MOX-Report No. 41/2022

A roughness penalty approach to estimate densities over two-dimensional manifolds

Arnone, A.; Ferraccioli, F.; Pigolotti, C.; Sangalli, L.M.

MOX, Dipartimento di Matematica
Politecnico di Milano, Via Bonardi 9 - 20133 Milano (Italy)

mox-dmat@polimi.it http://mox.polimi.it

To appear in Computational Statistics and Data Analysis, DOI: 10.1016/j.csda.2022.107527
Publisher version available at: https://doi.org/10.1016/j.csda.2022.107527

A roughness penalty approach to estimate densities over two-dimensional manifolds

Eleonora Arnone[a], Federico Ferraccioli[b], Clara Pigolotti[a], Laura M. Sangalli[a,*]

[a] *MOX - Department of Mathematics, Politecnico di Milano, Piazza Leonardo da Vinci 32, Milano, 20133, Italy*
[b] *Department of Statistical Sciences, University of Padova, Via 8 Febbraio, 2, Padova, 35122, Italy*

Abstract

An innovative nonparametric method for density estimation over general two-dimensional Riemannian manifolds is proposed. The method follows a functional data analysis approach, combining maximum likelihood estimation with a roughness penalty that involves a differential operator appropriately defined over the manifold domain, thus controlling the smoothness of the estimate. The proposed method can accurately handle point pattern data over complicated curved domains. Moreover, it is able to capture complex multimodal signals, with strongly localized and highly skewed modes, with varying directions and intensity of anisotropy. The estimation procedure exploits a discretization in finite element bases, enabling great flexibility on the spatial domain. The method is tested through simulation studies, showing the strengths of the proposed approach. Finally, the density estimation method is illustrated with an application to the distribution of earthquakes in the world.

Keywords: functional data analysis, differential regularization, finite element basis
2000 MSC: 62G07, 62G20, 58J65

---

1. Introduction

This work introduces an innovative method for density estimation over general two-dimensional Riemannian manifolds. The method follows a functional data analysis approach, combining maximum likelihood estimation with a roughness penalty, that involves a differential operator appropriately defined over the manifold domain, thus controlling the smoothness of the estimate. The proposed technique can deal with data scattered over two-dimensional curved domains with complicated geometries. Moreover, it accurately estimates complex multi-modal densities, with highly skewed modes or ridges. Figure 1 shows an example of a particular setting that can be efficiently treated with the proposed method. The points are the locations of earthquakes, of magnitude greater than 4.5, occurred across the world in the period from 1st March

---

[*] Supplementary material providing the code and data for the simulation studies is provided online.
[*] Corresponding author
Email address: `laura.sangalli@polimi.it` (Laura M. Sangalli)

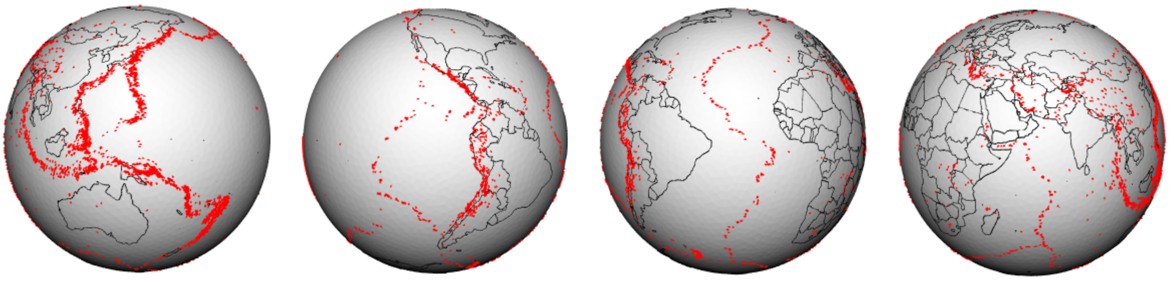

Figure 1: This figure shows the locations of earthquakes in the world which occur in the period from 1^{st} March 2019 to 1^{st} March 2020 with a magnitude greater then 4.5. The peculiarity of this dataset is that points are mainly on stretched regions and therefore the signal turns out to be very skewed and anisotropic.

2019 to 1^{st} March 2020. Most data are are observed along elongated narrow regions, that are the faults on the earth's crust.

Density estimation over parametric manifolds has attracted lots of attention. The special case of density estimation over a spherical domain has been extensively studied. Many examples can for instance be found in the literature on directional and axial statistics; see, e.g., Mardia [28], Watson [41] and Jupp and Mardia [23]. In particular, Hall et al. [20] and Fisher et al. [13] study kernel methods for axial or directional data. Axial or directional data also offers problems of density estimation over cylinders and tori. For instance, the study of temperatures and wind directions, that are a combination of scalar data (the temperature) and circular data (the wind direction) leads to problems of density estimation over cylinders [see, e.g., 22]. The study of the joint distribution of multiple circular data, such as joint distributions of angles, leads instead to densities over tori [see, e.g., 12, 8].

Fewer methods are instead available for density estimation on manifolds of generic shapes. Some theoretical studies concerning point data scattered over general compact manifolds without boundary are done in Hendriks [21], in which a method based on estimation with Fouries series is presented, and in Pelletier [33] and in Kim and Park [25], in which authors adapt kernel density estimation to the non-Euclidean setting. These works establish the good asymptotic properties of the proposed estimators; however their implementation for general manifolds is not straightforward and is not discussed. Berry and Sauer [4] instead presents a boundary correction for the kernel density estimator and provides an algorithm that works with manifolds with unknown boundary. Finally, the log-Gaussian Cox Processes proposed in Simpson et al. [39] can handle point data over domains with irregular shapes, including possibly curved domains. In a similar spirit, some recent works address density and point processes estimation on complicated domains such as networks [see, for example, 29, 36, 31, 30].

Here we propose a flexible density estimation method, based on a nonparametric maximum likelihood approach with a roughness penalty. Similar approaches have been considered for density estimation over one-dimensional domains by Silverman [38] and Good and Gaskins [17], over regular planar two-dimensional domains by Gu and Qiu [19] and Gu [18], and over bounded and irregularly shaped planar domains by

Ferraccioli et al. [11]. In particular, we here extend the method by Ferraccioli et al. [11] to the case of curved two-dimensional domains, considering both bounded manifolds and closed manifolds (i.e., manifolds that do not posses a boundary, as for instance spheres), possibly with non-trivial geometries. The method permits to handle point patterns arising over general two-dimensional Riemannian manifolds, as for instance data observed over complicated earth orographies or over life tissues with complex anatomies, such as the brain surface.

The extension from the planar domains considered in Ferraccioli et al. [11] to the general two-dimensional Riemannian manifolds here considered requires the definition of appropriate partial differential operators over the manifold, and suitable Sobolev spaces of functions on the manifold [see, e.g., 10]. In particular, the roughness penalty we here consider involves the Laplace-Beltrami operator defined on the manifold domain. This partial differential operator provides a measure of the local curvature of a function defined over a surface domain, enabling us to appropriately control the smoothness of the estimate. We demonstrate the good asymptotic properties of the method, proving the consistency of the proposed estimator. We hence describe a computational method to solve the estimation problem. Partial differential problems over Riemmanian manifolds requires appropriate numerical techniques [see, e.g., 10]. Specifically, we here resort to surface finite elements, which enable us to represent functions defined over triangulated surfaces. Non-planar triangular meshes enable us to appropriately represent curved domains with general and possibly complicated shapes, as shown in Figure 2. Moreover, these unstructured meshes and the localized support of finite element bases make the proposed technique able to accurately capture highly localized features and lower dimensional structures such as ridges, as demonstrated by simulation studies. Comparison to kernel density estimation over spherical domains highlights the superiority of the proposed method.

The method proposed is part of a class of models for the analysis of functional and of spatial data over complex multidimensional domains, that relies on statistical approaches with regularizing terms involving Partial Differential Equations (PDEs) [see, e.g., 37], and it is implemented in the `R` package `fdaPDE` [27, 35].

The article is organized as follows. In Section 2 we introduce the proposed model in its infinite-dimensional form. In Section 3 we derive its asymptotic properties. Section 4 details the discretization of the estimation problems through finite elements. Section 5 reports some simulation studies, which compare the proposed method to kernel density estimation on the sphere. Section 6 shows the application to the study of the distribution of earthquakes across the world, shown in Figure 1. Finally, Section 7 outlines some possible future developments of the proposed method.

2. Density estimation over two-dimensional manifolds

In this section we first introduce some concepts of differential geometry, giving the definition of the Laplace-Beltrami operator; we then present the model for density estimation over a two-dimensional manifold.

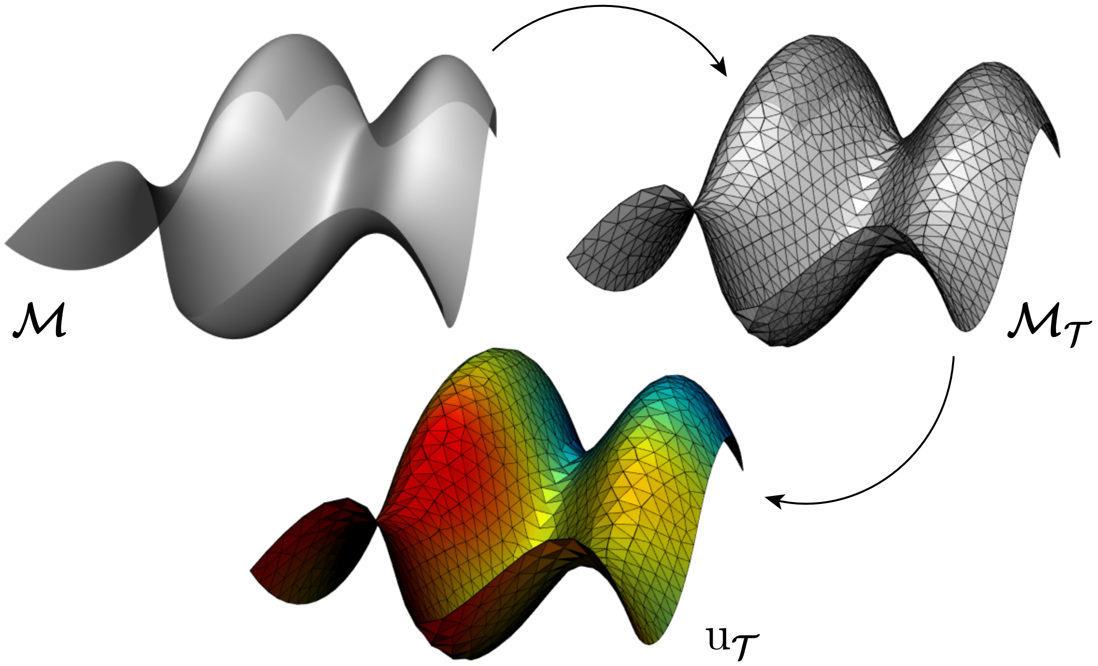

Figure 2: The discretization procedure. On the top left the true manifold; on the top right the domain approximated by a triangulation; on the bottom an example of a finite element function on the triangulated domain.

2.1. Review of some basic concepts from differential geometry

A manifold is a topological space that locally resembles Euclidean space, but globally can have different properties. More precisely, each point x of a d-dimensional manifold has a neighborhood that is homeomorphic to the Euclidean space of dimension d. In this more precise terminology, a manifold is referred to as a d-manifold.

In this work we consider two-dimensional Riemannian manifolds. A Riemannian manifold $\mathcal{M}$ is a differentiable manifold equipped with an inner product g_x defined on the tangent space $\mathcal{T}_x\mathcal{M}$ of each point $x \in \mathcal{M}$. Let $\varphi : \mathcal{U} \subset \mathbb{R}^2 \to \mathcal{V} \cap \mathcal{M}$, with $\mathcal{V}$ an open set in $\mathbb{R}^3$, be the bijective and smooth function that locally parametrizes the point $x \in \mathcal{M}$. Let $\theta \in \mathcal{U} \subset \mathbb{R}^2$ be such that $\theta = \varphi^{-1}(x)$. The couple $(\mathcal{V}, \varphi)$ is called local chart, and

$$\left\{ \frac{\partial \varphi}{\partial \theta_1}(\theta), \frac{\partial \varphi}{\partial \theta_2}(\theta) \right\}$$

defines a basis for the local tangent space $\mathcal{T}_x\mathcal{M}$. Finally, the collection of charts $\{(\mathcal{V}_i, \varphi_i)\}_{i\in I}$ such that $\bigcup_{i\in I}(\mathcal{V}_i \cap \mathcal{M}) = \mathcal{M}$ is called atlas. Figure 3 shows an example of a two-dimensional manifold with a local chart.

The scalar product g_x can be represented as the matrix $G = (g_{ij})_{i,j=1,2}$ such that

$$g_x(v, w) = \sum_{i,j=1}^{2} g_{ij} v_i w_j$$

for all $v = \sum_{i=1}^{2} v_i \frac{\partial \varphi}{\partial \theta_i}(\theta)$ and $w = \sum_{i=1}^{2} w_i \frac{\partial \varphi}{\partial \theta_i}(\theta)$, where v_i and w_i are the coordinates of v and w on the local chart. Considering the scalar product induced by the Euclidean

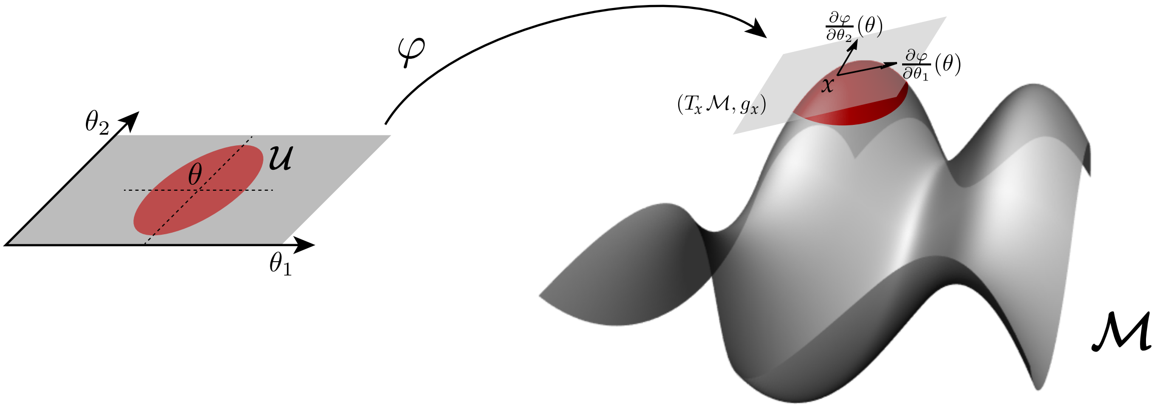

Figure 3: A two-dimensional Riemannian manifold with a graphical description of the tangent space and the local chart. $\mathcal{T}_x\mathcal{M}$ is the tangent space of $\mathcal{M}$ in x, with the associated scalar product g_x, while φ is the bijection from a neighborhood $\mathcal{U}$ of θ to a neighborhood $\mathcal{V} \cap \mathcal{T}_x\mathcal{M}$ of x.

embedding space $\mathbb{R}^3$, we have

$$g_{ij}(\theta) = \frac{\partial\varphi}{\partial\theta_i}(\theta) \cdot \frac{\partial\varphi}{\partial\theta_j}(\theta).$$

Let $G^{-1} = (g^{ij})_{i,j=1,2}$ be the inverse of the matrix G and $\det(G)$ its determinant. Let now $f : \mathcal{M} \longrightarrow \mathbb{R}$ be a real valued and twice differentiable function on the manifold $\mathcal{M}$. Let $\tilde{f} = f \circ \varphi$, then the gradient $\nabla_{\mathcal{M}}$ is defined as

$$\nabla_{\mathcal{M}} f(x) = \sum_{i,j=1}^{2} g^{ij}(\theta) \frac{\partial \tilde{f}}{\partial\theta_j}(\theta) \frac{\partial\varphi}{\partial\theta_j}(\theta).$$

In the case of a flat manifold $\mathcal{M}$, the last expression reduces to the expression of the gradient in $\mathbb{R}^2$, that is, $\nabla = (\frac{\partial}{\partial\theta_1}, \frac{\partial}{\partial\theta_2})$. In fact, in this case, φ is the identity map, implying that $\tilde{f} = f$ and $\frac{\partial\varphi}{\partial\theta_j}(\theta) = 1$, and g is the Euclidean metric, so that $g^{ij} = g_{ij} = 1$ if $i = j$ and $g^{ij} = g_{ij} = 0$ if $i \neq j$.

The Laplace-Beltrami operator $\Delta_{\mathcal{M}}$ is hence defined as

$$\Delta_{\mathcal{M}} f(x) = \nabla_{\mathcal{M}} \cdot \nabla_{\mathcal{M}} f(x) = \frac{1}{\sqrt{\det(G)(\theta)}} \sum_{i,j=1}^{2} \frac{\partial}{\partial\theta_j} g^{ij} \sqrt{\det(G)(\theta)} \frac{\partial \tilde{f}}{\partial\theta_j}(\theta).$$

The Laplace-Beltrami operator $\Delta_{\mathcal{M}}$ is a measure of the local curvature of the function. If $\mathcal{M}$ is a planar domain, thanks to the simplifications highlighted above, this operator coincides with the standard Laplacian Δ, defined on $\mathbb{R}^d$ as $\sum_{i=1}^{d} \frac{\partial^2}{\partial^2\theta_i}$.

Finally, we define the integral over $\mathcal{M}$. Let $\tilde{f} : \mathcal{U} \longrightarrow \mathbb{R}$ be an integrable function over $\mathcal{U}$. Then $f = \tilde{f} \circ \varphi^{-1}$ is an integrable function on $\mathcal{M}$ and the integral is defined as

$$\int_{\mathcal{M}} f = \int_{\mathcal{U}} \tilde{f}\sqrt{g}.$$

2.2. *Model description*

Let $x_1, \dots, x_n$ be independent observations drawn from a distribution with density f on a two-dimensional Riemannian manifold $\mathcal{M}$. The problem that we consider is to estimate f by a nonparametric model. The idea would be to maximize the likelihood function $L(f|x_1, \dots, x_n) = \prod_{i=1}^{n} f(x_i)$ in a proper functional space. However, without further assumptions, the maximum likelihood estimator for f is not well defined. In fact, the likelihood $L(f|x_1, \dots, x_n)$ is unbounded from above: its maximization over a rich enough functional space would thus return the trivial solution of sum of delta functions centered at the observations. We hence consider a regularization of the estimation problem, in order to restrict the class of possible solutions and avoid degenerated solutions. In particular, we propose to estimate f by minimizing the penalized negative log likelihood

$$-\frac{1}{n}\sum_{i=1}^{n}\log f(x_i) + \lambda \int_{\mathcal{M}} (\Delta_{\mathcal{M}} \log f)^2, \tag{1}$$

subject to:

1. $f \geq 0$ in $\mathcal{M}$,
2. $\int_{\mathcal{M}} f = 1$.

As explained in the previous section, the Laplace-Beltrami operator $\Delta_{\mathcal{M}}$ measures the local curvature of a function; therefore, the regularization in (1) controls the roughness of the estimate. The smoothing parameter $\lambda > 0$ balances the adherence to the data and the smoothness of the estimated density.

Following the approach in Silverman [38], Ferraccioli et al. [11], we consider the logarithm transformation $u = \log f$ and, instead of performing the constrained minimization problem above, we carry out the unconstrained minimization of

$$L(u) = -\frac{1}{n}\sum_{i=1}^{n} u(x_i) + \int_{\mathcal{M}} e^u + \lambda \int_{\mathcal{M}} (\Delta_{\mathcal{M}} u)^2. \tag{2}$$

The first and the last term in (2) come directly from (1), while the second term in (2) ensures that f integrates to one. The proof of the equivalence between the unconstrained problem (2) and the constrained problem (1) is reported in Appendix A. If $\mathcal{M}$ is a planar domain, the Laplace-Beltrami reduces to the standard Laplacian, and the estimation problem (2) coincides with the one considered in Ferraccioli et al. [11].

Let $L^2(\mathcal{M})$ denote the space of square integrable functions over $\mathcal{M}$ equipped with the norm $\|u\|^2_{L^2(\mathcal{M})} = \int_{\mathcal{M}} u^2$, and let $H^k(\mathcal{M})$ denote the Sobolev space

$$H^k(\mathcal{M}) := \left\{ u \in L^2(\mathcal{M}) \quad \text{s.t.} \quad D^\alpha u \in L^2(\mathcal{M}) \quad \forall |\alpha| \leq k \right\},$$

equipped with the norm $\|u\|^2_{H^k(\mathcal{M})} := \sum_{|\alpha|} \|D^\alpha u\|^2_{L^2(\mathcal{M})}$ where the symbol $D^\alpha u$ denote the weak derivative of order α [see, e.g., 5, 10, for the general theory of Sobolev spaces and Sobolev spaces on manifolds]. In this work we consider both bounded manifolds and closed manifolds (i.e., manifolds that do not posses a boundary, as for instance spheres). We define the space V as the space of functions in $H^2(\mathcal{M})$ with homogeneous Neumann conditions at the boundary of the domain, that is

$$V = \{u \in H^2(\mathcal{M}) : \frac{\partial u}{\partial \nu} = 0 \text{ on } \partial\mathcal{M}\},$$

where $\partial\mathcal{M}$ denotes the boundary of the manifold, ν denotes the normal unitary vector to the boundary, and $\frac{\partial u}{\partial \gamma}$ is the derivative of u in the normal direction. When $\mathcal{M}$ is a closed manifold, $V = H^2(\mathcal{M})$.

Theorem 1. *The penalized negative log likelihood* $L(u)$ *defined in equation* (2) *has a unique minimizer* $\hat{u}$ *in* V*.*

Proof. The proof is deferred to Appendix B. □

3. Asymptotic properties

We now study the asymptotic properties of the estimator when the number of observations goes to infinity. Let u_0 be the true log density function.

Assumption 1. *The log density* u_0 *is bounded above and below, and is such that* $\int_{\mathcal{M}} (\Delta_{\mathcal{M}} u_0)^2 < \infty$*.*

Assumption 1 guarantees that the weighted $L^2(\mathcal{M})$ norm with the density function $\exp(u_0)$ is equivalent to the standard $L^2(\mathcal{M})$ norm.

Denote with $D_{sKL}(u_1, u_2)$ the symmetrized Kullback–Leibler distance between u_1 and u_2, that is $\mu_{u_1}(u_1 - u_2) + \mu_{u_2}(u_2 - u_1)$, where $\mu_u(h)$ is the mean of $h(X)$ when X has log density u. Denote with $\mathrm{Var}_u(h)$ the variance of $h(X)$ when X has log density u. Following the same approach as in Silverman [38], Gu and Qiu [19], Ferraccioli et al. [11], we introduce u_*, an approximation of $\hat{u}$, which is the minimizer of

$$L_*(u) = -\frac{1}{n}\sum_{i=1}^{n} u(x_i) + 1 + \mu_{u_0}(u) + \frac{1}{2}\mathrm{Var}_{u_0}(u - u_0) + \lambda \int_{\mathcal{M}} (\Delta_{\mathcal{M}} u)^2 .$$

$L_*(u)$ is the quadratic form which has second order contact with the functional L at u_0. We make the following assumption on u_*.

Assumption 2. *For* u *in a convex set* B_0 *around* u_0 *containing* $\hat{u}$ *and* u_**, there exists a positive constant* c *such that* $c\,\mathrm{Var}_{u_0} \leq \mathrm{Var}_u$ *uniformly with respect to* u*.*

Assumption 2 is satisfied whenever the members of B_0 are bounded from above and below. The assumption requires that the functions near u_0 ($\hat{u}$ and u_*, in particular) satisfy the property in Assumption 1.

We are now able to state the consistency of the proposed density estimator.

Theorem 2. *Let* $\mathcal{M}$ *be a two-dimensional compact Riemannian manifold. Under Assumptions 1 and 2, as* $\lambda \to 0$ *and* $n\lambda^{1/2} \to \infty$*, the estimator* $\hat{u}$ *that minimizes* (2) *is consistent and*

$$D_{sKL}(\hat{u}, u_0) = O(n^{-1}\lambda^{-1/2} + \lambda). \tag{3}$$

Proof. The proof is deferred to Appendix C. □

4. Computation of the estimator

The infinite dimensional problem of minimizing functional (2) in the space V does not have an analytical solution. Therefore, it is necessary to compute an approximated solution. Following the standard approach in functional data analysis, we introduce a finite dimensional space $V_{\mathcal{T}}$, we hence approximate the functional (2) in this space, and finally we solve the resulting minimization problem. The rest of the section describes these steps in details.

4.1. Surface finite elements

We start the discretization procedure by approximating the original manifold $\mathcal{M}$ with $\mathcal{M}_{\mathcal{T}}$, that is a discretized version of the domain, composed by triangles. We denote with $\mathcal{T}$ the set of all the triangles τ composing $\mathcal{M}_{\mathcal{T}}$.

Meshes composed by quadrilateral or general polygonal elements can also be employed, but are not considered in this work. We approximate the infinite dimensional functional space V with the space:

$$V_{\mathcal{T}} = \{u_{\mathcal{T}} \in C(\mathcal{M}_{\mathcal{T}}) \quad \text{s.t.} \quad u_{\mathcal{T}}|_{\tau} \ \text{ is linear } \ \forall \tau \in \mathcal{T}\}\,.$$

$V_{\mathcal{T}}$ is the space of linear finite element functions. Polynomial of higher order can also be considered [see 10, for more information on surface finite elements]. In particular, linear and quadratic finite elements are implemented in the `fdaPDE` package.

We now introduce the finite element basis functions $\{\psi_i\}_{i=1,\cdots,K}$ for $V_{\mathcal{T}}$. Denote ξ_k the nodes of the mesh. In the case of linear finite elements, the nodes coincide with the vertices of the triangles. The linear finite elements basis functions $\{\psi_i\}_{i=1,\cdots,K}$ are defined as piecewise linear functions over $\mathcal{M}_{\mathcal{T}}$, such that $\psi_i(\xi_j) = \delta_{ij}$ for $i, j = 1 \ldots K$, where δ_{ij} is the Kronecker delta $\delta_{ij} = 1$ if $i = j$, $\delta_{ij} = 0$ otherwise. Each finite element basis function ψ_i has a small support, since it is zero over all the triangles that do not have ξ_i as a vertex.

Within this setting, any function u which belongs to the infinite dimensional space V can be approximated by a function $u_{\mathcal{T}} \in V_{\mathcal{T}}$, and it can be written as a linear combination of the basis functions:

$$u_{\mathcal{T}}(\mathbf{x}) = \sum_{k=1}^{K} u_k \psi_k(\mathbf{x}),$$

where the coefficients of the linear combination u_k are the values of the function u evaluated on the mesh nodes, that is $u_k = u(\xi_k)$ $\forall k = 1, \ldots, K$. Therefore, any function $u_{\mathcal{T}} \in V_{\mathcal{T}}$ is completely defined by specifying the values of the function at the K mesh nodes. In a more compact way, calling $\boldsymbol{\psi} := (\psi_1, \ldots, \psi_K)^\top$ and $\mathbf{u}_{\mathcal{T}} := (u(\xi_1), \ldots, u(\xi_K))^\top$, the function $u_{\mathcal{T}}(\mathbf{x})$ is

$$u_{\mathcal{T}}(\mathbf{x}) = \mathbf{u}_{\mathcal{T}}^\top \boldsymbol{\psi}(\mathbf{x}) \tag{4}$$

and the problem of finding the function u becomes a problem of finding the coefficients $\mathbf{u}_{\mathcal{T}}$.

Figure 2 is a schematic drawing of this procedure. The top left panel shows a manifold $\mathcal{M}$, the top right panel shows a triangular mesh of $\mathcal{M}$, while the bottom panel displays a function $u_{\mathcal{T}}$ in $V_{\mathcal{T}}$.

We point out that the number of basis functions (i.e., the number of mesh nodes) influences both the accuracy of the estimate and the accuracy of the approximation of the manifold itself. Therefore, if not provided with the data, the mesh should be chosen rich enough to ensure an adequate approximation of the domain geometry and to accurately capture the features of the signal over the domain.

4.2. *Discretization and minimization of the functional*

We now want to compute the functional $L(u)$ defined in equation (2) for functions in $V_{\mathcal{T}}$. More precisely, we compute an approximated version of the functional $L(u)$ that we denote by $L_{\mathcal{T}}(u_{\mathcal{T}})$.

Approximation of $-\frac{1}{n}\sum_{i=1}^{n} u(x_i)$. We introduce the $n \times K$ matrix Ψ which contains the values of the K basis functions $\{\psi_i\}_{i=1,\cdots,K}$ evaluated at the n data,

$$\Psi = \begin{bmatrix} \psi_1(x_1) & \dots & \psi_K(x_1) \\ \vdots & \vdots & \vdots \\ \psi_1(x_n) & \dots & \psi_K(x_n) \end{bmatrix}.$$

This is a sparse matrix thanks to the properties of the finite element bases. With this notation and using (4), the first term of the functional (2) can be expressed in a discretized way as

$$-\frac{1}{n}\sum_{i=1}^{n} u(x_i) \approx -\frac{1}{n}\mathbf{1}^\top \Psi \mathbf{u}_{\mathcal{T}},$$

where $\mathbf{1}$ denotes the n-vector with all entries equal to 1.

Approximation of $\int_{\mathcal{M}} \exp u(x) dx$. For the second term of the functional $L(u)$, we use (4) to obtain the discretized expression:

$$\int_{\mathcal{M}} \exp u(x) dx \approx \int_{\mathcal{M}_{\mathcal{T}}} \exp\left(\boldsymbol{\psi}(x)^\top \mathbf{u}_{\mathcal{T}}\right) dx = \sum_{\tau \in \mathcal{T}} \int_{\tau} \exp\left(\boldsymbol{\psi}(x)^\top \mathbf{u}_{\mathcal{T}}\right) dx.$$

The integrals over the triangles are approximated using a quadrature formula and have the expression

$$c_\tau \mathbf{w}^\top \exp\left(\Psi_\tau \mathbf{u}_{\mathcal{T}}\right),$$

where $\mathbf{w} \in \mathbb{R}^q$ in the vector of quadrature weights, $\Psi_\tau \in \mathbb{R}^{q \times K}$ is the matrix having as entries the evaluations of the basis functions at the q quadrature nodes in the triangle τ, and c_τ is the determinant of the transformation matrix that maps the reference triangle $\{(0,0),(0,1),(1,0)\}$ on the triangle τ.

Approximation of $\lambda \int_{\mathcal{M}} (\Delta_{\mathcal{M}} u)^2$. Let

$$R_0 := \int_{\mathcal{M}_{\mathcal{T}}} (\psi\psi^\top) \qquad \text{and} \qquad R_1 := \int_{\mathcal{M}_{\mathcal{T}}} (\nabla_{\mathcal{M}_{\mathcal{T}}}\psi)(\nabla_{\mathcal{M}_{\mathcal{T}}}\psi)^\top.$$

R_0 is called mass matrix and R_1 is called stiffness matrix. Relying on these matrices, it is possible to introduce a discretization of the regularization term in (2) that does not not involve second order derivatives, and is given by

$$\lambda \mathbf{u}_{\mathcal{T}}^\top R_1 R_0^{-1} R_1 \mathbf{u}_{\mathcal{T}}.$$

See Lila et al. [26] for details on the derivation of this discretization for an analogous estimation problem.

Summarizing, the discrete penalized negative log likelihood is

$$L_{\mathcal{T}}(\mathbf{u}_{\mathcal{T}}) = -\frac{1}{n}\mathbf{1}^\top \Psi \mathbf{u}_{\mathcal{T}} + \sum_{T\in\mathcal{T}} c_T \mathbf{w}^\top \exp\left(\Psi_T \mathbf{u}_{\mathcal{T}}\right) + \lambda \mathbf{u}_{\mathcal{T}}^\top R_1 R_0^{-1} R_1 \mathbf{u}_{\mathcal{T}}, \tag{5}$$

and the estimation problem becomes: find $\mathbf{u}_{\mathcal{T}} \in \mathbb{R}^K$ such that it minimizes $L_{\mathcal{T}}(\mathbf{u}_{\mathcal{T}})$.

Thanks to Lemma 2 in Appendix B, we know that the functional $L_{\mathcal{T}}(\mathbf{u}_{\mathcal{T}})$ is convex and therefore its minimization can be performed with standard optimization algorithms, such as gradient descent or quasi Newton methods. In particular, in the simulation studies and in the application we used a quasi Newton with BFGS direction. Since the functional is convex, the minimization converges regardless of the initialization, therefore a constant initialization (that corresponds to a uniform distribution) can be used. When the observed point pattern suggests a true density far from the uniform distribution, a data-driven initialization can significantly reduce the computational cost. In particular, an initialization such as the one detailed in Section 4.3 of Ferraccioli et al. [11], that relies on a heat-diffusion density estimation computed on the Voronoi tessellation associated with the triangulation, can save up to 40% of computational time, for highly non-uniform densities.

4.2.1. *Selection of the smoothing parameter*

The choice of the smoothing parameter λ is crucial for an accurate estimation. For this purpose, a k-fold cross-validation procedure is considered to select the optimal λ among a set of possible values.

We consider the squared L_2-norm of the difference between the true density $f(\mathbf{x})$ and the estimated one $\hat{f}_\lambda(\mathbf{x})$:

$$\begin{aligned} l(\lambda) &= \int_{\mathcal{M}} (\hat{f}_\lambda(\mathbf{x}) - f(\mathbf{x}))^2 d\mathbf{x} = \\ &= \int_{\mathcal{M}} \hat{f}_\lambda^2(\mathbf{x}) d\mathbf{x} - 2\int_{\mathcal{M}} \hat{f}_\lambda(\mathbf{x}) f(\mathbf{x}) d\mathbf{x} + \int_{\mathcal{M}} f^2(\mathbf{x}) d\mathbf{x}. \end{aligned}$$

The first term is easily computed thanks to the finite element formulation and the quadrature rules; the second term is approximated evaluating the estimated density on the validation data; the third term can instead be omitted since it does not depend on λ. Therefore, the cross-validation error is defined as

$$\hat{R}(\lambda) = \int_{\mathcal{M}_{\mathcal{T}}} \left(\hat{f}_{\lambda}^{-[k]}(x)\right)^2 dx - \frac{2}{m} \sum_{i \in [k]} \hat{f}_{\lambda}^{-[k]}(x_i),$$

where $[k]$ is the set of indices of data in the k-th fold, m is its cardinality, and $\hat{f}_{\lambda}^{-[k]}(x)$ is the density estimated without the data in the k-th fold.

5. Simulation studies

In this section we illustrate the good performances of the proposed nonparametric Density Estimator with Partial Differential Equation regularization (DE-PDE).

In Sections 5.1 and 5.2 we compare DE-PDE to Kernel Density Estimation (KDE). To enable the comparison with KDE, we take as domain the unitary sphere $\mathcal{S}^2$ in $\mathbb{R}^3$. We consider two different test functions. In Simulation 1 (Section 5.1) we consider a mixture of five Kent distributions, shown in the first column of Figure 4. In Simulation 2 (Section 5.2) we consider a highly localized and skewed density with elongated ridges, shown in the first column of Figure 5; this test function is used to mimic earthquakes data.

In Simulation 3 (Section 5.3) we instead consider a density on the manifold in Figure 2, to show the performances of the proposed DE-PDE on a general two-dimensional curved domain. In this context KDE cannot be applied.

The meshes that approximate the manifold domains are created using Gmsh [16]. The number of nodes is chosen according to the complexity of the manifold and of the point pattern, in order to obtain a good approximation of the domain as well as to capture the features of the signals highlighted by the point pattern. The proposed DE-PDE method is implemented using the function `DE.FEM` of the `R` package `fdaPDE`; the smoothing parameter is chosen by cross validation. KDE is implemented using the function `kde.fhat.cart` of the `R` package `retistruct` [40]. This function implements the method described in [20]; it considers a Fisherian density and automatically selects the concentration parameter with the function `kde.compute.concentration`. Each simulation is repeated 30 times. Estimation accuracy of the two methods is measured in terms of the Mean Integrated Squared Error with respect to true density function f_0, namely: $\text{MISE}(\hat{f}) = \mathbb{E}\int_{\mathcal{M}}(\hat{f} - f_0)^2$. The MISE is evaluated over 30 simulation repetitions and computed numerically on a fine grid over the manifold.

5.1. Simulation 1: mixture of five Kent distributions

We sample 800 data from the mixture of five Kent distributions displayed in the first column of Figure 4. The components of this mixture present anisotropies along different directions. The detailed definition of the density is reported in Appendix D.

DE-PDE estimates use a mesh with 606 nodes. Figure 4 shows the mean estimates obtained by DE-PDE and by KDE over 30 simulation repetitions and the boxplot of the MISE over the 30 repetitions. Both the mean estimates and the boxplots of the

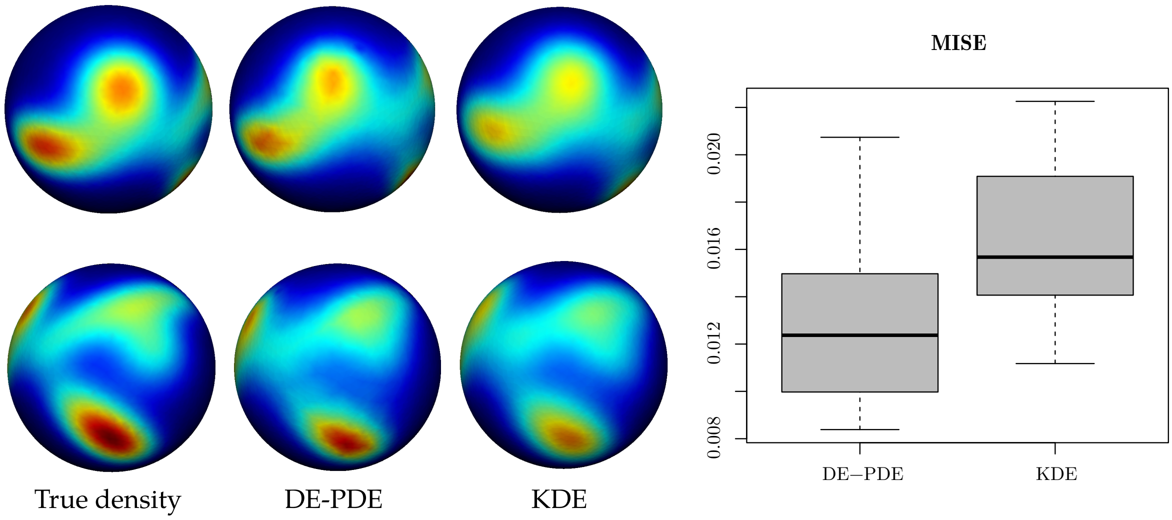

Figure 4: Simulation 1. On the left: two views of the true density (first column), of DE-PDE mean estimate over 30 repetitions (second column), and of KDE mean estimate over 30 repetitions (third column). On the right: boxplots of the MISE of DE-PDE estimates and KDE estimates over the 30 simulation repetitions.

MISE highlight the advantage of DE-PDE estimates over KDE estimates. In particular, DE-PDE appears to better capture skewed modes.

Analogous conclusions are obtained when considering different sample sizes. As expected, the MISE decreases as more data are sampled, both for DE-PDE and for KDE. The ordering of the methods remains the same, and the difference in performances increases as the sample size increases. In particular, with a sample size of $n = 200$ we get a median MISE of 0.0301 (with an IQR of 0.0098) for DE-PDE, and a median MISE of 0.0324 (with an IQR of 0.0107) for KDE; while, with a sample size of $n = 7500$ (approximately as in the application considered in Section 6) we get a median MISE of 0.0033 (with an IQR of 0.0011) for DE-PDE and a median MISE of 0.0046 (with an IQR of 0.0009) for KDE.

5.2. Simulation 2: highly localized and skewed density

The goal of this simulation study is to mimic the earthquake data. Earthquakes occur mainly on faults. Therefore our aim is to sample data that lies on elongated narrow regions over the sphere. To do this, data are generated from a mixture of 25 highly anisotropic Kent distributions, with means that lie along lines on the sphere, as detailed in Pigolotti [34]. The true density is displayed in the first column of Figure 5.

We generate 30 samples of 1000 data each from the true density, and compare DE-PDE to KDE. DE-PDE is implemented using a regular mesh with 3097 nodes. The results are shown Figure 5. The first column shows different views of the true density. The second and third column show, respectively, the mean estimates over 30 repetitions obtained by DE-PDE and KDE. In this setting the DE-PDE method is clearly better then KDE in representing the true density, as pointed out also by the boxplots of the MISE on the right panel of the same figure. The KDE method seems to flatten the density too much and it turns out to be worse than DE-PDE in assigning high density

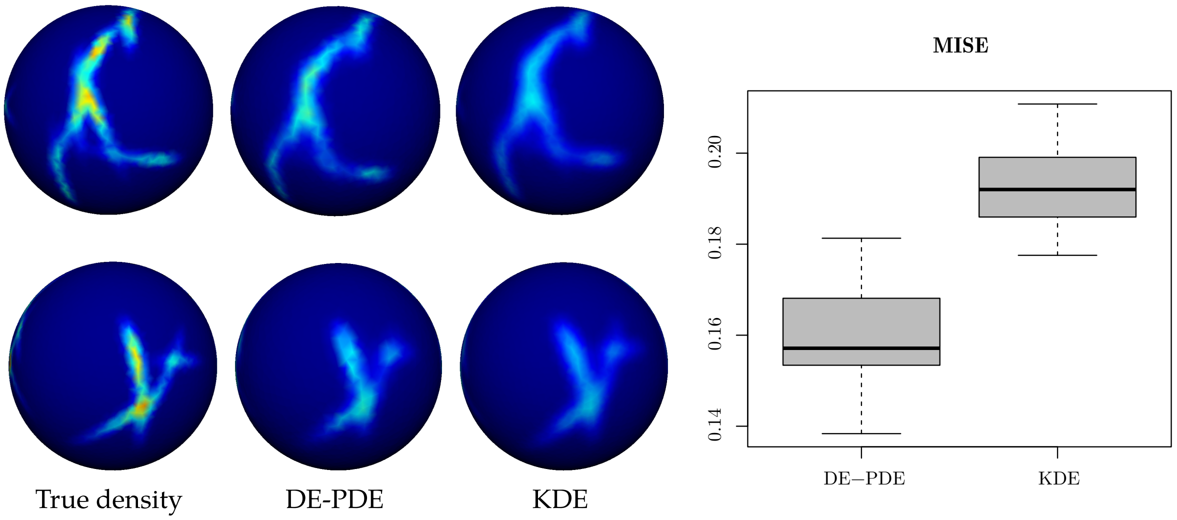

Figure 5: Simulation 2: results obtained with samples of 1000 data. On the left: first column: true density; middle column: DE-PDE mean estimate; third column: KDE mean estimate. The different rows show different components of the mixture. The results are averaged over 30 repetitions. On the right: boxplots of the MISE, over the 30 simulation repetitions, of the estimates provided by the two estimators.

along lines. The proposed DE-PDE method is instead particularly suited to work with data characterized by highly localized structures and skewed signals, thanks to the localized support of finite element basis and the associated unstructured meshes. We repeated the simulation with samples of 7500 data. Also in this case DE-PDE attains a significantly lower MISE and the difference in the performances of the two methods increases: we obtain a median MISE of 0.0742 (with an IQR of 0.0075) for DE-PDE, and a median MISE of 0.1587 (with an IQR of 0.0043) for KDE.

The tendency of kernels to over-smooth may result from the fact that the function `kde.fhat.cart` of the `R` package `retistruct`, that implements kernel density estimation on the sphere, is based on the von-Mises Fisher distribution and has only a scalar concentration parameter. This certainly implies a reduced flexibility in capturing highly localized features such as the one presented in this simulation and in the application to earthquake data. Moreover, the problem of optimal parameter selection has not been adequately addressed in the case of manifold domains. In contrast, for kernel density estimation over planar domains, it is possible to consider a full bandwidth matrix, thus allowing an accurate estimation of a richer class of signals [see, e.g., 32] and the problem of optimal parameter selection has been thoroughly studied [see, e.g., 6]. Nevertheless, the tendency of kernel estimates to over-smooth highly localized and skewed modes can be observed also over planar domains, as highlighted by the simulations in [11].

5.3. Simulation 3: manifold with boundary

We here consider the manifold in Figure 2. We sample 800 data from the true density shown in the left panel of Figure 6: the density is multimodal and presents modes far from the boundary as well as close to the boundary of the domain. The central panel of the Figure 6 shows a sample of the data.

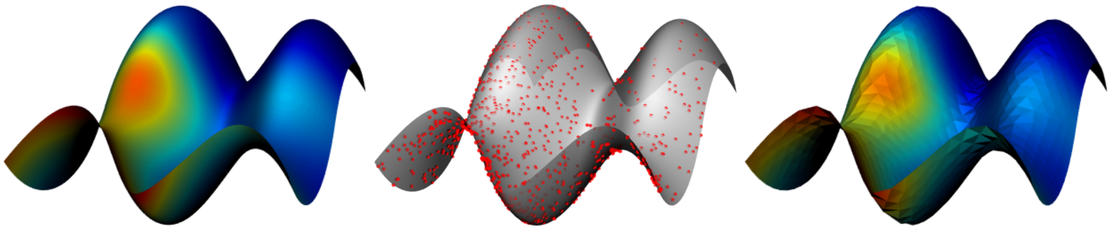
True density   Sample Data   DE-PDE

Figure 6: Simulation 3: results obtained with samples of 800 data. Left panel: true density; central panel: a sample of the data; right panel: DE-PDE mean estimate over 30 repetitions.

We use the mesh shown in Figure 2, that has 964 nodes. The right panel of Figure 6 shows the mean estimate over the 30 repetitions. The proposed DE-PDE is able to accurately estimate all the modes of the true density, also those falling close the boundary.

6. Earthquakes in the world

We apply the method to the problem of estimating the earthquakes distribution over the globe. The data are taken from the earthquake catalog of the USGS website (`https://earthquake.usgs.gov/earthquakes/search/`). Earthquakes with a magnitude greater than 4.5, occurred in the world in the period from 1st March 2019 to 1st March 2020, are considered. The dataset contains 7715 observations and locations are displayed as in Figure 7. The same data are displayed on the sphere in Figure 1.

The particularity of this dataset is that locations lie mainly on elongated narrow regions, that are the faults. Earthquakes in fact occur on these fractures of the earth's crust, where the mechanical stress induced by tectonic movements accumulates. The proposed DE-PDE method is particularly suitable for working with this kind of data, as shown by Simulation 2 in Section 5.2.

Figure 8, first row, shows the distribution of earthquakes in the world estimated with the proposed DE-PDE, on a mesh with 5016 nodes. The estimate is able to capture the distribution of earthquakes. A particularly high density is visualized in the area of the Pacific Ocean, between the Eurasian and the Pacific Plates, in northern hemisphere, and between the Indo-Australian and the Pacific Plates, in southern hemisphere; see the first panel of Figure 8. In particular the area around Japan is marked with an high estimated density; in fact, this is a well known high seismic risk area, along the Itoigawa-Shizuoka Tectonic Line and the Japan Median Tectonic Line, due to the presence of the boundaries of some of the major tectonic plates: the North American Plate, the Eurasian Plate and the Philippine Plate. As shown in second panel of the same figure, also the area nearby Indonesia is characterized by a significant density, along the Sunda megathrust and the Great Sumatran fault. These regions are indeed marked in history by important earthquakes and tsunamis, which caused lots of damages, such as the 2004 Indian Ocean earthquake and tsunami. The third panel in Figure 8 shows the estimated signal along the west coast of the American continent. In particular, in the North, the high density is along the Cascadia Subduction Zone and the Saint Andreas Fault, while in

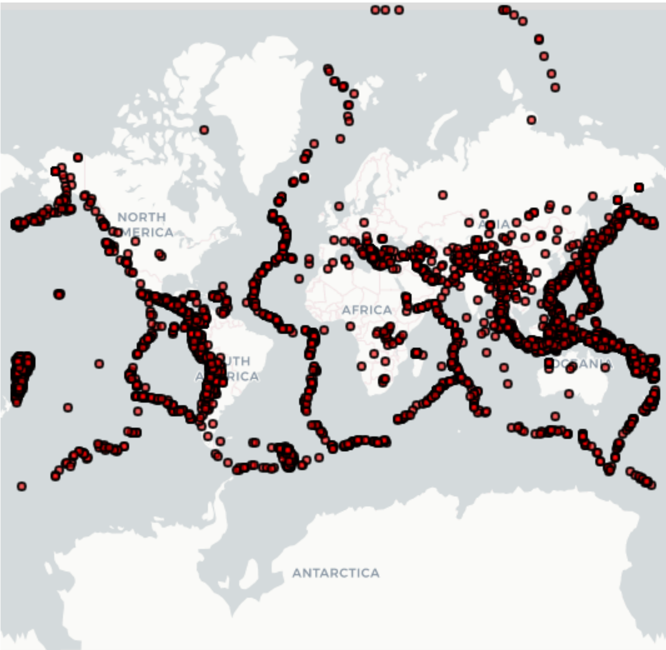

Figure 7: Locations of earthquakes with a magnitude greater then 4.5, in the period from 1st March 2019 to 1st March 2020. The dataset contains 7715 observations. The plot is made with the `R` package `mapview` [1].

the South it is along the Peru-Chile Trench, that separates the Nazca Plate and the South American Plate. Finally, in the last panel we can see that the estimate is also able to capture the lower signal along the Mid-Atlantic Ridge, that separates the American Plates from the Eurasian and African Plates.

The bottom panels of Figure 8 shows the same views of the estimate obtained by KDE. As already pointed out in Simulation 2, KDE is not able to accurately estimate highly localized and skewed modes and returns an over-smoothed density.

7. Conclusions and future developments

The proposed DE-PDE method is able to accurately capture highly complicated multi-modal densities over two-dimensional manifolds. DE-PDE is amenable to various extensions. A particularly interesting generalization concerns the development of a method to estimate time-dependent densities over two-dimensional manifolds. The modeling of densities over time permits the understanding of the evolution of underlying processes generating the data [see, e.g. 14, 42, 15, 9, in the case of densities over planar domains]. In particular, the proposed DE-PDE method could be generalized to space-time point data, either by considering two regularizations, one in time and one in space, or alternatively by using a unique regularization involving a time-dependent differential operator, in analogy to the spatio-temporal regression methods presented in Bernardi et al. [3] and Arnone et al. [2]. Another interesting line of research goes in the direction of intensity estimation in marked point processes. In the application to earthquake data this would for instance permit to include the magnitude of the

Figure 8: Distribution of earthquakes in the world. Estimate obtained by DE-PDE (top panels) and by KDE (bottom panels).

earthquake in the data analysis.

Appendix A The estimate as an unconstrained optimum: proof

We report the proof of the fact that the minimum of the functional (2) corresponds to the constrained minimum of the functional (1). The proof follows closely the one for the one dimensional case in Silverman [38]. Consider the space $V_0 = \{u \in V \text{ s.t. } \int e^u = 1\}$, and define the functional $L_0(u)$ as

$$L_0(u) = -\frac{1}{n}\sum_{i=1}^{n} u(\mathbf{x}_i) + \lambda \int_\Omega (\Delta u)^2.$$

Lemma 1. *The function* $\hat{u}$ *minimizes* $L_0(u)$ *over* $u \in V_0$ *if and only if* $\hat{u}$ *minimizes* $L(u)$ *over* V.

Proof. First of all, observe that if $u \in V_0$ we have $L(u) = L_0(u) + 1$, therefore the minimization of L in V_0 is equivalent to the minimization of L_0 in V_0. Take $u \in V$ and define $u^* = u - \log \int e^u$, so that $\int e^{u^*} = 1$, i.e., $u^* \in V_0$. Since u and u^* differ only by a constant, we have $\Delta u = \Delta u^*$, and therefore

$$\begin{aligned} L(u^*) &= -\frac{1}{n}\sum_{i=1}^{n} u^*(\mathbf{x}_i) + 1 + \lambda \int_\Omega (\Delta u^*)^2 \\ &= -\frac{1}{n}\sum_{i=1}^{n} u(\mathbf{x}_i) + \log \int e^u + 1 + \lambda \int_\Omega (\Delta u)^2 \\ &= L(u) - \int e^u + \log \int e^u + 1. \end{aligned}$$

Since $-t + \log t + 1 \leq 0$ for all $t > 0$, with equality only if $t = 1$, we have that $L(u^*) \leq L(u)$, with equality only if $\int e^u = 1$. This implies that the minimizer of L in V is in V_0, and therefore satisfies the constrain $\int e^u = 1$. □

Appendix B Proof of Theorem 1

Let V_0 denote the null space of the Laplace-Beltrami operator in V, i.e., $V_0 = \{u \in V : ||\Delta_{\mathcal{M}} u||_{L^2} = 0\}$. Let V_Δ denote the complementary space of V_0 in V, i.e., $V = V_0 \oplus V_\Delta$, where $\oplus$ denotes the direct sum. The proof of the Theorem 1 relies on the following two lemmas.

Lemma 2. *The functional* $J(u) = -\frac{1}{n}\sum_{i=1}^n u(x_i) + \int_{\mathcal{M}} e^{u(x)}$ *is continuous and strictly convex in* V.

Lemma 3. V_0 *is of finite dimension. Moreover* $||\Delta_{\mathcal{M}} \cdot ||_{L^2}$ *is a norm in the space* V_Δ*, equivalent to the* H^2 *norm.*

The proof of Lemma 2 can be derived analogously to the one in Ferraccioli et al. [11], that considers the simpler case of planar domains. Concerning Lemma 3, observe that the space V_0 is the eigenspace of the Laplace-Beltrami operator and thus is of finite dimension. Moreover, the eigenvalue zero has multiplicity one and the corresponding eigenfunction is the constant function over the domain. In addition, if $\mathcal{M}$ is a two-dimensional Riemannian manifold, then $||\Delta_{\mathcal{M}} \cdot ||_{L^2}$ is equivalent to $|| \cdot ||_{H^2}$ [see 7].

Thanks to the two lemmas above, we can leverage on Theorem 4.1 of Gu and Qiu [19]. This theorem states that the functional $L(u)$ in (2) has a unique minimizer in V if and only if $-\frac{1}{n}\sum_{i=1}^n u(x_i) + \int_{\mathcal{M}} e^{u(x)}$ has a minimizer in V_0. Since V_0 is a finite dimensional space, the latter condition is verified and therefore the functional $L(u)$ in (2) has a unique minimizer in V.

Appendix C Proof of Theorem 2

The proof of Theorem 2 relies on the following two lemmas.

Lemma 4. *Let* u_0 *be the true log likelihood and* $\hat{u}$ *the minimizer of* (2)*. Then*

$$D_{sKL}(u_0, \hat{u}) = 2\lambda \int_{\mathcal{M}} \hat{u}(u_0 - \hat{u}) + \left[\frac{1}{n}\sum_{i=1}^n (\hat{u} - u_0)(x_i) - \mu_{u_0}(\hat{u} - u_0)\right]. \tag{6}$$

Lemma 5. *Under Assumption 1, there exists an infinite set of functions* φ_k *such that*

$$\mathrm{Cov}(\varphi_k, \varphi_j) = \delta_{k,j} \quad \text{and} \quad \int_{\mathcal{M}} \Delta_{\mathcal{M}}\varphi_k \Delta_{\mathcal{M}}\varphi_j = \eta_k^2 \delta_{k,j}$$

where $\delta_{k,j}$ *is the Kronecker delta and* $0 \leq \eta_k \longrightarrow \infty$*. In additions, if* $\mathcal{M}$ *is a compact Riemannian manifold, there exist two positive constants* α *and* β *such that, for all* $k \geq 0$,

$$\eta_k = c_k k, \quad \alpha \leq c_k \leq \beta. \tag{7}$$

The proofs of Lemma 4 and 5 and of Theorem 2 can be derived along the lines of Ferraccioli et al. [11], that considers the case of bounded planar domains. In particular, the proof of Lemma 4 is derived setting $A_{u,h}(t) := -\frac{1}{n}\sum(u+th)(X_i) + \int \exp(u+th) + \lambda \int \Delta(u+th)^2$, differentiating it in t, and evaluating it in $t = 0$. The proof of Lemma 5 relies on the eigenvalue problem associated with the Laplace-Beltrami operator, and exploits results of functional analysis on manifolds; in particular, the rate in equation (7) can be found, e.g., in Chavel [7].

Appendix D Simulation 1: test function

The test function in Simulation 1 in Section 5 is a mixture of five Kent distributions [24] with the following scale parameters:

$$\begin{array}{lllll} k_1 = 18, & k_2 = 15, & k_3 = 20, & k_4 = 20, & k_5 = 20, \\ \beta_1 = 0, & \beta_2 = 7, & \beta_3 = 10, & \beta_4 = 7, & \beta_5 = 4, \end{array}$$

and position parameters:

$$\gamma_{1,1} = \begin{pmatrix} -0.5 \\ -0.5 \\ 0.8 \end{pmatrix}, \quad \gamma_{1,2} = \begin{pmatrix} -0.3 \\ -0.3 \\ 0.2 \end{pmatrix}, \quad \gamma_{1,3} = \begin{pmatrix} 0.5 \\ -0.5 \\ 0.8 \end{pmatrix}, \quad \gamma_{1,4} = \begin{pmatrix} 0.2 \\ -1 \\ 0 \end{pmatrix}, \quad \gamma_{1,5} = \begin{pmatrix} 0.6 \\ -0.5 \\ 0.3 \end{pmatrix},$$

$$\gamma_{2,1} = \begin{pmatrix} -0.7789 \\ 0.6157 \\ 0.1188 \end{pmatrix}, \quad \gamma_{2,2} = \begin{pmatrix} -0.8651 \\ 0.3803 \\ -0.3269 \end{pmatrix}, \quad \gamma_{2,3} = \begin{pmatrix} -0.6664 \\ -0.7432 \\ -0.0584 \end{pmatrix}, \quad \gamma_{2,4} = \begin{pmatrix} 0.5753 \\ -0.4629 \\ -0.6742 \end{pmatrix}, \quad \gamma_{2,5} = \begin{pmatrix} 0.7545 \\ -0.2314 \\ -0.6140 \end{pmatrix},$$

$$\gamma_{3,1} = \begin{pmatrix} -0.5695 \\ -0.6154 \\ -0.5448 \end{pmatrix}, \quad \gamma_{3,2} = \begin{pmatrix} 0.1482 \\ -0.4288 \\ -0.8911 \end{pmatrix}, \quad \gamma_{3,3} = \begin{pmatrix} 0.5753 \\ -0.4629 \\ -0.6742 \end{pmatrix}, \quad \gamma_{3,4} = \begin{pmatrix} 0.6364 \\ -0.0303 \\ -0.7707 \end{pmatrix}, \quad \gamma_{3,5} = \begin{pmatrix} 0.6364 \\ -0.0303 \\ -0.7707 \end{pmatrix}.$$

The mean directions $\gamma_{1,1}, \dots, \gamma_{1,5}$ are then normalized to lay on the unitary sphere. The mixing weights are set as:

$$\pi = \left(\tfrac{1}{5}, \ \tfrac{1}{5}, \ \tfrac{1}{5}, \ \tfrac{1}{5}, \ \tfrac{1}{5}\right).$$

References

[1] T. Appelhans. mapview, r package version 2.7., 2019. URL `https://cran.r-project.org/web/packages/mapview`.

[2] E. Arnone, L. Azzimonti, F. Nobile, and L. M. Sangalli. Modeling spatially dependent functional data via regression with differential regularization. *Journal of Multivariate Analysis*, pages 275–295, 2019.

[3] M. S. Bernardi, L. M. Sangalli, G. Mazza, and J. O. Ramsay. A penalized regression model for spatial functional data with application to the analysis of the production of waste in venice province. *Stochastic Environmental Research and Risk Assessment*, pages 23–28, 2017.

[4] T. Berry and T. Sauer. Density estimation on manifolds with boundary. *Computational Statistics and Data Analysis*, pages 1–17, 2017.

[5] H. Brezis. *Functional analysis, Sobolev spaces and partial differential equations*. Springer Science and Business Media, 2010.

[6] J. E. Chacón and T. Duong. *Multivariate kernel smoothing and its applications*. CRC Press, 2018.

[7] I. Chavel. *Eigenvalues in Riemannian Geometry*. Academic Press, Inc., 1984.

[8] M. Di Marzio, A. Panzera, and C. C. Taylor. Kernel density estimation on the torus. *Journal of Statistical Planning and Inference*, 141(6):2156–2173, 2011.

[9] P. J. Diggle, J. Mateu, and H. E. Clough. A comparison between parametric and non-parametric approaches to the analysis of replicated spatial point patterns. *Adv. in Appl. Probab.*, pages 331–343, 2000.

[10] G. Dziuk and C. M. Elliott. Finite element methods for surface pdes. *Acta Numerica*, pages 289–396, 2013.

[11] F. Ferraccioli, E. Arnone, L. Finos, J. O. Ramsay, and L. M. Sangalli. Nonparametric density estimation over complicated domains. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 83(2):346–368, 2021.

[12] N. I. Fisher. *Statistical analysis of circular data*. Cambridge University Press, 1995.

[13] N. I. Fisher, T. Lewis, and B. J. J. Embleton. *Statistical Analysis of Spherical Data*. Cambridge University Press, 1993.

[14] D. Gervini. Independent component models for replicated point processes. *Spat. Stat.*, 2016.

[15] D. Gervini. Doubly stochastic models for replicated spatio-temporal point processes. *ArXiv*, 2019.

[16] C. Geuzaine and J.-F. Remacle. Gmsh: A 3-d finite element mesh generator with built-in pre-and post-processing facilities. *International journal for numerical methods in engineering*, 79(11):1309–1331, 2009.

[17] I. J. Good and R. A. Gaskins. Density estimation and bump-hunting by the penalized likelihood method exemplified by scattering and meteorite data. *Journal of the American Statistical Association*, pages 42–56, 1980.

[18] C. Gu. Smoothing spline density estimation: A dimensionless automatic algorithm. *Journal of the American Statistical Association*, 88(422):495–504, 1993.

[19] C. Gu and C. Qiu. Smoothing spline density estimation: theory. *The Annals of Statistics*, pages 217–234, 1993.

[20] P. Hall, G. S. Watson, and J. Cabrera. Kernel density estimation with spherical data. *Biometrika*, pages 751–62, 1987.

[21] H. Hendriks. Nonparametric estimation of a probability density on a riemannian manifold using fourier expansions. *The Annals of Statistics*, pages 832–849, 1990.

[22] G. Henry, A. Munoz, and D. Rodriguez. Locally adaptive density estimation on riemannian manifolds. *SORT-Statistics and Operations Research Transactions*, pages 111–130, 2013.

[23] P. Jupp and K. Mardia. A unified view of the theory of directional statistics. *International Statistical Review*, pages 261–294, 1989.

[24] J. T. Kent. The fisher-bingham distribution on the sphere. *Journal of the Royal Statistical Society: Series B (Methodological)*, 44(1):71–80, 1982.

[25] Y. T. Kim and H. S. Park. Geometric structures arising from kernel density estimation on riemannian manifolds. *Journal of Multivariate Analysis*, pages 112–126, 2013.

[26] E. Lila, J. A. Aston, L. M. Sangalli, et al. Smooth principal component analysis over two-dimensional manifolds with an application to neuroimaging. *The Annals of Applied Statistics*, 10(4):1854–1879, 2016.

[27] E. Lila, L. M. Sangalli, E. Arnone, J. Ramsay, and L. Formaggia. *fdaPDE: Statistical Analysis of Functional and Spatial Data, Based on Regression with PDE Regularization*, 2020. URL https://CRAN.R-project.org/package=fdaPDE. R package version 1.0-9.

[28] K. Mardia. *Statistics of Directional Data*. Academic Press, 1972.

[29] G. McSwiggan, A. Baddeley, and G. Nair. Kernel density estimation on a linear network. *Scandinavian Journal of Statistics*, 44(2):324–345, 2017.

[30] M. M. Moradi and J. Mateu. First-and second-order characteristics of spatio-temporal point processes on linear networks. *Journal of Computational and Graphical Statistics*, 29(3):432–443, 2020.

[31] M. M. Moradi, O. Cronie, E. Rubak, R. Lachieze-Rey, J. Mateu, and A. Baddeley. Resample-smoothing of voronoi intensity estimators. *Statistics and computing*, 29 (5):995–1010, 2019.

[32] V. M. Panaretos and K. Konis. Nonparametric construction of multivariate kernels. *Journal of the American Statistical Association*, 107(499):1085–1095, 2012.

[33] B. Pelletier. Kernel density estimation on riemannian manifolds. *Statistics and Probability Letters*, pages 297–304, 2005.

[34] C. Pigolotti. Density estimation with differential regularization, 2020.

[35] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2020. URL `https://www.R-project.org/`.

[36] S. Rakshit, T. Davies, M. M. Moradi, G. McSwiggan, G. Nair, J. Mateu, and A. Baddeley. Fast kernel smoothing of point patterns on a large network using two-dimensional convolution. *International Statistical Review*, 87(3):531–556, 2019.

[37] L. M. Sangalli. Spatial regression with partial differential equation regularization. *International Statistical Review*, 2021. doi: 10.1111/insr.12444. URL `https://doi.org/10.1111/insr.12444`.

[38] B. W. Silverman. On the estimation of a probability density function by the maximum penalized likelihood method. *The Annals of Statistics*, page 795–810, 1982.

[39] D. Simpson, J. B. Illian, F. Lindgren, S. H. Sørbye, and H. Rue. Going off grid: Computationally efficient inference for log-gaussian cox processes. *Biometrika*, 103 (1):49–70, 2016.

[40] D. C. Sterratt and D. Lyngholm. retistruct: Retinal reconstruction program, r package version 0.6.2, 2019. URL `https://cran.r-project.org/web/packages/retistruct`.

[41] G. Watson. *Statistics on Spheres*. Wiley, 1983.

[42] Y. Yuan, F. E. Bachl, F. Lindgren, D. L. Borchers, J. B. Illian, S. T. Buckland, H. Rue, T. Gerrodette, et al. Point process models for spatio-temporal distance sampling data from a large-scale survey of blue whales. *The Annals of Applied Statistics*, 11 (4):2270–2297, 2017.

Supplementary material for: A roughness penalty approach to estimate densities over two-dimensional manifolds

Eleonora Arnone[a], Federico Ferraccioli[b], Clara Pigolotti[a], Laura M. Sangalli[a,*]

[a]*MOX - Department of Mathematics, Politecnico di Milano, Piazza Leonardo da Vinci 32, Milano, 20133, Italy*
[b]*Department of Statistical Sciences, University of Padova, Via 8 Febbraio, 2, Padova, 35122, Italy*

This document describes the contents of the supplementary material. The supplementary material contains the data and the scripts to reproduce the results in Section 5 of the paper.

1. R files

Simulation1.R the main to run the first simulation study (Section 5.1).

Simulation2.R the main to run the second simulation study (Section 5.2).

Simulation3.R the main to run the third simulation study (Section 5.3).

Generate.samples.R the scripts used to generate the samples in the data files. It is not needed to run the simulations in the paper, but it can be used to generate other samples from the densities used in the paper.

plot.R a function to plot functions on a manifold with the possibility to specify the range of the colormap (the default is that the range of the colormap is equal to the range of the function to be plotted).

sample.functions.R the functions to sample from the density used in the simulations.

2. Data files

simulation1.200data.txt the data for the 30 processes, with 200 observations each, considered in the Simulation 1, Section 5.1. The first 200 rows correspond to the first generated point process, the second 200 rows correspond to the second generated point process, and so on.

simulation1.800data.txt the data for the 30 processes, with 800 observations each, considered in the Simulation 1, Section 5.1. The first 800 rows correspond to the first generated point process, the second 800 rows correspond to the second generated point process, and so on.

---

[*] Corresponding author
Email address: `laura.sangalli@polimi.it` (Laura M. Sangalli)

simulation1.7500data.txt the data for the 30 processes, with 7500 observations each, considered in the Simulation 1, Section 5.1. The first 7500 rows correspond to the first generated point process, the second 7500 rows correspond to the second generated point process, and so on.

simulation2.1000data.txt the data for the 30 processes, with 1000 observations each, considered in the Simulation 2, Section 5.2. The first 1000 rows correspond to the first generated point process, the second 1000 rows correspond to the second generated point process, and so on.

simulation2.7500data.txt the data for the 30 processes, with 7500 observations each, considered in the Simulation 2, Section 5.2. The first 7500 rows correspond to the first generated point process, the second 7500 rows correspond to the second generated point process, and so on.

simulation3.800data.txt the data for the 30 processes, with 800 observations each, considered in the Simulation 3, Section 5.3. The first 800 rows correspond to the first generated point process, the second 800 rows correspond to the second generated point process, and so on.

simulation3.fullPoints.proj.RData a set of approximately 10^5 points, regularly spaced on the manifold domain of Simulation 3, used for the sampling on the manifold.

3. Mesh files

simulation1.sphere.vertices.txt the coordinates of the 606 vertices of the mesh in Simulation 1.

simulation1.sphere.triangles.txt the elements of the mesh with 606 nodes in Simulation 1.

simulation1.sphere.vertices.eval.txt the coordinates of the points to evaluate the MISE in Simulation 1.

simulation1.sphere.triangles.eval.txt the elements of a fine mesh to plot the results of Simulation 1.

simulation2.sphere.vertices.txt the coordinates of the 3097 vertices of the mesh in Simulation 2.

simulation2.sphere.triangles.txt the elements of the mesh with 3097 nodes in Simulation 2.

simulation3.surface.vertices.txt the coordinates of the 964 vertices of the mesh in Simulation 3.

simulation3.surface.triangles.txt the elements of the mesh with 964 nodes in Simulation 3.

MOX Technical Reports, last issues

Dipartimento di Matematica
Politecnico di Milano, Via Bonardi 9 - 20133 Milano (Italy)

40/2022 Fumagalli, A.; Patacchini, F. S.
Well-posedness and variational numerical scheme for an adaptive model in highly heterogeneous porous media

39/2022 Ferro, N.; Perotto, S.; Gavazzoni, M.
A new fluid-based strategy for the connection of non-matching lattice materials

37/2022 Boon, W. M.; Fumagalli, A.
A Reduced Basis Method for Darcy flow systems that ensures local mass conservation by using exact discrete complexes

38/2022 Burzacchi, A.; Landrò, M.; Vantini, S.
Object-oriented Classification of Road Pavement Type in Greater Maputo from Satellite Images

34/2022 Antonietti, P.F.; Vacca, G.; Verani, M.
Virtual Element method for the Navier-Stokes equation coupled with the heat equation

35/2022 Perotto, S.; Bellini, G.; Ballarin, F.; Calò, K.; Mazzi, V.; Morbiducci, U.
Isogeometric hierarchical model reduction for advection-diffusion process simulation in microchannels

36/2022 Vaccaro, F.; Brivio, S.; Perotto, S.; Mauri, A.G.; Spiga, S.
Physics-based Compact Modelling of the Analog Dynamics of HfOx Resistive Memories

33/2022 Africa, P.C.; Salvador, M.; Gervasio, P.; Dede', L.; Quarteroni, A.
A matrix-free high-order solver for the numerical solution of cardiac electrophysiology

30/2022 Bonetti S.; Botti M.; Antonietti P.F.
Discontinuous Galerkin approximation of the fully-coupled thermo-poroelastic problem

31/2022 Bortolotti, T; Peli, R.; Lanzano, G; Sgobba, S.; Menafoglio, A
Weighted functional data analysis for the calibration of ground motion models in Italy