discussion in Section 7 on K-functions for X and X_u.

Assuming $S = L$ is a linear network, let us return to the concept of IRMPS as defined by (20). Cronie et al. noticed that IRMPS is satisfied for the LGCP if $\inf \rho > 0$ and for all $u_1, u_2, u \in L$,

$$c(u_1, u_2) = c_1(\delta(u, u_1), (u, u_2)) \tag{30}$$

for some function c_1 (in our notation; see Cronie et al., 2020, Equation (29)). This statement is true due to (29), however, in our opinion (30) is a very strong condition, since we are not aware of any good examples unless L is isometric to a closed interval and δ is usual (Euclidean/geodesic/resistance) distance. Incidentally, in Cronie et al. (2020, Lemma 2) the metric δ is assumed to be origin independent; they do not define the meaning of 'origin independent' but we have been informed (by personal communication) that they actually mean that (30) should be satisfied and unfortunately when Cronie et al. (2020, in the text after Lemma 2) let $\delta = d_R$ be the resistance metric, they have misunderstood the meaning of origin independent as used in Anderes et al. (2020, Proposition 2). Moreover, even when using their

meaning of origin independent, the proof of Cronie et al. (2020, Lemma 2) is incorrect, since they claim that for any $u', u'', u_1 \in L$ we have $\delta(u', u_1) = \delta(u'', u_1)$, which is obviously not correct if $\delta = d_R$ (and it seems wrong in general no matter what the metric is).

5.2.2 | Interrupted Cox processes

Let X be an ICP, cf. (27). Then X conditioned on Π is obtained by an independent thinning of a Poisson process Z on S with intensity function $\rho_Z(u) = \rho(u)\,(1 + c(u,u))^{h/2}$, where the selection probabilities are given by Π. In the terminology of Stoyan (1979), X is an interrupted point process.

Lavancier and Møller (2016) studied the ICP (as well as many other examples of interrupted point processes) when $S = \mathbb{R}^k$. For our general state space setting we obtain in a similar way as in Lavancier and Møller (2016) that

$$g_0(t) = \left(\frac{(1+\sigma^2)^2}{(1+\sigma^2)^2 - \sigma^4 r_0(t)^2} \right)^{h/2}, \tag{31}$$

whilst third and higher-order moment results are less simple to express, but it can be proven that X is positively correlated at all orders. As σ increases from 0 to infinity, then q decreases from 1 to 0, whilst $g_0(t)$ increases from 1 to $(1 - r_0(t)^2) - h/2$ if $r_0(t) \neq 0$, which shows a trade-off between the degree of thinning and the degree of clustering. To understand how $g_0(t)$ depends on h it is natural to fix the value of $q \in (0,1)$. Then

$$g_0(t) = \left(1 + \left(1 - q^{h/2}\right)^2 r_0(t)^2 \right)^{-h/2}$$

is a strictly increasing function of h whenever $r_0(t) \neq 0$. Consequently, taking $h = 1$ is natural if we wish to model as much clustering as possible.

For the case $L = S$ and $c(u,v) = \sigma^2 \exp(-d(u,v)/\phi)$ given by an isotropic exponential covariance function, expressions of K_L when $h = 1, 2, \ldots, 5$ are given in (Christensen and Møller, 2020, Appendix A). Although these expressions were given for $d = d_{\mathcal{G}}$, they remain true for a general metric d because g_0 depends only on h, σ^2, and $r_0(t) = \exp(-t/\phi)$, cf. (31) and our comment after (17).

5.2.3 | Permanental point processes

Let X be an PPP, cf. (28). The case $S = \mathbb{R}^k$ is studied in Macchi (1975) and McCullagh and Møller (2006); using the parametrization in the latter paper, X is a PPP with parameters $\alpha = h/2$ and $C(u,v) = \sqrt{\rho(u)\rho(v)}\,c(u,v)/\alpha$. The process can also be defined in the general state space case, see Shirai and Takahashi (2003).

For $n = 1, 2, \ldots$ and $u_1, \ldots, u_n \in S$, define the α-weighted permanent by

$$\mathrm{per}_\alpha[c](u_1, \ldots, u_n) = \sum_\pi \alpha^{\#\pi} c(u_1, u_{\pi_1}) \cdots c(u_n, u_{\pi_n})$$

where the sum is over all permutations $\pi = (\pi_1, \ldots, \pi_n)$ of $(1, \ldots, n)$ and $\#\pi$ is the number of cycles. The usual permanent corresponds to $\alpha = 1$ (Minc, 1978), in which case X is also called a Boson process (Macchi, 1975). We have

$$\rho(u_1, \ldots, u_n) = \rho(u_1) \cdots \rho(u_n)\,\mathrm{per}_\alpha[c](u_1, \ldots, u_n)/\alpha^n$$

from which it can be verified that X is positively correlated at all orders. It also follows that the degree of clustering is a decreasing function of α. In particular,

$$g(u,v) = 1 + c(u,v)^2/\alpha.$$

This reflects the limitation of modelling clustering by a PPP, since c is a correlation function in this model: The pcf is bounded by $1 + 1/\alpha \leq 3$.

Valiant (1979) showed that exact computation of permanents of general matrices is a #P (sharp P) complete problem, so no deterministic polynomial time algorithm is available. For most statistical purposes, approximate computation of permanent ratios is sufficient, and analytic approximations are available for large α. However, as $\alpha \to \infty$, X tends to a Poisson process and the process becomes less and less interesting for the purpose of modelling clustering.

The PPP can be extended to the case where $\alpha \geq 0$ and c is not necessarily a covariance function (in which case we loose the connection to Gaussian and Cox processes), see McCullagh and Møller (2006) and Shirai and Takahashi (2003). Indeed the process also extends to the case where α is a negative integer (a (weighted) determinantal point processes). We return to this and further properties in Section 7).

6 | APPLICATION OF STATISTICAL INFERENCE PROCEDURES

(Some more text will be added here)

Christensen and Møller (2020) analysed several point pattern datasets given by spine locations on different dendrite trees which were identified by linear networks. For each dataset they fitted an inhomogeneous ICP with $h = 1$ and $c(u,v) = \sigma^2 \exp(-d_{\mathcal{G}}(u,v)/\phi)$ given by an isotropic exponential covariance function. Specifically, they used a two step estimation procedure: First, given a parametric model for ρ where the parameter ranges independently of (σ, β), the parameter is estimated by maximizing a first-order composite likelihood function (the Poisson likelihood), second a minimum contrast procedure based on g_0 was used for estimating $\sigma > 0$ and $\phi > 0$, where they plug-in the estimated intensity function. Simulation studies in Christensen and Møller (2020) indicated that 1) it can be difficult to estimate σ^2 and h simultaneously, as an increase in h can be balanced out by an increase in σ^2, 2) with the default bandwidth and kernel from the spatstat-package for estimating g_0, minimum contrast estimation based on g_0 generally performs better than if it is based on K (which is consistent with results in Lavancier and Møller, 2016), and 3) minimum contrast estimation is far more reliable than (second-order) composite likelihood estimation.

7 | CONCLUDING REMARKS

(Some more text will be added here)

References

Anderes, E., Møller, J. and Rasmussen, J. G. (2020) Isotropic covariance functions on graphs and their edges. *Annals of Statistics*, **48**, 2478–2503.

Ang, Q., Baddeley, A. and Nair, G. (2012) Geometrically corrected second order analysis of events on a linear network, with applications to ecology and criminology. *Scandinavian Journal of Statistics*, **39**, 591–617.

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

Baddeley, A., Nair, G., Rakshit, S. and McSwiggan, G. (2017) "Stationary" point processes are uncommon on linear networks. *Stat*, **6**, 68–78.

Baddeley, A., Rubak, E. and Turner, R. (2015) *Spatial Point Patterns: Methodology and Applications with R*. Boca Raton: Chapman and Hall/CRC Press.

Christensen, H. S. and Møller, J. (2020) Modelling spine locations on dendrite trees using inhomogeneous cox point processes. *Spatial Statistics*, **39**, Article 100478.

Coeurjolly, J.-F., Møller, J. and Waagepetersen, R. (2017) Palm distributions for log Gaussian Cox processes. *Scandinavian Journal of Statistics*, **44**, 192–203.

Cox, D. R. (1955) Some statistical methods connected with series of events. *Journal of the Royal Statistical Society: Series B (Methodological)*, **17**, 129–157.

Cronie, O., Moradi, M. and Mateu, J. (2020) Inhomogeneous higher-order summary statistics for point processes on linear networks. *Statistics and Computing*, **30**, 1221–1239.

Cuevas-Pacheco, F. and Møller, J. (2018) Log Gaussian Cox processes on the sphere. *Spatial Statistics*, **26**, 69–82.

Dai, F. and Xu, Y. (2013) *Approximation Theory and Harmonic Analysis on Spheres and Balls*. Springer Monographs in Mathematics. Springer, New York.

Daley, D. J. and Vere-Jones, D. (2003) *An Introduction to the Theory of Point Processes. Volume I: Elementary Theory and Methods*. second edn. Springer, New York.

Diggle, P. (2014) *Statistical Analysis of Spatial and Spatio-Temporal Point Patterns. 3rd ed*. Boca Raton, FL: CRC Press, Taylor and Francis Group.

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

Gneiting, T. (2001) Criteria of Pólya type for radial positive definite functions. *Proceedings of American Mathematical Society*, **129**, 2309–2318.

— (2013) Strictly and non-strictly positive definite functions on spheres. *Bernoulli*, **19**, 1327–1349.

Guttorp, P. and Thorarinsdottir, T. L. (2012) Bayesian inference for non-Markovian point processes. In *Advances and Challenges in Space-time Modelling of Natural Events* (eds. E. Porcu, J. Montero and M. Schlather), vol. 207 of *Lecture Notes in Statistics*, 79–102. Springer: Berlin Heidelberg.

Jorgensen, P. E. T. and Pearse, E. P. J. (2010) A Hilbert space approach to effective resistance metric. *Complex Analysis and Operator Theory*, **4**, 975–1013.

Kigami, J. (2003) Harmonic analysis for resistance forms. *Journal of Functional Analysis*, **204**, 399–444.

Klein, D. J. and M. Randić, M. (1993) Resistance distance. *Journal of Mathematical Chemistry*, **12**, 81–95.

Lavancier, F. and Møller, J. (2016) Modelling aggregation on the large scale and regularity on the small scale in spatial point pattern datasets. *Scandinavian Journal of Statistics*, **43**, 587–609.

Lawrence, T., Baddeley, A., Milne, R. K. and Nair, G. (2016) Point pattern analysis on a region of a sphere. *Stat*, **5**, 144–157.

Macchi, O. (1975) The coincidence approach to stochastic point processes. *Advances in Applied Probability*, **7**, 83–122.

McCullagh, P. and Møller, J. (2006) The permanental process. *Advances in Applied Probability*, **38**, 873–888.

Minc, H. (1978) *Permanents*. Reading, MA: Addison-Wesley.

Møller, J., Syversveen, A. R. and Waagepetersen, R. P. (1998) Log Gaussian Cox processes. *Scandinavian Journal of Statistics*, **25**, 451–482.

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

Møller, J. and Rubak, E. (2016) Functional summary statistics on the sphere with an application to determinantal point processes. *Spatial Statistics*, **18**, 4–23.

Møller, J. and Toftager, H. (2014) Geometric anisotropic spatial point pattern analysis and cox processes. *Scandinavian Journal of Statistics*, **41**, 414–435.

Møller, J. and Waagepetersen, R. (2007) Modern statistics for spatial point processes (with discussion). *Scandinavian Journal of Statistics*, **34**, 643–711.

— (2017) Some recent developments in statistics for spatial point patterns. *Annual Review of Statistics and Its Applications*, **4**, 317–342.

Mrkvička, T., Myllymäki, M., Jilik, M. and Hahn, U. (2020) A one-way ANOVA test for functional data with graphical interpretation. *Kybernetika*, **56**, 432–458.

Myllymäki, M. and Mrkvička, T. (2019) GET: Global envelopes in R. ArXiv preprint arXiv:1911.06583.

Myllymäki, M., Mrkvička, T., Grabarnik, P., Seijo, H. and Hahn, U. (2017) Global envelope tests for spatial processes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **79**, 381–404.

Pitt, L. D. (1971) A Markov property for Gaussian processes with a multidimensional parameter. *Archive for Rational Mechanics and Analysis*, **43**, 367–391.

Rakshit, S., Davies, T., Moradi, M., McSwiggan, G., Nair, G., Mateu, J. and Baddeley, A. (2019) Fast kernel smoothing of point patterns on a large network using two-dimensional convolution. *International Statistical Review*, **87**, 531–556.

Rakshit, S., Nair, G. and Baddeley, A. (2017) Second-order analysis of point patterns on a network using any distance metric. *Spatial Statistics*, **22**, 129–154.

Rue, H. and Held, L. (2005) *Theory of Gaussian Markov Random Fields*. Boca Raton: Chapman and Hall/CRC Press.

Rue, H., Martino, S. and Chopin, N. (2009) Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society: Series B (Methodological)*, **71**, 319–392.

Schoenberg, I. (1938) Metric spaces and completely monotone functions. *Annals of Mathematics*, **39**, 811–841.

— (1942) Positive definite functions on spheres. *Duke Mathematical Journal*, **9**, 96–108.

Shaw, T., Møller, J. and Waagepetersen, R. P. (2021) Globally intensity-reweighted estimators for K- and pair correlation functions. *Australian and New Zealand Journal of Statistics*, **63**, 93–118.

Shirai, T. and Takahashi, Y. (2003) Random point fields associated with certain Fredholm determinants i: fermion, Poisson and boson point processes. *Journal of Functional Analysis*, **205**, 414–463.

Stoyan, D. (1979) Interrupted point processes. *Biometrical Journal*, **21**, 607–610.

Valiant, L. G. (1979) The complexity of computing the permanent. *Theoretical Computer Science*, **8**, 189–201.

Van Lieshout, M. N. M. (2011) A J-function for inhomogeneous point processes. *Statistica Neerlandica*, **65**, 183–201.