

Nonparametric estimation of the intensity function of a spatial point process on a Riemannian manifold

BY S. WARD

Department of Mathematics, Imperial College London, London SW7 2AZ, U.K.
scott.ward12@imperial.ac.uk

5

H. S. BATTEY

Department of Mathematics, Imperial College London, London SW7 2AZ, U.K.
h.battey@imperial.ac.uk

AND E. A. K. COHEN

Department of Mathematics, Imperial College London, London SW7 2AZ, U.K.
e.cohen@imperial.ac.uk

10

SUMMARY

This paper is concerned with nonparametric estimation of the intensity function of a point process on a Riemannian manifold. It provides a first-order asymptotic analysis of the proposed kernel estimator for Poisson processes, supplemented by empirical work to probe the behaviour in finite samples and under other generative regimes. The investigation highlights the scope for finite-sample improvements by allowing the bandwidth to adapt to local curvature.

15

Some key words: Boundary-free manifold; Edge and shape correction; Kernel estimation; Point events.

1. INTRODUCTION

In the analysis of random collections of point events, a fundamental role is played by the intensity function, which determines the first-order properties of a spatial point process and is an essential component of second order analyses. It provides a complete characterization for the smaller class of Poisson processes.

20

Features of spatial point processes, as distinct from those along a time axis, are their inherent multidimensionality and the need to treat all directions equivalently, in contrast to the directionality of one-dimensional time. A further feature, sometimes ignored with little effect due to the scales involved, are the topological features of the space on which the point events occur.

25

In the present paper we are concerned with point processes on the surface of a Riemannian manifold, a situation of high relevance in cellular biology and microbiology, where super-resolution microscopy techniques can record the spatial arrangement of proteins and other molecules of interest on the cellular membranes of cells, bacteria and other microorganisms. At these scales the topology cannot be ignored, necessitating inferential procedures that adapt to local curvature. In this microbiological example, knowledge of the intensity functions of, say, two different molecular processes can guide scientific inference by suggesting possible dependencies between the processes, perhaps to be probed more formally. Alternatively, the intensity

30

35

estimates might be used as outcomes, blocking factors or concomitant variables in an experimental context, concerned with assessing the efficacy of one or more treatments.

Intensity estimation under this framework is unexplored. Recent relevant work is due to Roberson et al. (2014), Lawrence et al. (2016) and Møller & Rubak (2016), who considered functional summary statistics for point processes on the surface of a d -dimensional unit sphere. These summarize the global properties of the point process. Ward et al. (2021) extended the construction of such statistics to convex manifolds using the Mapping Theorem (e.g. Kingman, 1993, p.18) to map the point events on the manifold to the surface of the unit sphere, performed statistical analysis there using the rotational invariance of the sphere, and mapped the conclusions back to the manifold of interest. Whilst the present paper is concerned with more general processes and manifolds, we must similarly assume the implicit equation $g(x_1, \dots, x_d) = 0$ of the manifold is known in analytic form, or can be well approximated as illustrated in Section 7.

The closest related work is that concerned with nonparametric density estimation from independent and identically distributed (i.i.d.) observations constrained to the surface of a manifold. Pelletier (2005) extended some of the theory of kernel density estimation to accommodate i.i.d. observations on a finite volume boundary-free Riemannian manifold, while Kerkycharian et al. (2012) considered so-called needlet density estimation on compact homogeneous manifolds, motivated by applications in astrophysics. As with their Euclidean counterparts, the broad strategies appropriate for kernel density and kernel intensity estimation are rather similar, although the technical differences are considerable, most notably: the point process observations cannot be treated as i.i.d.; the number of event observations are, at least in the present context, treated as random; and the point process is frequently not observed over the entire manifold. The latter situation necessitates procedures that can seamlessly accommodate both boundary-free manifolds and manifolds with boundaries.

2. PRELIMINARIES

Consider a compact d -dimensional Riemannian manifold (\mathcal{M}, g) with Riemannian metric tensor g . Our treatment here is coordinate-free, i.e., avoiding a fixed basis in which to express all calculations. This formulation comes at the expense of greater abstraction but leads to a more compact notation. Most of the details are deferred to the Supplementary Material along with the proofs of the main results.

Let X be a point process over \mathcal{M} , most naturally viewed as a random set formed of elements of \mathcal{M} . To distinguish between points in a realization of X and any point in the space \mathcal{M} , we shall refer to the former as *events* and the latter as *points*. We use x both to specify points in \mathcal{M} and to index events in X , with the context ensuring no ambiguity.

For any Borel measurable subset $B \subseteq \mathcal{M}$, let $N_X(B)$ denote the number of events of X in B and let $d\text{vol}$ denote the d -dimensional Riemannian volume form on \mathcal{M} (see Supplementary Material). The intensity measure is defined as $\mu(B) = E\{N_X(B)\}$ and provided that μ is absolutely continuous with respect to $d\text{vol}$, there exists a function $\rho : \mathcal{M} \rightarrow \mathbb{R}$ called the *intensity function* such that

$$\mu(B) = \int_B \rho(x) d\text{vol}(x).$$

In other words, ρ is the Radon-Nikodym derivative of the intensity measure with respect to the Riemannian volume form. A more precise formalization avoiding ambiguity in the asymptotic

framework of Section 3 is

$$\rho(x) = \lim_{\delta_x \rightarrow 0} \text{vol}(\delta_x)^{-1} E\{N_X(\delta_x)\}, \quad (1)$$

where $\delta_x \subset \mathcal{M}$ is a region centred on x , and the notation $\delta_x \rightarrow 0$ means that the geodesic distance $d_g : \mathcal{M} \times \mathcal{M} \rightarrow \mathbb{R}_+$ between any two elements of δ_x tends to zero. This is a natural adaptation of the Euclidean definition of Cressie (2015) and retains all the usual properties. Under the constraint that X is simple, that is $\text{pr}\{N_X(\delta_x) > 1\} = o\{\text{vol}(\delta_x)\}$, $\rho(x)d\text{vol}(x)$ can be interpreted as the probability of a point occurrence in the infinitesimal volume $d\text{vol}(x)$ at $x \in \mathcal{M}$. 80

A point process X is said to be homogeneous if ρ is constant over \mathcal{M} and otherwise inhomogeneous. By Campbell's theorem (Daley & Vere-Jones, 2010), for any measurable nonnegative function $f : W \subseteq \mathcal{M} \rightarrow \mathbb{R}_+$, 85

$$E\left\{\sum_{x \in X \cap W} f(x)\right\} = \int_W f(x)\rho(x)d\text{vol}(x). \quad (2)$$

Poisson processes can be characterized in the same way on \mathcal{M} as in \mathbb{R}^d . Specifically, X is said to be a Poisson process with intensity function ρ if, for any Borel measurable subset $B \subseteq \mathcal{M}$, $N_X(B)$ is Poisson distributed with mean $\mu(B)$ and, for any disjoint Borel measurable subsets $A, B \subseteq \mathcal{M}$, $N_X(A)$ and $N_X(B)$ are independent random variables. This affords considerable simplification. In particular, for any measurable non-negative function $f : W \subseteq \mathcal{M} \rightarrow \mathbb{R}_+$, 90

$$\text{Var}\left\{\sum_{x \in X \cap W} f(x)\right\} = \int_W f^2(x)\rho(x)d\text{vol}(x). \quad (3)$$

3. INTENSITY ESTIMATION ON \mathcal{M}

Estimation of ρ is treated nonparametrically. As in simpler contexts (e.g. Bartlett, 1963; Cox, 1965, for events along a time axis) smoothing is required to achieve acceptable estimation variance. This entails some form of weighted averaging of nearby points, ideally with tapered weights for decreasing proximity. Intuitively, in regions of high curvature, neighbouring points appear closer in the Euclidean metric than the arc length of the shortest curve section between them, constrained to the surface of \mathcal{M} , namely the geodesic distance $d_g : \mathcal{M} \times \mathcal{M} \rightarrow \mathbb{R}_+$. This renders the standard Euclidean theory of kernel intensity estimation unusable. We pursue the natural approach of replacing the Euclidean metric in the kernel function by the geodesic distance, so that the kernel intensity estimator automatically adapts to local curvature. 95

A further complication in this setting is that a kernel function, typically non-compactly supported, centred at a particular point may not integrate to one over the manifold. This could be because the manifold has a boundary, or may only be observed over a convex compact subset of \mathcal{M} , a situation that is rather common in practice. A related problem arises when the manifold is of finite volume and boundary-free. Although this latter issue can be circumvented in certain special cases, for instance by using Fisher's (1953) density function as a kernel on the sphere or adopting finitely supported kernels (Pelletier, 2005), for more general manifolds and kernels a shape correction is needed, in effect to avoid double counting of points in the weighted average. Conveniently, the boundary correction required in the former situation is operationally the same as shape correction for finite-volume boundary-free manifolds. All cases can be encapsulated by defining a convex compact subset W of \mathcal{M} over which the point process is observed. The corrections introduced in the forthcoming discussion are then either edge or shape corrections, the latter corresponding to $W = \mathcal{M}$ with \mathcal{M} a finite-volume boundary-free manifold. 100
105
110

The intensity function estimator to be studied in the present paper is

$$\hat{\rho}_h(x) = \sum_{y \in X \cap W} \frac{c_h(x, y)^{-1}}{h^d} k \left\{ \frac{d_g(x, y)}{h} \right\}, \quad (4)$$

115 where W is as described above, $c_h(x, y)$ is the edge or shape correction and, for Euclidean norm $\|\cdot\|$, k is such that $k \circ \|\cdot\| : \mathbb{R}^d \rightarrow \mathbb{R}_+$ is a symmetric probability density function, specified for concreteness as Gaussian:

$$k\{d_g(\cdot, y)\} = (2\pi)^{-d/2} \exp\{-d_g^2(\cdot, y)/2\}.$$

In direct analogy to the corresponding corrections in \mathbb{R}^d (Diggle, 1985; Berman & Diggle, 1989; van Lieshout, 2012), $c_h(x, y)$ is defined either globally or locally as

$$120 \quad c_h^{\text{glo}}(x, y) = c_h(x) = \frac{1}{h^d} \int_W k \left\{ \frac{d_g(x, z)}{h} \right\} d\text{vol}(z), \quad (5)$$

$$c_h^{\text{loc}}(x, y) = c_h(y) = \frac{1}{h^d} \int_W k \left\{ \frac{d_g(z, y)}{h} \right\} d\text{vol}(z). \quad (6)$$

Specifically, the global correction depends only on the point at which the intensity is estimated, while the local correction adjusts for each event. The resulting estimator (4) is generally biased in finite samples regardless of which correction is used but, as shown in Proposition 1, the global version $\hat{\rho}_h^{\text{glo}}$ is unbiased for homogeneous point processes. The local version $\hat{\rho}_h^{\text{loc}}$ enjoys mass preservation for homogeneous and inhomogeneous processes, specifically,

$$\int_W \hat{\rho}_h^{\text{loc}}(x) d\text{vol}(x) = N_X(W), \quad (7)$$

as was demonstrated in the Euclidean case by van Lieshout (2012).

The estimators $\hat{\rho}_h^{\text{glo}}(x)$ and $\hat{\rho}_h^{\text{loc}}(x)$ are best justified by consideration of their first and second moment properties, stated as a series of Propositions of varying degrees of technical intricacy, and culminating in Proposition 3.

130 **PROPOSITION 1.** *Let (\mathcal{M}, g) be a Riemannian manifold and let X be a homogeneous spatial point process over \mathcal{M} with intensity function $\rho(x) = \rho$ for all $x \in \mathcal{M}$. Then for any h , $\hat{\rho}_h^{\text{glo}}$ is unbiased for ρ while $E\{\hat{\rho}_h^{\text{loc}}\} = \rho\eta$ with*

$$\eta = \int_W \frac{c_h^{\text{loc}}(\cdot, y)^{-1}}{h^d} k \left\{ \frac{d_g(x, y)}{h} \right\} d\text{vol}(y).$$

Proof. This is a special case of the more general result

$$135 \quad E\{\hat{\rho}_h^\bullet(x)\} = \int_W \frac{c_h^\bullet(x, y)^{-1}}{h^d} k \left\{ \frac{d_g(x, y)}{h} \right\} \rho(y) d\text{vol}(y), \quad \bullet \in \{\text{glo}, \text{loc}\},$$

which follows by Campbell's theorem. The result is immediate on noting the constancy of the intensity function. \square

Although $\hat{\rho}_h^{\text{loc}}$ has multiplicative bias η for homogenous processes (and is expected to exhibit pointwise bias for inhomogeneous processes too) the functional $\int_W \hat{\rho}_h^{\text{loc}}(x) d\text{vol}(x)$ is always unbiased for $\mu(W)$ by taking expectations in (7).

PROPOSITION 2. Let (\mathcal{M}, g) be a Riemannian manifold and let X be a Poisson process on \mathcal{M} . Then

$$\begin{aligned} \text{Var}\{\hat{\rho}_h^{\text{glo}}(x)\} &= c_h^{\text{glo}}(x, \cdot)^{-2} \int_W \left[\frac{1}{h^d} k \left\{ \frac{d_g(x, y)}{h} \right\} \right]^2 \rho(y) \, d\text{vol}(y), \\ \text{Var}\{\hat{\rho}_h^{\text{loc}}(x)\} &= \int_W c_h^{\text{loc}}(\cdot, y)^{-2} \left[\frac{1}{h^d} k \left\{ \frac{d_g(x, y)}{h} \right\} \right]^2 \rho(y) \, d\text{vol}(y). \end{aligned}$$

Proof. The result is by direct calculation using (3). □

For homogeneous Poisson processes, it follows from Proposition 2 that the variance is not constant over \mathcal{M} even though the intensity function is. This conclusion is equivalent to that of Rakshit et al. (2019) in the context of homogeneous point processes observed over linear networks.

Specification of the bandwidth h relies on a notional asymptotic regime in which the expected number of events $\mu(W)$ diverges. As in simpler contexts, the bias and variance are antagonistic as a function of h , and a suitable compromise between the two must be determined. In assessing the appropriate scaling of h with the expected number $\mu(W)$ of events, there are some subtleties that distinguish the present setting from the i.i.d. Euclidean case. In a Euclidean setting, one way to achieve $\mu(W) \rightarrow \infty$ is to consider an expanding W . This on its own is unsatisfactory, as the concentration of events around an arbitrary $x \in W$ could remain diffuse, as noted by Cucala (2008). The expanding W framework is also physically implausible in the context of boundary-free finite-volume manifolds where $W = \mathcal{M}$.

To ensure the target of inference is stable under the notional limiting operation $\mu(W) \rightarrow \infty$, the asymptotic properties of a suitably standardized version of (4) are considered, analogously to Cucala (2008). The standardized object of inference is $\rho_1(x) = \rho(x)/\mu(W)$, the relative concentration of events at each point of W ensuring that ρ_1 integrates to one over W . The corresponding estimators are

$$\hat{\rho}_{h,1}^\bullet(x) = \frac{1\{N_X(W) \neq 0\}}{N_X(W)} \sum_{y \in X \cap W} \frac{c_h^\bullet(x, y)^{-1}}{h^d} k \left\{ \frac{-d_g(x, y)}{h} \right\}, \quad \bullet \in \{\text{glo}, \text{loc}\}, \quad (8)$$

where $1(A)$ denotes the indicator function of the event A , and the relationship to the estimator in (4) is $\hat{\rho}_h^\bullet(x) = N_X(W) \hat{\rho}_{h,1}^\bullet(x)$. For Poisson processes the following proposition gives the pointwise asymptotic properties of $\hat{\rho}_{h,1}^\bullet$ for $\bullet \in \{\text{glo}, \text{loc}\}$.

PROPOSITION 3. Let (\mathcal{M}, g) be a Riemannian manifold. Suppose X is a Poisson process parameterized by $\rho = \{\rho(x) : x \in \mathcal{M}\}$ and observed over the bounded window $W \subseteq \mathcal{M}$. Provided that ρ_1 is smooth, for any $x \in W \subseteq \mathcal{M}$ and $\bullet \in \{\text{glo}, \text{loc}\}$,

$$\begin{aligned} E\{\hat{\rho}_{h,1}^\bullet(x)\} &\rightarrow \rho_1(x), \\ \text{Var}\{\hat{\rho}_{h,1}^\bullet(x)\} &\rightarrow 0, \end{aligned}$$

as $\mu(W) \rightarrow \infty$ provided that $h \rightarrow 0$ and $\mu(W)^{-1} = o(h^d)$.

This result supplies a degree of reassurance over the behaviour of the proposed estimators, as the conclusion coincides with that obtained in Euclidean space.

From a technical point of view there are some limitations of this analysis. Firstly, Proposition 3 is proved only for Poisson processes. It is supplemented by empirical work in Section 6, which probes the behaviour in finite samples and for other generative processes. Secondly, the conclusions are first-order asymptotic in nature, and not optimized to exploit the interaction between the process and the manifold. Since the volume of a ball of radius r at two points x and y on a

180 general manifold is not necessarily equal for $x \neq y$, the expected number of events in such a ball is, in general, not constant over the manifold. Intuitively then, an optimal estimator would have a bandwidth that adapted to the local curvature, thereby producing a better separation between clustering induced by the process and that induced by the geometry. Similar reasoning would lead one to allow asymmetric localization via elongated “balls” and so on.

185 4. PRACTICAL GUIDE TO BANDWIDTH SELECTION

While Proposition 3 specifies the properties of h under a notional asymptotic regime, and thereby provides some theoretical reassurance over the proposed intensity estimator, the practical problem of choosing the bandwidth for a given sample size is always present, as in almost all areas of nonparametric estimation.

190 One approach to selecting h is through a critical inspection of intensity plots in order to balance local and global features in the data (Møller & Waagepetersen, 2004). Other approaches involve optimization criteria. Baddeley et al. (2015, p. 176) suggest selecting the h that maximises the cross-validated Poisson log likelihood, which in the present setting is

$$\ell_{\text{cv}}(h|X) = \sum_{x \in X} \log \{ \hat{\rho}_h^{-x}(x) \} - \int_{\mathcal{M}} \hat{\rho}_h(x) d\text{vol}(x), \quad (9)$$

195 where $\hat{\rho}_h^{-z}(x) = h^{-d} \sum_{y \in X \setminus \{z\}} k\{-d_g(x, y)/h\} c_h^{-1}(x, y)$ is an estimate of ρ constructed as in (4) but without the observation $z \in X$. Application of Campbell’s Theorem shows ℓ_{cv} is unbiased for the log likelihood function

$$\ell(\rho; X) = \sum_{x \in X} \log \{ \rho(x) \} - \int_{\mathcal{M}} \rho(x) d\text{vol}(x).$$

A nonparametric bandwidth selection procedure that can be readily extended to the Riemannian setting is given in Cronie & Van Lieshout (2018). On assuming that the intensity function is positive everywhere on \mathcal{M} and applying Campbell’s formula (2) to ρ^{-1} ,

$$E \left\{ \sum_{x \in X} \rho^{-1}(x) \right\} = \text{Vol}(W).$$

Replacement of ρ by its estimate $\hat{\rho}_h$ points to a choice of h that minimizes

$$F(h) = \{T(\hat{\rho}_h) - \text{Vol}(W)\}^2, \quad (10)$$

200 where $T(\hat{\rho}_h) = \sum_{x \in X} \hat{\rho}_h^{-1}(x)$. In addition to being relatively free of modelling assumptions, (10) is less burdensome to compute than (9). The existence of a minima of F can be shown by consideration of its continuity and limiting properties. Proposition 4 extends Theorem 1 of Cronie & Van Lieshout (2018) to \mathcal{M} .

PROPOSITION 4. *Let (\mathcal{M}, g) be a Riemannian manifold and let X be a point process observed through a bounded window $W \subseteq \mathcal{M}$. After disregarding the trivial case $X \cap W = \emptyset$, global and local corrections (5) or (6) both yield T continuous in $h \in (0, \infty)$. This conclusion also holds when no correction is used, i.e. $c_h^\bullet(x, y) = 1$. In all cases, $\lim_{h \rightarrow 0} T(\hat{\rho}_h) = 0$. For correction given by (5) and (6) $\lim_{h \rightarrow \infty} T(\hat{\rho}_h) = \text{Vol}(W)$ and if no correction is used $\lim_{h \rightarrow \infty} T(\hat{\rho}_h) = \infty$.*

210 The intermediate value theorem dictates that when $c_h(x, y) = 1$ there exists a minima for F , whilst if a correction is used a minimum occurs when $h \rightarrow \infty$. This is consistent with the Euclidean approach considered by Cronie & Van Lieshout (2018). The recommendation of the

present paper is also to optimize F with no correction, including it instead after h has been selected.

5. NUMERICAL CONFIRMATION IN A TEST CASE

Numerical validity of the proposed procedure is checked empirically using an example in which standard methods are approximately valid in the limit as a key parameter becomes small, but more generally handled by the approach developed in Section 3.

We consider a Poisson process on the unit square with intensity function

$$\rho(x_1, x_2) = \frac{N}{2\pi\sigma^2 K} \exp\left\{-\frac{(x_1 - \frac{1}{2})^2 + (x_2 - \frac{1}{2})^2}{2\sigma^2}\right\},$$

where $N > 0$ and K is chosen to ensure the expected number of points in $[0, 1]^2$ is N . With the unit square considered as the unit subset of the plane $x_3 = 0$ in \mathbb{R}^3 , a Poisson process on a bounded Euclidean manifold, $\mathcal{M} = W$ say, is obtained by rotating the plane $x_3 = 0$ through an angle of θ about the x_2 -axis, giving the intensity function

$$\rho_W(x_1, x_2, x_3) = \begin{cases} \rho\{(x_1^2 + x_3^2)^{1/2}, x_2\} & x_3 = x_1 \tan(\theta); \\ 0 & \text{otherwise.} \end{cases}$$

The first approach, as described in Sections 3 and 4, involves direct application of the recommended procedure on W to estimate ρ_W using a local correction. The second, for comparison, follows a cross-validation approach previously considered in Baddeley et al. (2015), whereby a standard bivariate Euclidean kernel intensity estimator is first applied to the orthogonal projection on the plane $x_3 = 0$, with the fitted intensity $\hat{\rho}_{\text{proj}}$ then mapped back onto W as

$$\hat{\rho}_W(x_1, x_2, x_3) = \begin{cases} \hat{\rho}_{\text{proj}}(x_1, x_2)\{1 + \tan^2(\theta)\}^{-1/2} & x_3 = x_1 \tan(\theta); \\ 0 & \text{otherwise.} \end{cases}$$

The two situations are depicted in Fig. 1(a).

The sample mean integrated squared error (MISE) of each approach is computed using 10,000 simulated replicates of the point pattern with parameters $\sigma^2 = 0.01$ and $N = 500$ for a range of values of θ . Fig. 1(b) shows that the MISE remains constant for increasing θ when estimation is performed directly on the manifold, whereas the projection approach differs considerably for large θ , but coincides, as expected, for small θ .

6. SIMULATIONS

Point patterns are simulated on the surface of three ellipsoids of increasing eccentricity: manifolds \mathcal{E}_1 , \mathcal{E}_2 and \mathcal{E}_3 , respectively. A common local chart used to describe an ellipsoid \mathcal{E} is $x \equiv (x_1, x_2, x_3) = \{a \sin(\theta) \cos(\phi), b \sin(\theta) \sin(\phi), c \cos(\theta)\}$ where $\theta \in [0, \pi]$ and $\phi \in [0, 2\pi)$. Manifold \mathcal{E}_1 is a sphere of radius $a = b = c = (4\pi)^{-1/2}$, \mathcal{E}_2 has $a = b = 0.8(4\pi)^{-1/2}$, and \mathcal{E}_3 has $a = b = 0.6(4\pi)^{-1/2}$. To enable comparison, the value of c for \mathcal{E}_2 and \mathcal{E}_3 is set to ensure that they each have unit Riemannian volume measure (surface area).

The intensity function is estimated using point patterns sampled from three Poisson process models. Details and results for log Gaussian Cox processes and Strauss processes are presented in Supplementary Materials Section 3, alongside a detailed explanation of how the processes were simulated in Supplementary Materials Section 4. The three Poisson process models considered are: (PP1) homogeneous Poisson process, i.e. with intensity function $\rho_1(x) = \rho_1$; (PP2) inhomogeneous Poisson process with log-linear intensity function $\rho_2(x) =$

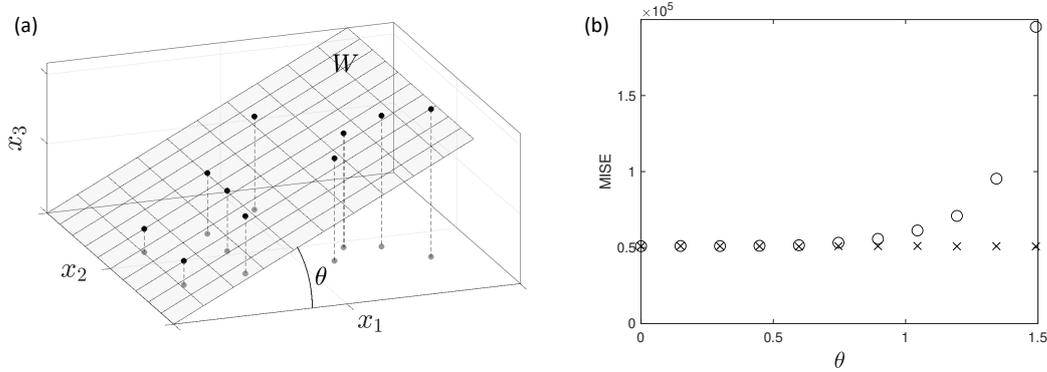


Fig. 1. (a) Example point pattern on W (black) with its orthogonal projection on the plane $x_3 = 0$ (grey). (b) Sample mean integrated squared error (MISE) of the intensity estimate across 10,000 simulations for the projected point pattern (o) and working directly on the manifold (x)

$\exp(3 + \alpha_2 x_1)$; (PP3) inhomogeneous Poisson process with log-modulation intensity function $\rho_3(x) = \exp\{2 + \alpha_3 \cos(8x_2)\}$. Parameters ρ_1 , α_2 and α_3 each take three values to give a low, medium and high number of expected events.

240 The results from this 3^3 factorial experiment are presented in Table 1, where the last two columns display the integrated squared error of the estimate, $\|\hat{\rho} - \rho\|^2$, averaged over Monte Carlo replications and standardized by the square of the expected number of events $\|\rho\|^2$ to make the rows comparable. The norm is the $L_2(\mathcal{M})$ norm, i.e. with $\mathcal{M} \in \{\mathcal{E}_1, \mathcal{E}_2, \mathcal{E}_3\}$,

$$\begin{aligned}
 \|\hat{\rho} - \rho\|^2 &= \int_{\mathcal{M}} \{\hat{\rho}(x) - \rho(x)\}^2 d\text{vol}(x) \\
 &= \int_0^\pi \int_0^{2\pi} [\hat{\rho}\{\psi^{-1}(\theta, \phi)\} - \rho\{\psi^{-1}(\theta, \phi)\}]^2 \{\det(g_{ij})\}^{1/2} d\theta d\phi, \quad (11)
 \end{aligned}$$

where ψ is the local chart for \mathcal{M} and where (g_{ij}) is the matrix representation of the metric under the corresponding local coordinate system. For the chosen chart it can be shown that

$$\det(g_{ij}) = \sin^2(\theta) a^2 b^2 \{1 - (1 - c^2/a^2) \sin^2(\theta) \cos^2(\phi) - (1 - c^2/b^2) \sin^2(\theta) \sin^2(\phi)\}.$$

The bandwidth is selected using the two methods outlined in Section 4, referred to here as *cross validation* (CV) (Baddeley et al., 2015) and *nonparametric* (NP) (Cronie & Van Lieshout, 2018).

250 Intensity function estimates are then computed using the local correction. The integral in (11) is computed using a numerical approximation.

In the Poisson setting outlined here, the CV method for bandwidth selection outperforms the NP method, while inspection of the results for log Gaussian Cox and Strauss processes shows the opposite is true. This is unsurprising since the CV method is based on a Poisson likelihood.

255 The simulation results are consistent with the Euclidean analysis considered in Cronie & Van Lieshout (2018).

7. APPLICATION TO THE BEILSCHMIEDIA PENDULA DATASET

In applying the proposed estimator to the Beilschmiedia Pendula data set (Hubbell, 1983; Condit et al., 1996; Condit, 1998), a number of important practical considerations are isolated.

Table 1. Performance of kernel intensity estimators

Manifold	Poisson model	Process parameters	Expected number of events	$\{\hat{E}(\ \hat{\rho} - \rho\ ^2 / \ \rho\ ^2)\}^{1/2}$	
				CV	NP
\mathcal{E}_1	PP1	$\rho_1 = 50$	50.00	0.285	0.307
\mathcal{E}_1	PP1	$\rho_1 = 150$	150.0	0.182	0.220
\mathcal{E}_1	PP1	$\rho_1 = 300$	300.0	0.128	0.185
\mathcal{E}_1	PP2	$\alpha_2 = 10$	59.57	0.306	0.598
\mathcal{E}_1	PP2	$\alpha_2 = 18$	317.2	0.179	0.765
\mathcal{E}_1	PP2	$\alpha_2 = 22$	802.2	0.141	0.806
\mathcal{E}_1	PP3	$\alpha_3 = 3$	49.98	0.462	0.624
\mathcal{E}_1	PP3	$\alpha_3 = 4$	116.1	0.378	0.671
\mathcal{E}_1	PP3	$\alpha_3 = 5$	280.2	0.305	0.706
\mathcal{E}_2	PP1	$\rho_1 = 50$	50.00	0.285	0.303
\mathcal{E}_2	PP1	$\rho_1 = 150$	150.0	0.178	0.224
\mathcal{E}_2	PP1	$\rho_1 = 300$	300.0	0.137	0.186
\mathcal{E}_2	PP2	$\alpha_2 = 10$	43.60	0.343	0.538
\mathcal{E}_2	PP2	$\alpha_2 = 18$	153.8	0.236	0.717
\mathcal{E}_2	PP2	$\alpha_2 = 22$	313.7	0.183	0.763
\mathcal{E}_2	PP3	$\alpha_3 = 3$	58.60	0.449	0.562
\mathcal{E}_2	PP3	$\alpha_3 = 4$	135.9	0.380	0.617
\mathcal{E}_2	PP3	$\alpha_3 = 5$	326.7	0.303	0.659
\mathcal{E}_3	PP1	$\rho_1 = 50$	50.00	0.292	0.297
\mathcal{E}_3	PP1	$\rho_1 = 150$	150.0	0.197	0.229
\mathcal{E}_3	PP1	$\rho_1 = 300$	300.0	0.147	0.187
\mathcal{E}_3	PP2	$\alpha_2 = 10$	32.34	0.432	0.511
\mathcal{E}_3	PP2	$\alpha_2 = 18$	75.05	0.313	0.662
\mathcal{E}_3	PP2	$\alpha_2 = 22$	123.1	0.279	0.718
\mathcal{E}_3	PP3	$\alpha_3 = 3$	74.88	0.447	0.478
\mathcal{E}_3	PP3	$\alpha_3 = 4$	175.7	0.368	0.514
\mathcal{E}_3	PP3	$\alpha_3 = 5$	423.1	0.290	0.562

Square root of the mean integrated squared error from a 3^3 factorial experiment. The mean is taken over 100 Monte Carlo replicates.

These data provide locations of 3605 trees in a 1000m by 500m rectangular sampling region of a tropical rainforest. Each event is characterized by its longitude, latitude and elevation and is assumed to lie on a two-dimensional Riemannian manifold within \mathbb{R}^3 , inheriting the canonical metric tensor by the embedding. As with most practical settings, there is no analytical formula for the manifold. Instead, it was approximated using a 201×101 regular grid of longitude and latitude coordinates each with an elevation. Let U be the union of the recorded events and the manifold grid points. The manifold was approximated as a triangular mesh $\mathcal{M} = (U, T)$, where $T = \{F_m; m = 1, \dots, M\}$ is the set of triangular faces of the mesh with each element represented by the three elements of U that form its vertices. This was constructed by implementing a 2D Delaunay triangular mesh on the longitude and latitude, with each vertex then raised by its respective elevation. The full mesh and an illustrative subsection are shown in Fig. 3(a) and Fig. 3(b), respectively. Including the events as vertices of the triangular mesh aids the computation of event-to-point geodesic distances, here conducted with the Fast Marching Algorithm (Peyre, 2021).

260

265

270

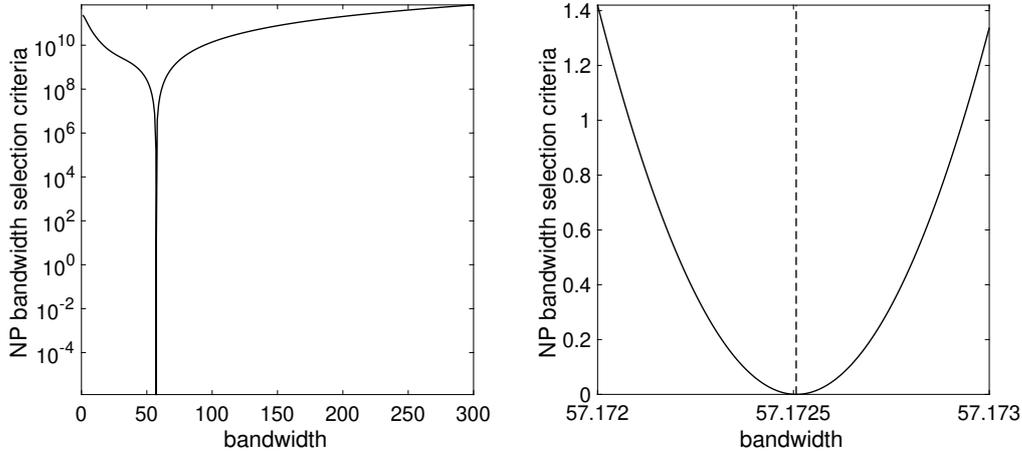


Fig. 2. (a) Nonparametric (NP) bandwidth selection criteria $F(h)$ for the Beilschmiedia Pendula Dataset, plotted with a log scale y -axis. (b) $F(h)$ in the locality of its minimum, plotted with a linear scale y -axis.

The kernel intensity estimator of (4) was constructed using the local correction with (6) becoming

$$c_h(y) = \frac{1}{h^d} \sum_{m=1}^M \int_{F_m} k \left\{ \frac{d_g(z, y)}{h} \right\} d\lambda_m(z)$$

on replacement of $d\text{vol}(\cdot)$ by $d\lambda_m(\cdot)$ for the surface area element over the m th face of the mesh. This is approximated by

$$\frac{1}{h^d} \sum_{m=1}^M k \left\{ \frac{d_g(z_m, y)}{h} \right\} \lambda_m(F_m),$$

where z_m is a representative point of F_m , here computed as the arithmetic average of its three vertices. At this junction, a second triangular mesh was constructed in an identical manner to the first but including $\{z_1, \dots, z_M\}$ as additional vertices such that all required geodesic distances could be computed with the Fast Marching Algorithm.

To avoid modelling the data generating process, only the nonparametric (NP) approach to bandwidth selection was used. If the CV method was to be applied, the integral in (9) would instead be approximated as

$$\int_{\mathcal{M}} \hat{\rho}_h(x) d\text{vol}(x) \approx \sum_{m=1}^M \hat{\rho}_h(z_m) \lambda_m(F_m).$$

As recommended, the bandwidth was selected without correction, which was only applied subsequently in the construction of the kernel intensity estimate. The selection criterion was evaluated at bandwidths $h \in \{1, 2, \dots, 300\}$ in the units of metres. Additional refinement around the minimizing value of h gave a final bandwidth choice of 57.17 m (2 d.p.). Fig. 2 shows a well behaved convex function with a pronounced minimum. The resulting NP intensity estimate is shown in Fig. 3(c). Fig. 3(d) displays the relative difference between this and a simple alternative. The latter, written $\hat{\rho}_{\text{flat}}$ constructs the intensity estimate on the plane and projects it onto the landscape using local gradients, as in (Baddeley et al., 2015, p. 176). The elevation scale has

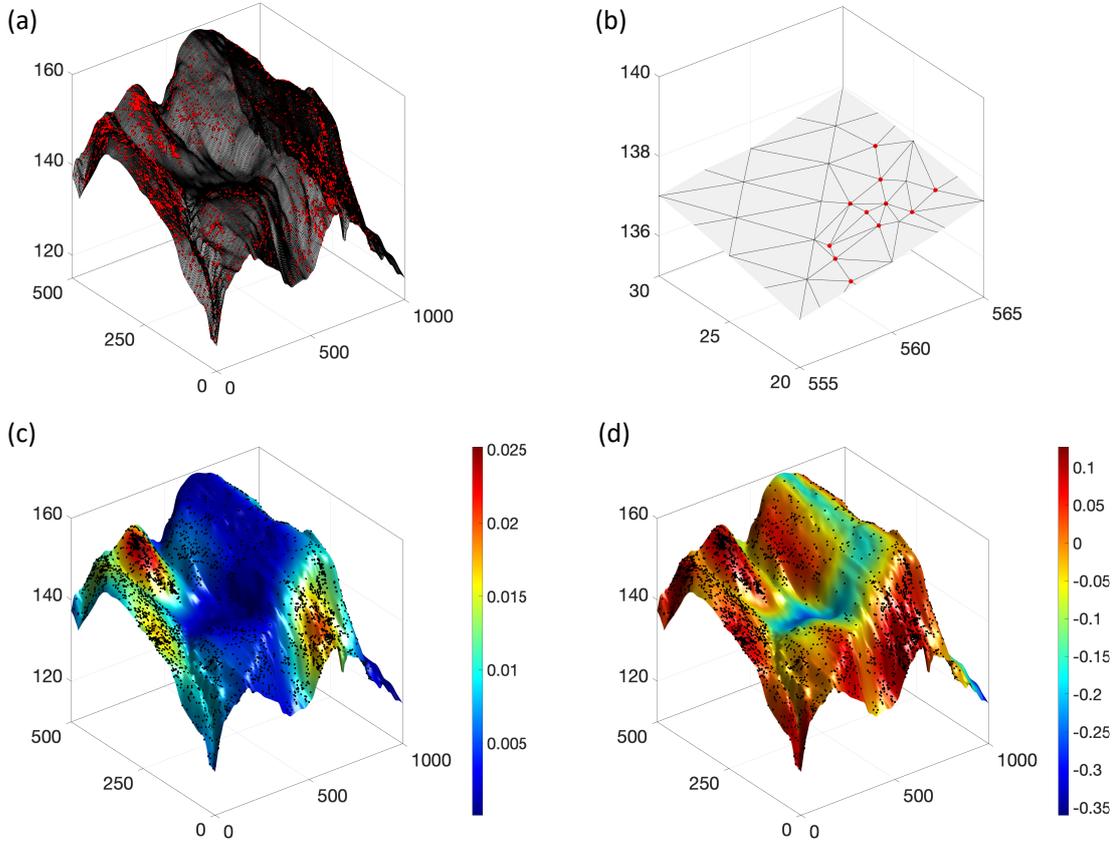


Fig. 3. (a) *Beilschmiedia Pendula* events (red dots) on the triangular mesh representation of the geographical surface. (b) Small portion of (a), depicting the events (red dots) and constructed triangular mesh. (c) NP intensity estimate in units m^{-1} depicted using the presented colormap, shown with the events (black dots). (d) Relative difference between the estimated intensity directly on the manifold and estimated intensity assuming a flat surface, $\{\hat{\rho}_{\text{manifold}}(x)/\hat{\rho}_{\text{flat}}(x)\} - 1$.

been magnified to aid visualization. For a manifold without a boundary, such as a sphere or an ellipsoid, it is unclear how $\hat{\rho}_{\text{flat}}$ could be implemented.

290

8. DISCUSSION AND OPEN PROBLEMS

The constructions presented in the present work have first-order asymptotic guarantees for the estimation of intensity functions of Poisson processes observed over general Riemannian manifolds, with or without boundaries. Their properties under other generative point processes have been assessed by simulation. As discussed in Section 3, intuitive reasoning suggests that the proposed estimator is not optimal in finite samples. A finite-sample theoretical analysis seems challenging and may involve extension of the classical probability inequalities.

295

ACKNOWLEDGEMENTS

We are grateful to Niall Adams for joint supervision of S.W.'s Ph.D. at Imperial College
 300 London, and Karthik Bharath for a helpful discussion. Financial support was provided by the
 Wellcome Trust (to S.W.) and the UK Engineering and Physical Sciences Research Council (to
 H.S.B.).

REFERENCES

- BADDELEY, A., RUBAK, E. & TURNER, R. (2015). *Spatial Point Patterns: Methodology and Applications with R*.
 305 CRC press.
- BARTLETT, M. S. (1963). The spectral analysis of point processes. *J. Roy. Statist. Soc. Ser. B* **25**, 264–296.
- BERMAN, M. & DIGGLE, P. (1989). Estimating weighted integrals of the second-order intensity of a spatial point
 process. *J. Roy. Statist. Soc. Ser. B* **51**, 81–92.
- CONDIT, R. (1998). *Tropical Forest Census Plots*. Springer Berlin, Heidelberg.
- 310 CONDIT, R., HUBBELL, S. P. & FOSTER, R. B. (1996). Changes in tree species abundance in a neotropical forest:
 impact of climate change. *J. Trop. Ecol.*, 231–256.
- COX, D. R. (1965). On the estimation of the intensity function of a stationary point process. *J. Roy. Statist. Soc. Ser.*
B **27**, 332–337.
- CRESSIE, N. (2015). *Statistics for Spatial Data*. John Wiley & Sons.
- 315 CRONIE, O. & VAN LIESHOUT, M. N. (2018). A non-model-based approach to bandwidth selection for kernel
 estimators of spatial intensity functions. *Biometrika* **105**, 455–462.
- CUCALA, L. (2008). Intensity estimation for spatial point processes observed with noise. *Scand. J. Stat.* **35**, 322–334.
- DALEY, D. & VERE-JONES, D. (2010). *An Introduction to the Theory of Point Processes*, vol. I. Springer New York,
 2nd ed.
- 320 DIGGLE, P. (1985). A kernel method for smoothing point process data. *J. Roy. Statist. Soc. Ser. C* **34**, 138–147.
- HUBBELL, S. P. (1983). Diversity of canopy trees in a neotropical forest and implications for conservation. In *Tropi-*
cical Rain Forest: Ecology and Management, T. Whitmore, A. Chadwick & A. Sutton, eds. The British Ecological
 Society, pp. 25–41.
- KERKYACHARIAN, G., NICKL, R. & PICARD, D. (2012). Concentration inequalities and confidence bands for
 325 needle density estimators on compact homogeneous manifolds. *Probab. Theory Relat. Fields* **153**, 363–404.
- KINGMAN, J. F. C. (1993). *Poisson Processes*. Oxford University Press.
- LAWRENCE, T., BADDELEY, A., MILNE, R. K. & NAIR, G. (2016). Point pattern analysis on a region of a sphere.
Stat **5**, 144–157.
- MØLLER, J. & RUBAK, E. (2016). Functional summary statistics for point processes on the sphere with an application
 330 to determinantal point processes. *Spat. Stat.* **18**, 4–23.
- MØLLER, J. & WAAGEPETERSEN, R. P. (2004). *Statistical Inference and Simulation for Spatial Point Processes*.
 Florida: CRC Press.
- PELLETIER, B. (2005). Kernel density estimation on Riemannian manifolds. *Statist. Prob. Lett.* **73**, 297–304.
- PEYRE, G. (2021). Toolbox Fast Marching. [https://www.mathworks.com/matlabcentral/](https://www.mathworks.com/matlabcentral/fileexchange/6110-toolbox-fast-marching)
 335 [fileexchange/6110-toolbox-fast-marching](https://www.mathworks.com/matlabcentral/fileexchange/6110-toolbox-fast-marching), [Accessed: 13 September 2022].
- RAKSHIT, S., DAVIES, T., MORADI, M. M., MCSWIGGAN, G., NAIR, G., MATEU, J. & BADDELEY, A. (2019).
 Fast kernel smoothing of point patterns on a large network using two-dimensional convolution. *Int. Stat. Rev.* **87**,
 531–556.
- ROBESON, S. M., LI, A. & HUANG, C. (2014). Point-pattern analysis on the sphere. *Spat. Stat.* **10**, 76–86.
- 340 VAN LIESHOUT, M.-C. N. (2012). On estimation of the intensity function of a point process. *Methodol. Comput.*
Appl. Probab. **14**, 567–578.
- WARD, S., COHEN, E. A. K. & ADAMS, N. (2021). Testing for complete spatial randomness on three dimensional
 bounded convex shapes. *Spat. Stat.* **41**.

[Received on 2 January 2017. Editorial decision on 1 April 2017]

Supplementary material for “Nonparametric estimation of the intensity function of a spatial point process on a Riemannian manifold”

BY S. WARD

Department of Mathematics, Imperial College London, London SW7 2AZ, U.K.
scott.ward12@imperial.ac.uk

5

H. S. BATTEY

Department of Mathematics, Imperial College London, London SW7 2AZ, U.K.
h.battey@imperial.ac.uk

E. A. K. COHEN

Department of Mathematics, Imperial College London, London SW7 2AZ, U.K.
e.cohen@imperial.ac.uk

10

SUMMARY

The Supplementary Material provides differential geometric background, proofs of the main results in the paper, and simulations for log Gaussian Cox processes and Strauss processes.

15

1. BACKGROUND ON COORDINATE-FREE DIFFERENTIAL GEOMETRY

1.1. Smooth manifolds

A d -dimensional manifold is a topological space that can be locally approximated at any point $x \in \mathcal{M}$ by a subset of \mathbb{R}^d . That is, for any $x \in \mathcal{M}$ there is an open set $U_x \subset \mathcal{M}$ such that $x \in U_x$, an open set $V_x \subset \mathbb{R}^d$ and a map $\psi_x : U_x \rightarrow V_x$ that is required to be continuous and bijective with a continuous inverse (a *homeomorphism*). The pair (U_x, ψ_x) is called a *local chart*, or simply a *chart*. The purpose of ψ_x is to attach coordinates to points in U_x , and the purpose of charts is to allow entire manifolds to be described with injective parameterizations. For instance, a circle described by the parameterized curve $\gamma : \mathbb{R} \rightarrow \mathbb{R}^2$ would otherwise be problematic, because the circle $\gamma(t) = (\cos t, \sin t)$ is not homeomorphic to \mathbb{R} . In order to obtain an injective parameterization γ should be restricted, and the whole circle can be recovered by considering at least two intervals, $t \in U_1 = (0, 2\pi)$ and $t \in U_2 = (-\pi, \pi)$, say. A collection of charts $\{(U_j, \psi_j)\}_{j \in \mathcal{J}}$ such that $\cup_{j \in \mathcal{J}} U_j = \mathcal{M}$ is called an *atlas* of \mathcal{M} . Thus, for two charts (U_x, ψ_x) and (U_y, ψ_y) and a point $z \in U_x \cap U_y \subset \mathcal{M}$, on defining $v_x = \psi_x(z)$ and $v_y = \psi_y(z)$, we have $v_y = \psi_y \circ \psi_x^{-1}(v_x)$, where \circ denotes composition of functions. For any $i, j \in \mathcal{J}$ such that $U_i \cap U_j \neq \emptyset$, $\psi_i \circ \psi_j^{-1}$ are called *transition maps*. These provide a way of transitioning between coordinate systems assigned by ψ_i and ψ_j .

20

25

30

We restrict attention to *orientable* manifolds with the property that Jacobian determinants of the transition maps are all positive. The class of orientable manifolds includes those most commonly encountered in physical contexts, including spheres and ellipsoids. It excludes certain exotic examples such as Möbius strips and Klein bottles.

35

If for any local chart (U_x, ψ_x) , ψ_x is smooth, i.e. has continuous partial derivatives of all orders, and has a smooth inverse, then ψ_x is called a *diffeomorphism*, ψ_x^{-1} is called a *local parameterization* of \mathcal{M} around x and \mathcal{M} is called a *smooth* (or *differentiable*) *manifold*. It follows that for overlapping charts, $\psi_i \circ \psi_j^{-1}$ is a diffeomorphism. In other words the determinants of the Jacobians of transformation involved in a change of coordinates are non-zero.

A function $f : \mathcal{M} \rightarrow \mathbb{R}$ is defined to be smooth if and only if $f \circ \psi^{-1}$ is smooth for every chart (U, ψ) in the atlas. There are of course many choices of atlas (local parameterizations) and a function f that is smooth with respect to one atlas is smooth with respect to any (e.g. Lee, 2003). More generally, the geometric properties of the manifolds are consistent regardless of the choice of local parameterization made.

For ψ_x^{-1} a local parameterization of \mathcal{M} around x , let $u_x = \psi_x(x)$. The *tangent space* $T_x\mathcal{M}$ to \mathcal{M} at x is the space of all tangent vectors v of \mathcal{M} at x , where v is defined as a tangent vector at x if there exists a parameterized curve $\gamma : (-\varepsilon, \varepsilon) \rightarrow \mathbb{R}^m$ such that $\gamma(t) \in \mathcal{M}$ for all $t \in (-\varepsilon, \varepsilon)$, $\gamma(0) = x$ and $\gamma'(0) = \{\gamma'_1(0), \dots, \gamma'_d(0)\} = v$. An equivalent definition of $T_x\mathcal{M}$ is the image of the derivative of ψ_x^{-1} at u_x , written $(D\psi_x^{-1})(u_x)$. Thus, if \mathcal{M} is embedded in \mathbb{R}^m , then $(D\psi_x^{-1})(u_x) : \mathbb{R}^d \rightarrow \mathbb{R}^m$ and $T_x\mathcal{M}$ is a d -dimensional subspace of \mathbb{R}^m . It is clear in the latter case that $T_x\mathcal{M}$ is a vector space of dimension d . This in fact holds even when there is no embedding space for \mathcal{M} . Thus $T_x\mathcal{M}$ possesses a basis, which changes for every point x on the manifold.

1.2. Geodesics and integration on Riemannian manifolds

For a fixed coordinate system, i.e. fixed chart (U, ψ) , we may write the local coordinates for the point x as $\psi(x) = \{\omega^1(x), \dots, \omega^d(x)\}$, the indices being identifiers, not powers. Define the infinitesimal squared distance ds^2 between two neighbouring points x and $x + dx$ in this coordinate system by the inner product

$$ds^2 = g_{ij}d\omega^i d\omega^j = \sum_{i,j} g_{ij}d\omega^i d\omega^j, \quad (\text{S1})$$

where the first equality uses the Einstein convention of summing over indices that appear in lower (covariant) and upper (contravariant) positions. The coefficients g_{ij} are functions of the coordinate system used. If the quadratic differential form in (S1) is positive definite, it is called a *Riemannian metric* and any manifold characterized by such a metric is called a *Riemannian manifold*. In Euclidean space using Cartesian coordinates, ds^2 is simply a sum of squared components from each coordinate axis. However (S1) is more general and allows for non-orthogonal coordinate systems.

A geodesic of a surface in Euclidean 3-space is familiar as a curve whose curvature relative to the surface is zero at all points. Geodesics on a Riemannian manifold are generalizations of this as a curve whose first curvature relative to \mathcal{M} is zero at all points. They are the shortest piecewise continuously differentiable curves between any given pair of points, and the length of such a curve is called the *geodesic distance*. We write the geodesic distance between two points $x, y \in \mathcal{M}$ as $d_g(x, y)$.

The natural volume element on an oriented Riemannian manifold of dimension d is the *Riemannian volume form* or *Levi-Civita tensor* given, in the aforementioned local coordinates by

$$d\text{vol} = \{\det(g_{ij})\}^{1/2} d\omega^1 \wedge \dots \wedge d\omega^d,$$

where $d\omega^1 \wedge \dots \wedge d\omega^d$ is the usual volume element on \mathbb{R}^d given by the exterior (or wedge) product in the coordinates specified by $\psi(x) = \{\omega^1(x), \dots, \omega^d(x)\}$, and (g_{ij}) is the matrix with (i, j) th entry g_{ij} . Clearly ds^2 , $d\text{vol}$, and any other geometric properties of \mathcal{M} are invariant to the

choice of coordinate system, and so it is not necessary in the coordinate-free treatment to specify g_{ij} explicitly.

80

Let K be any compact subset of \mathcal{M} . Then the volume of K is

$$\text{Vol}(K) = \int_K d\text{vol}.$$

The Riemannian volume form is the analogue of d -dimensional Lebesgue measure on Euclidean space.

A coordinate-free calculus on manifolds entails the notions of pushforward and pullbacks. For a smooth map, $f : \mathcal{N} \rightarrow \mathcal{M}$, say, the pushforward defines the corresponding map between the vectors in the tangent spaces and the pullback defines the converse. Along with the logarithmic and exponential maps, to be defined below, this allows operations such as changes of variables to be formulated on the tangent spaces and mapped back to the manifold. Let $\gamma_{x,v}$ be the geodesic such that $\gamma_{x,v}(0) = x \in \mathcal{M}$ and $\gamma'_{x,v}(0) = v \in T_x\mathcal{M}$. The exponential map is defined as $\exp_x(v) = \gamma_{x,v}(1) \in \mathcal{M}$ and the logarithmic map as $\log_x(q) = \exp_x^{-1}(q)$, $q \in \mathcal{M}$.

85

90

Let (U, ψ) be a local chart. The pullback $(\psi^{-1})^*\omega$ of the d -form ω on U can be written as $f(x)dx^1 \wedge \dots \wedge dx^d$ for some $f : \mathbb{R}^d \rightarrow \mathbb{R}$ where $dx^1 \wedge \dots \wedge dx^d = dx$, say, is the volume element on \mathbb{R}^d written in terms of the exterior product of the local coordinate system. The integral of ω on U is

$$\int_U \omega = \int_{\psi(U)} (\psi^{-1})^*\omega = \int_{\psi(U)} f(x)dx.$$

2. PROOFS

95

2.1. Preliminary lemmata

LEMMA S1. Suppose that (\mathcal{M}, g) is a Riemannian manifold. Consider the edge correction factors given by Equations (5) and (6) of the main text with a Gaussian kernel

$$k\{d_g(\cdot, y)\} = (2\pi)^{-d/2} \exp\{-d_g^2(\cdot, y)/2\}.$$

For fixed $x, y \in \mathcal{M}$ these satisfy

$$\begin{aligned} c_h(x, y) &\rightarrow \frac{K}{(2\pi)^{d/2}} & (h \rightarrow 0), \\ h^d c_h(x, y) &\rightarrow \frac{\text{Vol}(W)}{(2\pi)^{d/2}} & (h \rightarrow \infty), \end{aligned}$$

100

where $K = \int_{\mathbb{R}^d} \exp(-\|x\|^2/2)dx < \infty$.

Proof of Lemma S1. Since the edge correction factors are effectively symmetric in x and y we need only consider one of Equations (5) and (6). We use Equation (5):

$$c_h(x, y) = c_h(x) = \frac{1}{h^d} \int_W k \left\{ \frac{d_g(x, z)}{h} \right\} d\text{vol}(z). \quad (\text{S2})$$

Let $B_{\mathcal{M}}(x, r)$ be a ball of radius r centred at $x \in \mathcal{M}$, so that by definition the minimum geodesic distance between x and any point in $\mathcal{M} \setminus B_{\mathcal{M}}(x, r)$ is r . Consider the decomposition of (S2) with k replaced by the Gaussian kernel:

105

$$\frac{1}{(2\pi h^2)^{d/2}} \left\{ \int_{W \setminus B_{\mathcal{M}}(x, r)} + \int_{W \cap B_{\mathcal{M}}(x, r)} \right\} \exp \left\{ -\frac{d_g^2(x, z)}{2h^2} \right\} d\text{vol}(z) = \text{I} + \text{II},$$

say. Then

$$I \leq (2\pi h^2)^{-d/2} \exp(-r^2/2h^2) \text{Vol}\{W \setminus B_{\mathcal{M}}(x, r)\} = O\{h^{-d} \exp(-h^2)\}.$$

Gauss's Lemma (e.g. Carmo, 1992) maps a ball $B_{\mathcal{M}}(x, r)$ on \mathcal{M} to one on the tangent plane $T_x \mathcal{M}$ through the logarithmic map: $\log_x\{B_{\mathcal{M}}(x, r)\} = B_{T_x \mathcal{M}}(0, r)$. The calculations below entail an additional restriction to W and therefore for typographical reasons we introduce the set

$$A(x, r) = \log_x\{B_{\mathcal{M}}(x, r) \cap W\} = B_{T_x \mathcal{M}}(0, r) \cap \log_x(W).$$

A consequence of Gauss's lemma is that the exponential map is radially symmetric, i.e. $d_g^2\{x, \exp_x(z)\} = \|z\|^2$. In the control over II, we thus have

$$\begin{aligned} & \frac{1}{(2\pi h^2)^{d/2}} \int_{W \cap B_{\mathcal{M}}(x, r)} \exp\left\{\frac{-d_g^2(x, z)}{2h^2}\right\} d\text{vol}(z) \\ &= \frac{1}{(2\pi h^2)^{d/2}} \int_{A(x, r)} \exp\left[\frac{-d_g^2\{x, \exp_x(z)\}}{2h^2}\right] \theta_x\{\exp_x(z)\} dz \\ &= \frac{1}{(2\pi h^2)^{d/2}} \int_{A(x, r)} \exp\left\{\frac{-\|z\|^2}{2h^2}\right\} \theta_x\{\exp_x(z)\} dz. \end{aligned}$$

As defined in the statement of the lemma, $K = \int_{\mathbb{R}^d} \exp(-\|x\|^2/2) dx < \infty$, where the finiteness follows from the Nash embedding theorem (Nash, 1956). Observe that $\lim_{h \rightarrow 0} (1/K h^d) \exp(-\|z\|^2/2h^2)$ is a Dirac delta at z . Since $\theta_x\{\exp_x(z)\}$ depends smoothly on z , $\theta_x\{\exp_x(0)\} = \theta_x(x) = 1$ and is defined over the compact space of $B_{T_x \mathcal{M}}(0, r)$. The Dirac-delta property $\int_{\mathbb{R}^d} f(x) \delta(dx) = f(0)$ in the limit as $h \rightarrow 0$ allows us to write the previous displayed equation as

$$\frac{K}{(2\pi)^{d/2}} \theta_x(0) = \frac{K}{(2\pi)^{d/2}},$$

as required.

The behaviour as $h \rightarrow \infty$ is established using a similar argument to Cronie & Van Lieshout (2018). For the Gaussian kernel $\lim_{h \rightarrow \infty} k\{d_g(x, y)/h\} = (2\pi)^{-d/2}$, therefore,

$$\begin{aligned} \lim_{h \rightarrow \infty} h^d c_h(x, y) &= \lim_{h \rightarrow \infty} \int_W \frac{1}{(2\pi)^{d/2}} \exp\left\{-\frac{d_g(x, y)}{2h^2}\right\} d\text{vol}(z) \\ &= \int_W \frac{1}{(2\pi)^{d/2}} \left[\lim_{h \rightarrow \infty} \exp\left\{-\frac{d_g(x, y)}{2h^2}\right\} \right] d\text{vol}(z) \\ &= \int_W \frac{1}{(2\pi)^{d/2}} d\text{vol}(z) = \frac{\text{Vol}(W)}{(2\pi)^{d/2}}, \quad \square \end{aligned}$$

where the second line follows by the dominated convergence theorem.

LEMMA S2. For the Gaussian kernel, the edge correction factors as defined by Equations (5) and (6) are continuous in $h \in (0, \infty)$.

Proof of Lemma S2. If $c_h(\mathbf{x}, \mathbf{y}) = 1$, the conclusion follows directly.

As before, by symmetry it is sufficient to consider Equation (5). Fix $\epsilon, h_0 > 0$ and with $x, z \in \mathcal{M}$ both treated initially as fixed, $f_x(z, h) = (2\pi)^{-d/2} \exp\{-d_g^2(x, z)/2h^2\}$ is continuous in h by continuity of $1/h^2$ and \exp , since the composition of continuous functions is continuous. It

follows that for $\epsilon' = \epsilon/\text{Vol}(\mathcal{M})$ there exists $\delta > 0$ such that $|h - h_0| < \delta$ implies

$$|f_x(z, h) - f_x(z, h_0)| < \epsilon' = \frac{\epsilon}{\text{Vol}(W)}.$$

It follows that, for fixed $x \in \mathcal{M}$

$$\begin{aligned} \left| \int_W \{f_x(z, h) - f_x(z, h_0)\} d\text{vol}(z) \right| &\leq \int_W |f_x(z, h) - f_x(z, h_0)| d\text{vol}(z) \\ &< \int_W \frac{\epsilon}{\text{Vol}(W)} d\text{vol}(z) = \epsilon. \end{aligned}$$

We have shown that $\int_W f_x(z, h) d\text{vol}(z)$ and therefore $c_h(x, y)$ is continuous in $h \in (0, \infty)$. \square

2.2. Proof of Proposition 3

Proof. Observe from Equation (8) that $\hat{\rho}_{h,1}^\bullet(x)$ is of the form

$$Z(x) = \frac{1\{N_X(W) \neq 0\}}{N_X(W)} \sum_{y \in X \cap W} f(x, y).$$

135

In order to show pointwise unbiasedness and consistency we require the following identities given by Cucala (2007) translated to the setting of Riemannian manifolds,

$$E\{Z(x)\} = \left\{1 - e^{-\mu(W)}\right\} \int_W f(x, y) \rho_1(y) d\text{vol}(y), \quad (\text{S3})$$

$$\begin{aligned} \text{Var}(Z) &= A\{\mu(W)\} \int_W f^2(x, y) \rho_1(y) d\text{vol}(y) - \\ &\quad \left\{ \int_W f(x, y) \rho_1(y) d\text{vol}(y) \right\} \left[A\{\mu(W)\} - e^{-\mu(W)} - e^{-2\mu(W)} \right], \quad (\text{S4}) \end{aligned} \quad 140$$

where X is a Poisson process with intensity function ρ ,

$$A\{\mu(W)\} = E \left[\frac{1\{N_X(W) \neq 0\}}{N_X(W)} \right],$$

$\rho_1(x) = \rho(x)/\mu(\mathcal{M})$ and $f : \mathcal{M} \mapsto \mathbb{R}$ is measurable and nonnegative.

We first show asymptotic unbiasedness. For $\bullet \in \{\text{glo}, \text{loc}\}$ and from Equation (S3) we have

$$\begin{aligned} E\{\hat{\rho}_{h,1}^\bullet(x)\} &= \left\{1 - e^{-\mu(W)}\right\} \int_{\mathcal{M}} 1(y \in W) \frac{c_h^{-1}(x, y)}{h^d} k \left\{ \frac{d_g(x, y)}{h} \right\} \rho_1(y) d\text{vol}(y) \quad (\text{S5}) \\ &= \left\{1 - e^{-\mu(W)}\right\} \left[\int_{B_{\mathcal{M}}(x, r)} 1(y \in W) \frac{c_h^{-1}(x, y)}{h^d} k \left\{ \frac{d_g(x, y)}{h} \right\} \rho_1(x) d\text{vol}(y) \right. \\ &\quad \left. + \int_{\mathcal{M} \setminus B_{\mathcal{M}}(x, r)} 1(y \in W) \frac{c_h^{-1}(x, y)}{h^d} k \left\{ \frac{d_g(x, y)}{h} \right\} \rho_1(y) d\text{vol}(y) \right] \end{aligned} \quad 145$$

where $0 < r < r^*$, r^* being the global injectivity radius of \mathcal{M} . Applying an identical argument to that used in the proof of Lemma S1 to show that the first term of Equation (S5) goes to 0, we can show that the second term here also goes to 0 as $h \rightarrow 0$. Thus we shall focus on the integral

150 in the first term,

$$\begin{aligned}
& \int_{B_{\mathcal{M}}(x,r)} 1(y \in W) \frac{c_h^{-1}(x,y)}{h^d} k \left\{ \frac{d_g(x,y)}{h} \right\} \rho_1(y) d\text{vol}(y) \\
&= \int_{B_{T_x \mathcal{M}}(0,r)} \frac{c_h^{-1}\{x, \exp_x(y)\}}{(2\pi)^{d/2} h^d} \exp \left[-\frac{d_g^2\{x, \exp_x(y)\}}{2h^2} \right] g(y) dy \\
&= \int_{B_{T_x \mathcal{M}}(0,r)} \frac{c_h^{-1}\{x, \exp_x(y)\}}{(2\pi)^{d/2} h^d} \exp \left(-\frac{\|y\|^2}{2h^2} \right) g(y) dy = I(h),
\end{aligned}$$

say, where $g(y) = 1\{\exp_x(y) \in W\} \rho_1\{\exp_x(y)\} \theta_x\{\exp_x(y)\}$. By Lemma S1 for any $0 < \epsilon < K/(2\pi)^{d/2}$ we can find a small enough h' such that for $h < h'$, $K/(2\pi)^{d/2} - \epsilon < c_h(x,y) < K/(2\pi)^{d/2} + \epsilon$. Setting $K' = K/(2\pi)^{d/2}$ we have the following lower and upper bounds on $I(h)$

155

$$\begin{aligned}
L(h) &= \int_{B_{T_x \mathcal{M}}(0,r)} \frac{1}{(2\pi)^{d/2} h^d (K' + \epsilon)} \exp \left(-\frac{\|y\|^2}{2h^2} \right) g(y) dy \\
&< I(h) < \int_{B_{T_x \mathcal{M}}(0,r)} \frac{1}{(2\pi)^{d/2} h^d (K' - \epsilon)} \exp \left(-\frac{\|y\|^2}{2h^2} \right) g(y) dy = U(h).
\end{aligned}$$

Consider the lower bound,

$$\begin{aligned}
L(h) &= \frac{K}{(2\pi)^{d/2} (K' + \epsilon)} \int_{B_{T_x \mathcal{M}}(0,r)} \frac{1}{K h^d} \exp \left(-\frac{\|y\|^2}{2h^2} \right) g(y) dy \\
&= \frac{K'}{(K' + \epsilon)} \int_{B_{T_x \mathcal{M}}(0,r)} \frac{1}{K h^d} \exp \left(-\frac{\|y\|^2}{2h^2} \right) g(y) dy.
\end{aligned}$$

Then, on letting $g(0) = 1\{\exp_x(0) \in W\} \rho_1\{\exp_x(0)\} \theta_x\{\exp_x(0)\} = \rho_1(x)$,

$$\lim_{h \rightarrow 0} L(h) = \frac{K'}{(K' + \epsilon)} g(0) = \frac{K'}{(K' + \epsilon)} g(0) = \frac{K'}{(K' + \epsilon)} \rho_1(x).$$

160 An analogous argument for $U(h)$ gives

$$\frac{K'}{(K' + \epsilon)} \rho_1(x) < \lim_{h \rightarrow 0} I(h) < \frac{K'}{(K' - \epsilon)} \rho_1(x).$$

Then since this holds for all ϵ we have

$$\lim_{h \rightarrow 0} I(h) = \rho_1(y),$$

and since $1 - e^{-\mu(W)} \rightarrow 1$ as $\mu(W) \rightarrow \infty$, asymptotic unbiasedness follows.

By Equation (S4),

$$\begin{aligned}
165 \quad \text{Var}\{\hat{\rho}_{h,1}^\bullet(x)\} &= A\{\mu(W)\} \int_{\mathcal{M}} 1(y \in W) \frac{c_h^{-2}(x,y)}{h^{2d}} k^2 \left\{ \frac{d_W(x,y)}{h} \right\} \rho_1(y) d\text{vol}(y) \\
&- \left[\int_{\mathcal{M}} 1(y \in W) \frac{c_h^{-1}(x,y)}{h^n} k \left\{ \frac{d_g(x,y)}{h} \right\} \rho_1(y) d\text{vol}(y) \right] \left[A\{\mu(W)\} - e^{-\mu(W)} - e^{-2\mu(W)} \right].
\end{aligned} \tag{S6}$$

Notice that

$$\begin{aligned} & \int_{\mathcal{M}} 1(y \in W) \frac{c_h^{-1}(x, y)}{h^n} k \left\{ \frac{d_g(x, y)}{h} \right\} \rho_1(y) \, d\text{vol}(y) \\ &= \frac{E \left\{ \hat{\rho}_{h,1}^\bullet(x) \right\}}{1 - e^{-\mu(W)}} = O(1), \quad (h \rightarrow 0, \mu(W) \rightarrow \infty) \end{aligned}$$

Next consider $A\{\mu(W)\}$.

$$\begin{aligned} A\{\mu(W)\} &= E \left[\frac{1\{N_X(W) \neq 0\}}{N_X(W)} \right] = \sum_{n=1}^{\infty} \frac{e^{-\mu(W)} \{\mu(W)\}^n}{n \cdot n!} \\ &\leq \sum_{n=1}^{\infty} \frac{2e^{-\mu(W)} \{\mu(W)\}^n}{(n+1)!} = \frac{2e^{-\mu(W)}}{\mu(W)} \sum_{n=1}^{\infty} \frac{\{\mu(W)\}^{n+1}}{(n+1)!} \\ &= \frac{2e^{-\mu(W)}}{\mu(W)} \sum_{n=2}^{\infty} \frac{\{\mu(W)\}^n}{n!} \leq \frac{2e^{-\mu(W)}}{\mu(W)} \sum_{n=0}^{\infty} \frac{\{\mu(W)\}^n}{n!} = \frac{2}{\mu(W)}, \end{aligned}$$

where the first inequality follows since $k \geq 1$ implies $1/k \leq 2/(k+1)$. Thus $A\{\mu(W)\} \rightarrow 0$ as $\mu(W) \rightarrow \infty$, showing that the second line of Equation (S6) tends to 0 as $h \rightarrow 0$ and $\mu(W) \rightarrow \infty$. 170

It only remains to control the integral in the first term of Equation (S6). By Lemma S1, for any $0 < \epsilon < \min\{r^*, K/(2\pi)^{d/2}\}$ there exists a h' such that for $h < h'$, $0 < K/(2\pi)^{d/2} - \epsilon < c_h(x, y)$. Consider $h < h'$ and decompose the integral. As in previous arguments, the integral over $\mathcal{M} \setminus B_{\mathcal{M}}(x, r)$ tends to zero. Let $K' = K/(2\pi)^{d/2}$. Then

$$\begin{aligned} & \int_{B_{\mathcal{M}}(x, h)} 1(y \in W) \frac{c_h^{-2}(x, y)}{h^{2d}} k^2 \left\{ \frac{d_g(x, y)}{h} \right\} \rho_1(y) \, d\text{vol}(y) \\ &= \int_{B_{T_x \mathcal{M}}(0, h)} 1\{\exp_x(y) \in W\} \frac{c_h^{-2}\{x, \exp_x(y)\}}{(2\pi)^{d/2} h^{2d}} \exp\left(-\frac{\|y\|^2}{h^2}\right) \rho_1\{\exp_x(y)\} \theta_x\{\exp_x(y)\} \, dy \\ &\leq \frac{1}{(K' - \epsilon)^2} \int_{B_{T_x \mathcal{M}}(0, h)} \frac{1\{\exp_x(y) \in W\}}{(2\pi)^{d/2} h^{2d}} \exp\left(-\frac{\|y\|^2}{h^2}\right) \rho_1\{\exp_x(y)\} \theta_x\{\exp_x(y)\} \, dy \\ &\leq \frac{1}{(K' - \epsilon)^2} \int_{B_{T_x \mathcal{M}}(0, h)} \frac{1}{(2\pi)^{d/2} h^{2d}} \exp\left(-\frac{\|y\|^2}{h^2}\right) \rho_1\{\exp_x(y)\} \theta_x\{\exp_x(y)\} \, dy \\ &= \frac{1}{(K' - \epsilon)^2} \int_{B_{T_x \mathcal{M}}(0, 1)} \frac{h^d}{(2\pi)^{d/2} h^{2d}} \exp(-\|z\|^2) \rho_1\{\exp_x(hz)\} \theta_x\{\exp_x(hz)\} \, dz \\ &= \frac{1}{h^d (K' - \epsilon)^2} \int_{B_{T_x \mathcal{M}}(0, 1)} \frac{1}{(2\pi)^{d/2}} \exp(-\|z\|^2) \rho_1\{\exp_x(hz)\} \theta_x\{\exp_x(hz)\} \, dz. \end{aligned}$$

where we have used a change of variables $z = y/h$. Both $\rho_1\{\exp_x(hz)\}$ and $\theta_x\{\exp_x(hz)\}$ are bounded for small h since $\lim_{h \rightarrow 0} \rho_1\{\exp_x(hz)\} = \rho_1(x)$ and $\lim_{h \rightarrow 0} \theta_x\{\exp_x(hz)\} = 1$. Since $\exp(-\|z\|^2) \leq \exp(-\|0\|^2) = 1$, the whole integral is bounded $h \rightarrow 0$. Thus, $\text{Var}\{\hat{\rho}_{h,1}^\bullet(x)\} \rightarrow 0$ as $h \rightarrow 0, \mu(W) \rightarrow \infty$ provided that $A\{\mu(W)\}/h^d \rightarrow 0$ under the same notional limiting operation. 175 \square

2.3. Proof of Proposition 4

Proof. Parts of the argument follow Cronie & Van Lieshout (2018). 180

By Lemma S2, $\hat{\rho}_h$ is continuous in h . Further, since X is non-empty $\hat{\rho}_h(x) \geq h^{-d}k\{d_g(x, x)/h\}c_h^{-1}(x, y) > 0$ so that $\hat{\rho}_h^{-1}(x)$ exists and is continuous. Thus T is continuous. For $x \in X$ and $\bullet \in \{\text{glo}, \text{loc}\}$,

$$\begin{aligned}\hat{\rho}_h^\bullet(x) &= \sum_{y \in X \cap W} \frac{1}{(2\pi h^2)^{d/2}} \exp\left\{\frac{-d_g^2(x, y)}{2h^2}\right\} c_h^{-1}(x, y) \\ &\geq \frac{c_h^{-1}(x, x)}{h^d} \exp\left\{\frac{-d_g^2(x, x)}{2h^2}\right\} = \frac{1}{h^d c_h(x, x)},\end{aligned}\quad \square$$

185 where the inequality follows on splitting the sum of positive real numbers into that over $y = x \in X$ and the remainder, and discarding the latter. The previous display implies $\hat{\rho}_h^{-1}(x) \leq h^d c_h(x, x)$. Thus

$$T(\hat{\rho}_h) \leq h^d \sum_{x \in X \cap W} c_h(x, x).$$

Consider a particular realization of X . By Lemma S1, edge correction terms tend to $K/(2\pi)^{d/2}$ for $K = \int_{\mathbb{R}^d} \exp(-\|x\|^2) dx < \infty$ therefore $T(\hat{\rho}_h) \rightarrow 0$ as $h \rightarrow 0$.

190 Under the notional limiting operation $h \rightarrow \infty$ we have

$$\frac{c_h^{-1}(x, y)}{(2\pi)^{d/2} h^d} \exp\left\{\frac{-d_g^2(x, y)}{2h^2}\right\} \rightarrow \begin{cases} \frac{1}{\text{Vol}(W)}, & \text{if } c_h(x, y) \text{ is given by Equation (5) or (6);} \\ 0, & \text{if } c_h(x, y) = 1, \end{cases}$$

by Lemma S1, leading to

$$195 \quad T(\hat{\rho}_h) \rightarrow \begin{cases} \sum_{x \in X} \left\{ \sum_{y \in X} 1/\text{Vol}(W) \right\}^{-1} = \text{Vol}(W), & \text{if } c_h(x, y) \text{ is given by Equation (5) or (6);} \\ \infty, & \text{if } c_h(x, y) = 1. \end{cases}$$

3. ADDITIONAL SIMULATION RESULTS

3.1. Poisson process

Fig. S1 shows manifolds \mathcal{E}_1 , \mathcal{E}_2 and \mathcal{E}_3 , together with examples of the log-linear and log-modulation intensity functions from the simulation study in Section 6 of the main text. Details on simulating a homogeneous and inhomogeneous Poisson process on the surface of an ellipsoid are contained in Sections 4.1 and 4.2, respectively.

3.2. Log-Gaussian Cox Process

205 The simulation study in Section 6 of the main text is extended to the log-Gaussian Cox process (LGCP). Point patterns are again simulated on the ellipsoids \mathcal{E}_1 , \mathcal{E}_2 and \mathcal{E}_3 . Intensity function estimates are then computed using the cross-validations (CV) (Baddeley et al., 2015) and non-parametric (NP) (Cronie & Van Lieshout, 2018) bandwidth selection methods. The intensity function is estimated from point patterns sampled from three models of LGCP.

210 To define a LGCP on a manifold \mathcal{M} , first define a Gaussian random field on \mathcal{M} . The random function $U : \mathcal{M} \mapsto \mathbb{R}$ is a Gaussian random field (GRF) if for any $n \in \mathbb{N}$, $x_1, \dots, x_n \in \mathcal{M}$ and $a_1, \dots, a_n \in \mathbb{R}$, $\sum_{i=1}^n a_i U(x_i)$ is normally distributed. For GRF U , define random field Z :

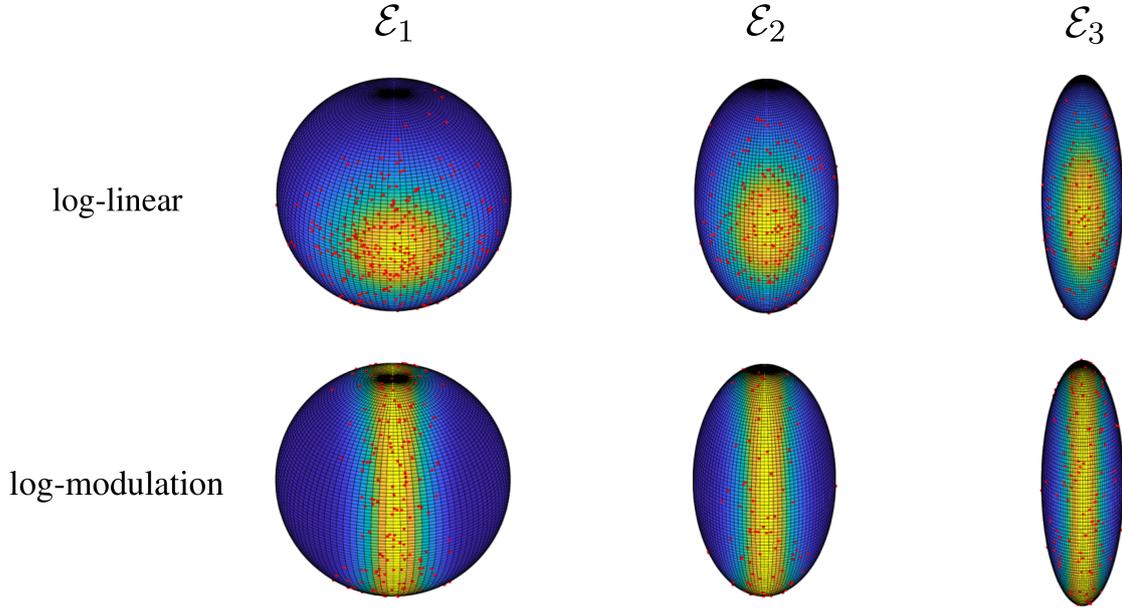


Fig. S1. Example point patterns of a log-linear and log-modulation Poisson process with parameters $\alpha_2 = 18$ and $\alpha_3 = 4$, respectively. The intensity is illustrated with the heatmap, with yellow representing high intensity and blue representing low intensity.

$\mathcal{M} \mapsto \mathbb{R}$ as $Z(x) = \exp\{U(x)\}$. Point process X is then said to be a LGCP on \mathcal{M} if X given $Z = z$ is a Poisson process with intensity z .

A GRF is defined via its mean function $\mu : \mathcal{M} \mapsto \mathbb{R}$ and covariance function $c : \mathcal{M} \times \mathcal{M} \mapsto \mathbb{R}$. Three models of LGCP are considered, labelled LGCP1, LGCP2 and LGCP3, with respective mean functions

$$\mu_1(x) = \log(\xi), \quad \text{homogeneous}; \quad (S7)$$

$$\mu_2(x) = 4 + 3x_1, \quad \text{log-linear}; \quad (S8)$$

$$\mu_3(x) = 6 \cos(8x_2) - 1, \quad \text{log-modulation}. \quad (S9)$$

For all three models the covariance function is of exponential form

$$c(x, y) = \sigma^2 \exp \left[-\frac{\{(x - y)^T(x - y)\}^{1/2}}{2\gamma^2} \right].$$

Model parameters ξ , σ^2 and γ are varied and the results are presented in Table S1. Note, following Møller & Waagepetersen (1998), the intensity function of a LGCP with driving random field $\exp\{U(x)\}$ is

$$\rho(x) = \exp \left\{ \mu(x) + \frac{c(x, x)}{2} \right\}.$$

Hence, the intensity function of the models considered here will be $\rho(x) = \exp \{ \mu(x) + \sigma^2/2 \}$. Examples from the log-linear (LGCP2) and log-modulation (LGCP3) models on \mathcal{E}_1 , \mathcal{E}_2 and \mathcal{E}_3 are presented in Fig. S2. Details on simulating a LGCP on an ellipse are given in Section 4.3.

215

220

225

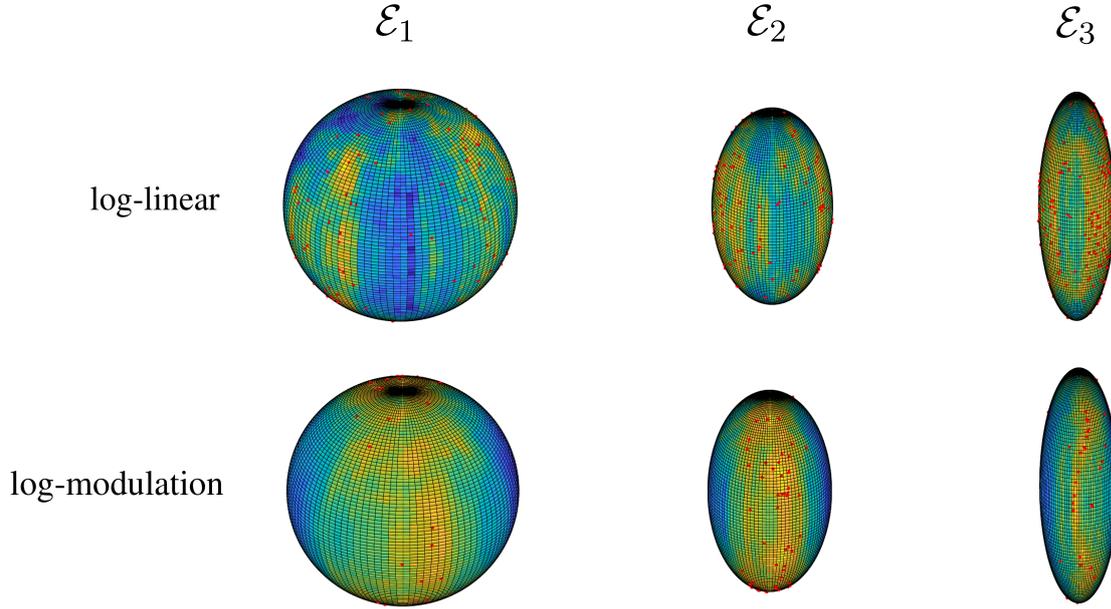


Fig. S2. Example point patterns of a log-linear and log-modulation LGCP all with parameters $(\sigma^2, \gamma^2) = (2 \log(2), 1/10)$. The generating GRF U is illustrated with the heatmap, with yellow representing high intensity and blue representing low intensity.

3.3. Strauss Process

The final simulation study is analogous to the previous ones only with Strauss processes used to generate the events.

230 Strauss and, more broadly, Markov processes have been briefly discussed on general differentiable manifolds in Jensen & Nielsen (2001) albeit with the primary focus being \mathbb{R}^n . Let N_{lf} be the set of locally finite point configurations on \mathcal{M} , with \mathcal{N}_{lf} the sigma-algebra generated by sets constructed from N_{lf} . Furthermore, let Z be a unit rate Poisson process over \mathcal{M} with accompanying measure P_Z . A Strauss process is defined to be the point process with density

$$f(s) \propto \beta^{n(s)} \exp \left[-\alpha \sum_{\{x,y\} \in s} 1_{\{d_g(x,y) \leq R\}} \right] \quad (\text{S10})$$

for $s \in N_{lf}$, where $n(s)$ is the cardinality of s and $\beta, \alpha, R > 0$, such that for $F \in \mathcal{N}_{lf}$

$$P(F) = \int_F f(s) dP_Z(s).$$

235 The constraint $\alpha > 0$ is sufficient for the density to be integrable and dictates the strength of the repulsive effect between events. The range of this interaction is determined by R .

The Papangelou conditional intensity for $s \in N_{lf}$ at $x \in \mathcal{M}$ is defined as $\lambda(x, s) = f(s \cup x) / f(s)$ (Møller & Waagepetersen, 2004), which for the Strauss process is

$$\lambda(x, s) = \frac{f(s \cup x)}{f(s)} = \beta \exp \left[-2\alpha \sum_{y \in s} 1_{\{d_g(x,y) \leq R\}} \right].$$

While this natural extension of the Strauss process from \mathbb{R}^n to \mathcal{M} is formed by replacing the canonical Euclidean metric with the Riemannian metric over \mathcal{M} , it is worth noting that the induced distance metric d_g is no longer stationary. For example, consider \mathbb{R}^n endowed with its canonical metric d . For any $x, y, z \in \mathbb{R}^n$ we have $d(x, y) = d(x + z, y + z)$, however, this is a special property of Euclidean space that does not generalize to manifolds. Consequently, a Strauss process on a manifold is in general non-stationary and need not even be homogeneous. As a result, there is, in general, no closed-form formula for the intensity function, which is problematic for calculating the ISE. On noting that the intensity of a point process X with Papangelou conditional intensity λ is given by $\rho(x) = E\{\lambda(x, X)\}$ (Møller & Waagepetersen, 2004, Proposition 6.2), it is possible to simulate realizations X_1, \dots, X_M and thereby approximate ρ by

$$\frac{1}{M} \sum_{i=1}^M \lambda(x, X_i). \tag{S11}$$

Additionally, $E\{N_X(\mathcal{M})\}$ is estimated as $(1/M) \sum_{i=1}^M N_{X_i}(\mathcal{M})$. This replaces the step of calculating the mean number of points using integration, as in the Poisson and LGCP studies.

In order to consider different forms of inhomogeneity beyond that imposed by the Strauss model of (S10), the process is thinned using one of the three following retention probabilities,

$$\begin{aligned} p_1(x) &= 1, && \text{pseudo-homogeneous;} \\ p_2(x) &= \frac{4 + 3x_1}{4 + 3a}, && \text{linear;} \\ p_3(x) &= \frac{8 + 6 \cos(8x_1)}{14}, && \text{modulation,} \end{aligned}$$

where a is the minor axis length of the ellipsoid. The three thinning models are labelled as SP1, SP2 and SP3, respectively. Note that SP1 is the original Strauss process and hence referred to as pseudo-homogeneous owing to the fact that its inhomogeneity is purely metric induced.

Model parameters β , α and R are varied and the results are presented in Table S2. Examples from the linear (SP2) and modulation (SP3) models on \mathcal{E}_1 , \mathcal{E}_2 and \mathcal{E}_3 are presented in Fig. S3. The number of replicates used to approximate ρ with (S11) is $M = 10^5$. Details on simulating a Strauss process on an ellipse are given in Section 4.4.

4. SIMULATING SPATIAL POINT PROCESSES ON MANIFOLDS

4.1. Homogeneous Poisson process on an ellipsoid

Simulation of the more complex processes discussed in this work build on that of the homogeneous Poisson process. Let X denote such a process, with intensity $\rho \in \mathbb{R}_+$ on an ellipsoid, \mathcal{M} , having semi-major axis lengths (a, b, c) and surface area $S_{a,b,c}$. We first simulate the total number of points as $N_X(\mathcal{M}) \sim \text{Poisson}(\rho S_{a,b,c})$. Conditional on the realized value, these points are distributed uniformly on the surface of the ellipsoid using the rejection sampling technique discussed in Kopytov & Mityushov (2013). Kopytov & Mityushov (2013) propose to simulate points under a density f in the local chart of the manifold such that under transformation by the inverse chart, the resulting points are uniform on the original manifold. Without loss of generality we can consider simulating only over a single chart within the atlas of a manifold, as the manifold can be partitioned into disjoint sets such that each set in the partition is contained within at least one local chart.

More precisely, let (\mathcal{M}, g) be a Riemannian manifold and suppose that uniformly distributed points over the compact set $W \subset \mathcal{M}$ are required. Suppose further that there exists a local chart

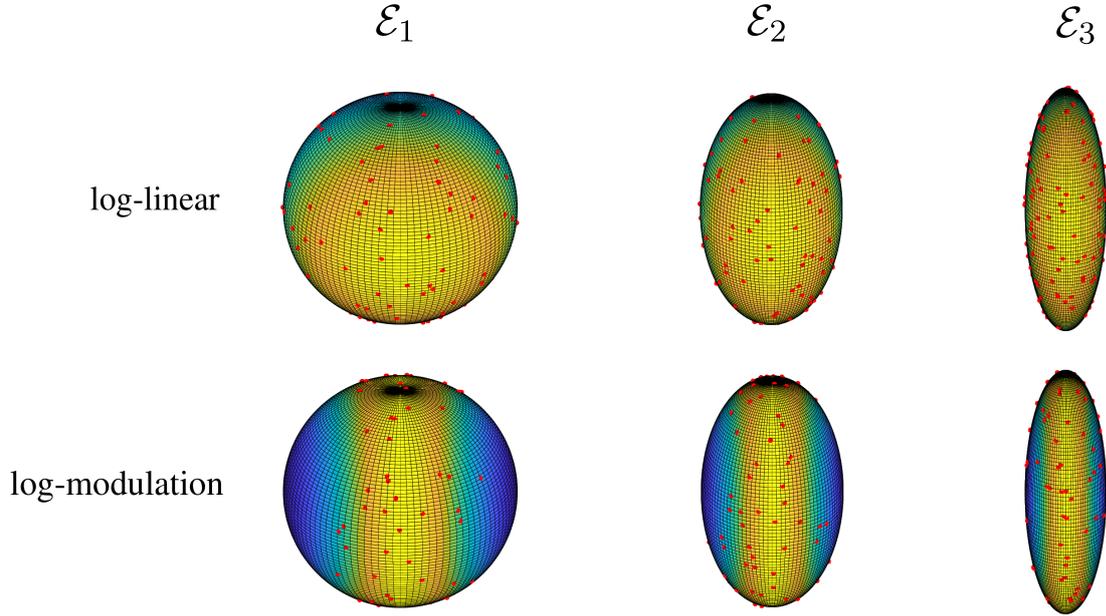


Fig. S3. Example point patterns of a linear and modulation Strauss process all with parameters $(\beta, \alpha, R) = (400, 0.1, 0.1)$. The intensity is illustrated with the heatmap, with yellow representing high intensity and blue representing low intensity.

(U_x, ψ_x) such that $W \subset U_x$ for $x \in \mathcal{M}$ and define $V_x = \psi_x(U_x)$. For ease of exposition, partition U_x such that $U_{(x,1)} = W$ and $U_{(x,2)} = U_x \setminus W$ and restrict attention to the local chart (W, ψ_x) . Similarly, define $V_{(x,1)} = \psi(W)$ and $V_{(x,2)} = \psi(U_{(x,2)})$. Kopytov & Mityushov (2013) show uniform points over W are obtained by simulating points on $V_{(x,1)} = \psi(W)$ according to the density,

$$f(z) = \frac{\sqrt{\det(g_{ij})}}{\text{Vol}(W)}, \quad z \in V_{(x,1)} \quad (\text{S12})$$

and mapping them to W using $x = \psi^{-1}(z)$. To simulate a point z according to Equation S12, Kopytov & Mityushov (2013) propose a rejection sampling algorithm.

A local chart for the ellipsoid is presented at the beginning of Section 6 in the main paper, i.e. $x \equiv (x_1, x_2, x_3) = \{a \sin(\theta) \cos(\phi), b \sin(\theta) \sin(\phi), c \cos(\theta)\}$ where $\theta \in [0, \pi]$ and $\phi \in [0, 2\pi]$. Thus,

$$\det(g_{ij}) = \sin^2(\theta) a^2 b^2 \{1 - (1 - c^2/a^2) \sin^2(\theta) \cos^2(\phi) - (1 - c^2/b^2) \sin^2(\theta) \sin^2(\phi)\}$$

and points in $\psi(\mathcal{M})$ are sampled according to the density

$$f(\theta, \phi) = \frac{[\sin^2(\theta) a^2 b^2 \{1 - (1 - c^2/a^2) \sin^2(\theta) \cos^2(\phi) - (1 - c^2/b^2) \sin^2(\theta) \sin^2(\phi)\}]^{1/2}}{\text{Vol}(\mathcal{M})},$$

using the rejection algorithm of Kopytov & Mityushov (2013). Algorithm S1 gives details.

Algorithm S1. Simulation of a homogeneous Poisson process on an ellipsoid

```

Require: intensity  $\rho$ , semi-major axis lengths  $(a, b, c)$ 
Sample  $N \sim \text{Poisson}(\rho S_{a,b,c})$ 
Set  $n = 0$ 
While  $n < N$ 
  Sample  $R \sim U(0, 1)$ ,  $\theta \sim U(0, \pi)$  and  $\phi \sim U(0, 2\pi)$ 
  If  $f(\theta, \phi) > R \max_{\theta, \phi} f(\theta, \phi)$ 
    Set  $X_{n+1} = \{a \sin(\theta) \cos(\phi), b \sin(\theta) \cos(\phi), c \cos(\theta)\}$ 
     $n = n + 1$ 
  End If
End While

```

4.2. Inhomogeneous Poisson process on an ellipsoid

290

A realization of an inhomogeneous Poisson process can be constructed from a homogeneous one by the process of independent thinning. This is discussed in the Euclidean case in (Møller & Waagepetersen, 2004, Section 3.2.3). Similarly, for an inhomogeneous Poisson process defined over a manifold \mathcal{M} with intensity $\rho : \mathcal{M} \mapsto \mathbb{R}_+$ such that $\rho_0 = \sup_{x \in \mathcal{M}} \rho(x) < \infty$, we first simulate a homogeneous Poisson process with intensity ρ_0 over \mathcal{M} and remove an event at $x \in \mathcal{M}$ with probability $1 - p(x) = 1 - \rho(x)/\rho_0$. See Algorithm S2 for implementation when \mathcal{M} is an ellipsoid.

295

Algorithm S2. Simulation of an inhomogeneous Poisson process on an ellipsoid

```

Require: intensity  $\rho(\cdot)$ , semi-major axis lengths  $(a, b, c)$ 
Calculate  $\rho_0 = \sup_{x \in \mathcal{M}} \rho(x)$ 
Sample homogeneous Poisson process,  $\tilde{X}$ , with intensity  $\rho_0$  using Algorithm S1
For  $X_i \in \tilde{X}$ 
  Sample  $R \sim U(0, 1)$ 
  If  $R < p(X_i)$ 
    Keep  $X_i$ 
  Else
    Discard  $X_i$ 
  End If
End For

```

4.3. Log Gaussian Cox process on an ellipsoid

Let $U(x)$ be a Gaussian random field (GRF) over \mathcal{M} . Conditional on the realization $u(x)$, one realization of a log Gaussian Cox process (LGCP) is obtained as a realization of an inhomogeneous Poisson process with intensity function $\exp\{u(x)\}$. In view of the discussion in Section 4.2, the only missing element is simulation of a GRF over \mathcal{M} . For the simulation study considered in the paper this is accomplished by simulating a GRF $\tilde{U}(x)$ over \mathbb{R}^3 using the RandomFields (Schlather et al., 2020) package available in R and defining $U(x) = \tilde{U}(x)$ for $x \in \mathcal{M}$, which is still a GRF but now restricted to \mathcal{M} . Since it is not possible to store a GRF as a function defined over all of points in \mathcal{M} , evaluation is on a finite grid of points on \mathcal{M} . Specifically, let $I = \{i_1, \dots, i_m\}$ for $i_j \in \mathcal{M}$ and $m \in \mathbb{N}$. The GRF is sampled at the points in I giving

300

305

$\{U_1, \dots, U_m\}$ where $U_j = U(i_j)$. The GRF over \mathcal{M} is then approximated as

$$U^*(x) = \sum_{j=1}^m \mathbb{1}\{\text{Ne}_g(x, i_j, I) = 1\}U_j, \quad (\text{S13})$$

310 where $\text{Ne}_g(x, i_j, I) = 1$ if x 's nearest neighbour defined by the metric g in I is i_j and 0 otherwise. By appropriate selection of the mean and covariance function, the resulting GRF is almost surely continuous (see (Møller & Waagepetersen, 2004, Section 5.6.1)) so that $U^*(x)$ is a close approximation to $U(x)$ for $x \in \mathcal{M}$ provided that the grid is sufficiently granular. The finite chart I in the simulations from the main paper is defined using equidistantly spaced points in $\theta \in [0, \pi)$ and $\phi \in [0, 2\pi)$ for the local chart defined at the beginning of Section 6.

Algorithm S3. Simulation of a LGCP on an ellipsoid

Require: Finite grid over ellipsoid I , mean function $\mu(\cdot)$, covariance function $c(\cdot, \cdot)$, and semi-major axis lengths, (a, b, c)
 Sample the GRF, U , with mean function μ and covariance function c at points I
 Define the approximation U^* to U as defined by Equation S13
 Using Algorithm S2, sample an inhomogeneous Poisson process, X , with intensity $\rho(x) = \exp\{U^*(x)\}$

4.4. Strauss process on an ellipsoid

Simulation of the Strauss process uses a spatial birth-death-move Metropolis Hastings (SB-DMMH) algorithm. The main difference from the Euclidean case (Møller & Waagepetersen, 320 2004, Chapter 7) arises from the proposal distribution for whether a birth or move step is initiated, which is defined over \mathcal{M} instead of \mathbb{R}^d .

The SB-DMMH algorithm has three different Metropolis steps: a birth, death and move step each with their own Hasting's ratio. These are briefly discussed in turn. For every step, suppose there is an initial configuration X_i for the point process. The algorithm yields X_{i+1} . 325 Let $(y_1, \dots, y_n) \in \mathcal{M}^n$ be a specific ordering of the points in X_i , where $n = N_{X_i}(\mathcal{M})$. For example, if there are three points in the pattern then $(y_1, y_2, y_3) \neq (y_2, y_3, y_1)$ even though $\{y_1, y_2, y_3\} = \{y_2, y_3, y_1\}$. Here, $f(\cdot)$ is the density of a Strauss process, defined up to a constant of proportionality in (S10).

• **Move Step:** Given a configuration $X_i \neq \emptyset$ and denoting $\bar{y} = (y_1, \dots, y_n)$, $n = N_{X_i}(\mathcal{M})$, to 330 be a specific ordering of X_i , the move step picks an initial point at random $y_j \in X_i$ and a proposal point y^* sampled from a proposal distribution $q_{\text{move},j}(\bar{y}, \cdot)$ say, where $q_{\text{move},j}(\bar{y}, \cdot) : \mathcal{M} \mapsto \mathbb{R}_+$ is a density over \mathcal{M} that depends on the current state of the chain, X_i and the initial choice of point y_j . In Algorithm S4 the proposal distribution is uniform over the ellipsoid, i.e. $q_{\text{move},j}(\bar{y}, x) = 1/\text{Vol}(\mathcal{M})$, which is operationalized using Kopytov & Mityushov (2013)'s 335 procedure. The Hasting's ratio is then constructed as

$$r_{\text{move},i}(\bar{y}, y^*) = \frac{f\{(X_i \setminus y_i) \cup y^*\} q_{\text{move},j}\{(y_1, \dots, y_{i-1}, y^*, y_{i+1}, \dots, y_n), y_j\}}{f(X_i) q_{\text{move},j}(\bar{y}, y^*)}.$$

With probability

$$\alpha_{\text{shift},i}(\bar{y}, y^*) = \min\{1, r_{\text{move},i}(\bar{y}, y^*)\},$$

y^* replaces y_j and $X_{i+1} = (X_i \setminus y_j) \cup y^*$, and with the complementary probability y_j is retained so that $X_{i+1} = X_i$.

- **Birth-Death Step:** Given a configuration X_i , the birth-death step either adds or removes a point, or makes no change. Let x be the current state of the Markov chain. Define $N_f = \{x \subset \mathcal{M} : n_x(\mathcal{M}) < \infty\}$, then the birth probability is $p : N_f \mapsto \mathbb{R}$, i.e. $p(x)$ denotes the probability of having a birth given the current state of the chain x . Define $q_{\text{birth}}(x, \cdot) : \mathcal{M} \mapsto \mathbb{R}_+$ to be the proposal distribution for the location of a birth. If the current state of the chain is $X_i = x \in N_f$, a birth step has possibilities $X_{i+1} = x$ or $X_{i+1} = x \cup y^*$ where y^* is sampled from $q_{\text{birth}}(x, \cdot)$. The death probability is $1 - p(x)$, under which, if $x = \emptyset$ nothing happens, while if $x \neq \emptyset$ then a point in x is selected for removal using the discrete probability $q_{\text{death}}(x, \cdot)$. Thus if a death step is chosen and the current state is not the empty set then the two possibilities are $X_{i+1} = X_i$ or $X_{i+1} = X_i \setminus y_j$ if y_j is the point chosen for removal.

The Hasting's ratio for a birth step is

$$r_{\text{birth}}(X_i, y^*) = \frac{f(X_i \cup y^*)\{1 - p(X_i \cup y^*)\}q_{\text{death}}(X_i \cup y^*, y^*)}{f(X_i)p(X_i)q_{\text{birth}}(X_i, y^*)},$$

and so the acceptance ratio for a birth step is $\alpha_{\text{birth}}(X_i, y^*) = \min\{1, r_{\text{birth}}(X_i, y^*)\}$. Thus if a birth step is initiated then $X_{i+1} = X_i \cup y^*$ with probability $\alpha_{\text{birth}}(X_i, y^*)$, otherwise $X_{i+1} = X_i$.

The Hasting's ratio for a death step is

$$r_{\text{death}}(X_i, y^*) = \frac{f(X_i \setminus y_j)p(X_i \setminus y_j)q_{\text{birth}}(X_i \setminus y_j, y_j)}{f(X_i)\{1 - p(X_i)\}q_{\text{death}}(X_i, y_j)},$$

and so the acceptance ratio for a death step is $\alpha_{\text{death}}(X_i, y^*) = \min\{1, r_{\text{death}}(X_i, y^*)\}$. Thus if a death step is initiated then $X_{i+1} = X_i \setminus y_j$ with probability $\alpha_{\text{death}}(X_i, y^*)$, otherwise $X_{i+1} = X_i$ where y_j is drawn from $q_{\text{death}}(x, \cdot)$.

In order to implement the SBDMMH an initial configuration X_0 is needed. The implementation in the paper simulates X_0 as a homogeneous Poisson process with intensity β , where β is as in (S10). Such an initialization is chosen since $E_{X_0}[N_{X_0}(\mathcal{M})] \geq E_X[N_X(\mathcal{M})]$. To show this let X be the Strauss process of interest. Since $\alpha > 0$,

$$E_{X_0}\{N_{X_0}(\mathcal{M})\} = \beta \text{Vol}(\mathcal{M}) = \int_{\mathcal{M}} \beta d\text{Vol}(x) \geq \int_{\mathcal{M}} \beta \exp(-2\alpha) d\text{Vol}(x)$$

and since

$$\begin{aligned} E_X \left(\exp \left[-2\alpha \sum_{y \in X} 1\{d_g(x, y) \leq R\} \right] \right) &\leq \exp(-2\alpha) \\ E_{X_0}\{N_{X_0}(\mathcal{M})\} &\geq \int_{\mathcal{M}} E_X \left(\beta \exp \left[-2\alpha \sum_{y \in X} 1\{d_g(x, y) \leq R\} \right] \right) d\text{Vol}(x) \\ &= \int_{\mathcal{M}} E_X \{ \lambda(X, x) \} d\text{Vol}(x) = \int_{\mathcal{M}} \rho(x) d\text{Vol}(x) = E_X \{ N_X(\mathcal{M}) \}. \end{aligned}$$

It follows that the expected number of points of X_0 is greater than X and thus a plot of $N_{X_i}(\mathcal{M})$ against the iteration number i helps determine when the Markov chain is close to its stationary

365 distribution. As convergence is approached, the plot shows a steady decrease until the true expected number of points have been reached. If the Markov chain is instead initiated at $X_0 = \emptyset$, a plot of $N_{X_i}(\mathcal{M})$ against the iteration number i shows a steady increase. Such plots were inspected but not included in the paper.

Algorithm S4 gives a the manifold version of SBDMMH used in the paper, where
 370 $q_{\text{move},i}(\bar{y}, x) = q_{\text{birth}}(s, x) = 1/\text{Vol}(\mathcal{M})$, $q_{\text{death}}(x) = 1/n_x(\mathcal{M})$, $\bar{y} = (y_1, \dots, y_n) \in \mathcal{M}^n$, $s \in N_f$, $x \in \mathcal{M}$ and $q = 1/2$, (the probability for deciding between a move step or a birth/death step), and $p(x) = 1/2$ (the probability for deciding between a birth or death step). Algorithm S4 was used to simulate the pseudo-homogeneous Strauss processes, and inhomogeneity was introduced by applying independent thinning, as discussed above. The SBDMMH algorithm
 375 was run with a burn-in sample of 10^6 iterations, and the chain was subsampled every 1000 iterations thereafter.

Algorithm S4. Simulation of a Strauss process on an ellipsoid

Require: Parameters of Strauss process (β, α, R) , $0 \leq q < 1$,
 probability of birth or death step $p(x)$, birth proposal distribution $q_{\text{birth}}(x, \cdot)$,
 death proposal distribution $q_{\text{death}}(x, \cdot)$, shift proposal distribution $q_{\text{move},i}(\bar{y}, \cdot)$,
 and semi-major axis lengths (a, b, c)
 Initialise X_0 as either a homogenous Poisson process with intensity β using
 Algorithm S1 or set as \emptyset .
 Set $i = 0$
 For $i = 1, 2, \dots$
 Set $n = N_{X_i}(\mathcal{M})$ and \bar{y} to be a specific configuration of X_i
 Sample $R_1, R_2, R_3 \sim U(0, 1)$
 If $R_1 \leq q$ (This is a move step)
 Draw $J \sim \text{Uniform}(\{1, \dots, n\})$, and conditioned on $J = j$
 sample $y^* \sim q_{\text{move},j}(\bar{y}, \cdot)$
 If $R_3 \leq \alpha_{\text{move}}(\bar{y}, y^*)$
 Set $X_{i+1} = (y_1, \dots, y_{j-1}, y^*, y_{j+1}, \dots, y_n)$
 Else
 Set $X_{i+1} = \bar{y}$
 End If
 Else
 If $R_2 \leq p(x)$ (This is a birth step)
 Sample $y^* \sim q_{\text{birth}}(X_i, \cdot)$
 If $R_3 \leq \alpha_{\text{birth}}(X_i, y^*)$
 Set $X_{i+1} = X_i \cup y^*$
 Else
 Set $X_{i+1} = X_i$
 End If
 Else (This is a death step)
 Sample $y^* \sim q_{\text{death}}(x, \cdot)$
 If $R_3 \leq \alpha_{\text{death}}(X_i, y^*)$
 Set $X_{i+1} = X_i \setminus y^*$
 Else
 Set $X_{i+1} = X_i$
 End If

End If
 End If
 End For

REFERENCES

- BADDELEY, A., RUBAK, E. & TURNER, R. (2015). *Spatial Point Patterns: Methodology and Applications with R*. CRC press. 380
- CARMO, M. P. D. (1992). *Riemannian Geometry*. Birkhäuser.
- CRONIE, O. & VAN LIESHOUT, M. N. (2018). A non-model-based approach to bandwidth selection for kernel estimators of spatial intensity functions. *Biometrika* **105**, 455–462.
- CUCALA, L. (2007). *Espacements bidimensionnels et données entachées d'erreurs dans l'analyse des processus ponctuels spatiaux*. Ph.D. thesis, Université des Sciences Sociales - Toulouse I, 385
- JENSEN, E. B. V. & NIELSEN, L. S. (2001). A review on inhomogeneous Markov point processes. *Lecture Notes-Monograph Series*, 297–318.
- KOPYTOV, N. P. & MITYUSHOV, E. A. (2013). The method for uniform distribution of points on surfaces in multi-dimensional Euclidean space. <http://www.intellectualarchive.com/?link=item&id=1170>, [Accessed: 15 January 2023]. 390
- LEE, J. M. (2003). *Introduction to Smooth Manifolds*. Springer-Verlag, New York.
- MØLLER, J. & WAAGEPETERSEN, R. P. (1998). Log Gaussian Cox processes. *Scand. J. Stat.* **25**, 451–482.
- MØLLER, J. & WAAGEPETERSEN, R. P. (2004). *Statistical Inference and Simulation for Spatial Point Processes*. Florida: CRC Press.
- NASH, J. (1956). The imbedding problem for Riemannian manifolds. *Ann. Math.* **63**, 20–63. 395
- SCHLATHER, M., MALINOWSKI, A., OESTING, M., BOECKER, D., STROKORB, K., ENGELKE, S., MARTINI, J., BALLANI, F., MOREVA, O., AUDEL, J., MENCK, P. J., GROSS, S., OBER, U., RIBEIRO, P., RIPLEY, B. D., SINGLETON, R., PFAFF, B. & R CORE TEAM (2020). *RandomFields: Simulation and Analysis of Random Fields*. R package version 3.3.8.

[Received on 2 January 2017. Editorial decision on 1 April 2017]

Table S1. *Performance of kernel intensity estimators*

Manifold	LGCP model	Process parameters	Expected number of events	$\{\hat{E}(\ \hat{\rho} - \rho\ ^2 / \ \rho\ ^2)\}^{1/2}$	
				CV	NP
\mathcal{E}_1	LGCP1	$(\xi, \sigma^2, \gamma^2) = (10, 2 \log(2), 1/10)$	20.00	0.924	0.635
\mathcal{E}_1	LGCP1	$(\xi, \sigma^2, \gamma^2) = (50, 2 \log(2), 1/10)$	100.0	0.988	0.586
\mathcal{E}_1	LGCP1	$(\xi, \sigma^2, \gamma^2) = (10, 2 \log(5), 1/10)$	50.00	2.211	0.916
\mathcal{E}_1	LGCP1	$(\xi, \sigma^2, \gamma^2) = (50, 2 \log(5), 1/10)$	250.0	2.771	0.904
\mathcal{E}_1	LGCP1	$(\xi, \sigma^2, \gamma^2) = (10, 2 \log(2), 1/50)$	20.00	0.484	0.427
\mathcal{E}_1	LGCP1	$(\xi, \sigma^2, \gamma^2) = (50, 2 \log(2), 1/50)$	100.0	0.496	0.310
\mathcal{E}_1	LGCP1	$(\xi, \sigma^2, \gamma^2) = (10, 2 \log(5), 1/50)$	50.00	1.025	0.465
\mathcal{E}_1	LGCP1	$(\xi, \sigma^2, \gamma^2) = (50, 2 \log(5), 1/50)$	250.0	1.770	0.465
\mathcal{E}_1	LGCP2	$(\sigma^2, \gamma^2) = (2 \log(2), 1/10)$	122.7	1.015	0.568
\mathcal{E}_1	LGCP2	$(\sigma^2, \gamma^2) = (2 \log(2), 1/50)$	122.7	0.517	0.287
\mathcal{E}_1	LGCP2	$(\sigma^2, \gamma^2) = (2 \log(5), 1/10)$	306.8	2.222	0.753
\mathcal{E}_1	LGCP2	$(\sigma^2, \gamma^2) = (2 \log(5), 1/50)$	306.8	2.281	0.460
\mathcal{E}_1	LGCP3	$(\sigma^2, \gamma^2) = (2 \log(2), 1/10)$	68.86	1.056	0.836
\mathcal{E}_1	LGCP3	$(\sigma^2, \gamma^2) = (2 \log(2), 1/50)$	68.86	0.683	0.761
\mathcal{E}_1	LGCP3	$(\sigma^2, \gamma^2) = (2 \log(5), 1/10)$	172.2	2.424	0.977
\mathcal{E}_1	LGCP3	$(\sigma^2, \gamma^2) = (2 \log(5), 1/50)$	172.2	1.652	0.793
\mathcal{E}_2	LGCP1	$(\xi, \sigma^2, \gamma^2) = (10, 2 \log(2), 1/10)$	20.00	1.083	0.692
\mathcal{E}_2	LGCP1	$(\xi, \sigma^2, \gamma^2) = (50, 2 \log(2), 1/10)$	100.0	1.284	0.672
\mathcal{E}_2	LGCP1	$(\xi, \sigma^2, \gamma^2) = (10, 2 \log(5), 1/10)$	50.00	2.890	1.200
\mathcal{E}_2	LGCP1	$(\xi, \sigma^2, \gamma^2) = (50, 2 \log(5), 1/10)$	250.0	3.475	1.180
\mathcal{E}_2	LGCP1	$(\xi, \sigma^2, \gamma^2) = (10, 2 \log(2), 1/50)$	20.00	0.537	0.414
\mathcal{E}_2	LGCP1	$(\xi, \sigma^2, \gamma^2) = (50, 2 \log(2), 1/50)$	100.0	0.532	0.344
\mathcal{E}_2	LGCP1	$(\xi, \sigma^2, \gamma^2) = (10, 2 \log(5), 1/50)$	50.00	0.829	0.427
\mathcal{E}_2	LGCP1	$(\xi, \sigma^2, \gamma^2) = (50, 2 \log(5), 1/50)$	250.0	1.373	0.402
\mathcal{E}_2	LGCP2	$(\sigma^2, \gamma^2) = (2 \log(2), 1/10)$	118.3	1.135	0.625
\mathcal{E}_2	LGCP2	$(\sigma^2, \gamma^2) = (2 \log(2), 1/50)$	118.3	0.541	0.293
\mathcal{E}_2	LGCP2	$(\sigma^2, \gamma^2) = (2 \log(5), 1/10)$	295.7	3.519	1.170
\mathcal{E}_2	LGCP2	$(\sigma^2, \gamma^2) = (2 \log(5), 1/50)$	295.7	1.411	0.370
\mathcal{E}_2	LGCP3	$(\sigma^2, \gamma^2) = (2 \log(2), 1/10)$	80.06	2.265	1.026
\mathcal{E}_2	LGCP3	$(\sigma^2, \gamma^2) = (2 \log(2), 1/50)$	80.06	0.711	0.736
\mathcal{E}_2	LGCP3	$(\sigma^2, \gamma^2) = (2 \log(5), 1/10)$	200.1	2.426	1.052
\mathcal{E}_2	LGCP3	$(\sigma^2, \gamma^2) = (2 \log(5), 1/50)$	200.1	1.509	0.777
\mathcal{E}_3	LGCP1	$(\xi, \sigma^2, \gamma^2) = (10, 2 \log(2), 1/10)$	20.00	0.959	0.722
\mathcal{E}_3	LGCP1	$(\xi, \sigma^2, \gamma^2) = (50, 2 \log(2), 1/10)$	100.0	0.966	0.650
\mathcal{E}_3	LGCP1	$(\xi, \sigma^2, \gamma^2) = (10, 2 \log(5), 1/10)$	50.00	1.852	0.960
\mathcal{E}_3	LGCP1	$(\xi, \sigma^2, \gamma^2) = (50, 2 \log(5), 1/10)$	250.0	2.076	0.937
\mathcal{E}_3	LGCP1	$(\xi, \sigma^2, \gamma^2) = (10, 2 \log(2), 1/50)$	20.00	0.573	0.464
\mathcal{E}_3	LGCP1	$(\xi, \sigma^2, \gamma^2) = (50, 2 \log(2), 1/50)$	100.0	0.600	0.363
\mathcal{E}_3	LGCP1	$(\xi, \sigma^2, \gamma^2) = (10, 2 \log(5), 1/50)$	50.00	1.004	0.487
\mathcal{E}_3	LGCP1	$(\xi, \sigma^2, \gamma^2) = (50, 2 \log(5), 1/50)$	250.0	1.408	0.447
\mathcal{E}_3	LGCP2	$(\sigma^2, \gamma^2) = (2 \log(2), 1/10)$	114.4	0.986	0.647
\mathcal{E}_3	LGCP2	$(\sigma^2, \gamma^2) = (2 \log(2), 1/50)$	114.4	0.579	0.337
\mathcal{E}_3	LGCP2	$(\sigma^2, \gamma^2) = (2 \log(5), 1/10)$	286.1	2.320	0.976
\mathcal{E}_3	LGCP2	$(\sigma^2, \gamma^2) = (2 \log(5), 1/50)$	286.1	1.461	0.462
\mathcal{E}_3	LGCP3	$(\sigma^2, \gamma^2) = (2 \log(2), 1/10)$	103.5	0.962	0.799
\mathcal{E}_3	LGCP3	$(\sigma^2, \gamma^2) = (2 \log(2), 1/50)$	103.5	0.704	0.683
\mathcal{E}_3	LGCP3	$(\sigma^2, \gamma^2) = (2 \log(5), 1/10)$	258.8	1.614	0.929
\mathcal{E}_3	LGCP3	$(\sigma^2, \gamma^2) = (2 \log(5), 1/50)$	258.8	1.646	0.753

Square root of the mean integrated squared error from a 3^3 factorial experiment. The mean is taken over 100 Monte Carlo replicates.

Table S2. Performance of kernel intensity estimators

Manifold	SP model	Process parameters	Expected number of events	$\{\hat{E}(\ \hat{\rho} - \rho\ ^2 / \ \rho\ ^2)\}^{1/2}$	
				CV	NP
\mathcal{E}_1	SP1	$(\beta, \alpha, R) = (400, 0.01, 0.1)$	326.9	0.113	0.175
\mathcal{E}_1	SP1	$(\beta, \alpha, R) = (400, 0.1, 0.1)$	158.8	0.105	0.178
\mathcal{E}_1	SP1	$(\beta, \alpha, R) = (400, 0.01, 0.2)$	230.5	0.115	0.172
\mathcal{E}_1	SP1	$(\beta, \alpha, R) = (400, 0.1, 0.2)$	74.07	0.166	0.213
\mathcal{E}_1	SP2	$(\beta, \alpha, R) = (400, 0.01, 0.1)$	177.2	0.177	0.165
\mathcal{E}_1	SP2	$(\beta, \alpha, R) = (400, 0.1, 0.1)$	85.94	0.198	0.195
\mathcal{E}_1	SP2	$(\beta, \alpha, R) = (400, 0.01, 0.2)$	124.8	0.181	0.171
\mathcal{E}_1	SP2	$(\beta, \alpha, R) = (400, 0.1, 0.2)$	40.12	0.286	0.272
\mathcal{E}_1	SP3	$(\beta, \alpha, R) = (400, 0.01, 0.1)$	147.7	0.355	0.594
\mathcal{E}_1	SP3	$(\beta, \alpha, R) = (400, 0.1, 0.1)$	71.85	0.389	0.603
\mathcal{E}_1	SP3	$(\beta, \alpha, R) = (400, 0.01, 0.2)$	104.2	0.385	0.596
\mathcal{E}_1	SP3	$(\beta, \alpha, R) = (400, 0.1, 0.2)$	33.46	0.489	0.628
\mathcal{E}_2	SP1	$(\beta, \alpha, R) = (400, 0.01, 0.1)$	20.00	0.109	0.167
\mathcal{E}_2	SP1	$(\beta, \alpha, R) = (400, 0.1, 0.1)$	100.0	0.109	0.179
\mathcal{E}_2	SP1	$(\beta, \alpha, R) = (400, 0.01, 0.2)$	50.00	0.123	0.178
\mathcal{E}_2	SP1	$(\beta, \alpha, R) = (400, 0.1, 0.2)$	250.0	0.164	0.214
\mathcal{E}_2	SP2	$(\beta, \alpha, R) = (400, 0.01, 0.1)$	118.3	0.173	0.166
\mathcal{E}_2	SP2	$(\beta, \alpha, R) = (400, 0.1, 0.1)$	118.3	0.203	0.201
\mathcal{E}_2	SP2	$(\beta, \alpha, R) = (400, 0.01, 0.2)$	295.7	0.204	0.192
\mathcal{E}_2	SP2	$(\beta, \alpha, R) = (400, 0.1, 0.2)$	295.7	0.280	0.272
\mathcal{E}_2	SP3	$(\beta, \alpha, R) = (400, 0.01, 0.1)$	80.06	0.278	0.406
\mathcal{E}_2	SP3	$(\beta, \alpha, R) = (400, 0.1, 0.1)$	80.06	0.321	0.424
\mathcal{E}_2	SP3	$(\beta, \alpha, R) = (400, 0.01, 0.2)$	200.1	0.315	0.415
\mathcal{E}_2	SP3	$(\beta, \alpha, R) = (400, 0.1, 0.2)$	200.1	0.416	0.468
\mathcal{E}_3	SP1	$(\beta, \alpha, R) = (400, 0.01, 0.1)$	20.00	0.121	0.172
\mathcal{E}_3	SP1	$(\beta, \alpha, R) = (400, 0.1, 0.1)$	100.0	0.123	0.180
\mathcal{E}_3	SP1	$(\beta, \alpha, R) = (400, 0.01, 0.2)$	50.00	0.131	0.177
\mathcal{E}_3	SP1	$(\beta, \alpha, R) = (400, 0.1, 0.2)$	250.0	0.197	0.227
\mathcal{E}_3	SP2	$(\beta, \alpha, R) = (400, 0.01, 0.1)$	114.4	0.184	0.181
\mathcal{E}_3	SP2	$(\beta, \alpha, R) = (400, 0.1, 0.1)$	114.4	0.206	0.204
\mathcal{E}_3	SP2	$(\beta, \alpha, R) = (400, 0.01, 0.2)$	286.1	0.200	0.194
\mathcal{E}_3	SP2	$(\beta, \alpha, R) = (400, 0.1, 0.2)$	286.1	0.280	0.272
\mathcal{E}_3	SP3	$(\beta, \alpha, R) = (400, 0.01, 0.1)$	103.5	0.248	0.251
\mathcal{E}_3	SP3	$(\beta, \alpha, R) = (400, 0.1, 0.1)$	103.5	0.290	0.280
\mathcal{E}_3	SP3	$(\beta, \alpha, R) = (400, 0.01, 0.2)$	258.8	0.267	0.262
\mathcal{E}_3	SP3	$(\beta, \alpha, R) = (400, 0.1, 0.2)$	258.8	0.355	0.340

Square root of the mean integrated squared error from a 3^3 factorial experiment. The mean is taken over 100 Monte Carlo replicates.