MOX-Report No. 32/2021

# Spatial regression with partial differential equation regularization

Sangalli, L.M.

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

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

# Spatial regression with partial differential equation regularization

Laura M. Sangalli

MOX - Dipartimento di Matematica, Politecnico di Milano

**Abstract**

This work gives an overview of an innovative class of methods for the analysis of spatial and of functional data observed over complicated two-dimensional domains. This class is based on regression with regularizing terms involving partial differential equations. The associated estimation problems are solved resorting to advanced numerical analysis techniques. The synergical interplay of approaches from statistics, applied mathematics and engineering endows the methods with important advantages with respect to the available techniques, and makes them able to accurately deal with data structures for which the classical techniques are unfit. Spatial regression with differential regularization is illustrated via applications to the analysis of eco-color doppler measurements of blood-flow velocity, and to functional magnetic resonance imaging signals associated with neural connectivity in the cerebral cortex.

**Keywords:** functional data analysis, spatial data analysis, high dimensional and complex data, partial differential equations, finite elements.

## 1 Introduction

This work reviews an innovative class of methods for the analysis of data that display complicated spatial or spatio-temporal dependencies. The complexity in the structure of spatial or spatio-temporal variation may be due to different reasons. In some cases, the complexity originates from the complex physics of the phenomenon under study. As an illustrative application, we will discuss the study of the velocity field of blood flow in human arteries, with data from eco-color doppler and magnetic resonance imaging. In other cases, the complexity is due to an external source that generates strong anisotropies and non-stationarities in the observed quantity of interest; this is the case, for instance, of streams or prevailing winds in the analysis of environmental and climate data. In yet other cases, the complex spatial variation might be influenced by the complicated conformation of the domain where the data are observed. As an example, Figure 1 shows the spatial distribution of all the crimes reported in 2012 in the city of Portland, Oregon, USA, each crime location being indicated by a dot. The domain of interest for this application, i.e., the territory of the municipality, is characterized by a strong concavity, generated by the presence of the Willamette river that cuts the city in two parts, connected by a few bridges downtown. As apparent in the figure, the variation of the phenomenon is not smooth across the river: in the north of the municipality, for instance, a rather high criminality affects the area East of the river, while a very low criminality is observed on the West of the Willamette river; this sharp difference cannot be fully explained by population density and other census information. For this reason, when analysing these data, it is important to also take appropriately into account the domain conformation. In other applied problems the domain is a two-dimensional manifold. Data distributed over non-planar domains with complicated geometries are in fact nowadays common in varied contexts. Figure 2 illustrates an example from the neurosciences, concerning the study of high-dimensional neuroimaging signals associated with neuronal activity in the cerebral cortex, a highly convoluted thin sheet of neural tissue that constitutes the outermost part of the brain, and where most neural activity is focused; see the left panel in Figure 2. Due to the highly convoluted

nature of the cortex, functionally distinct areas, that are far apart along the cortex, may in turn be close in three-dimensional space. For this reason, neglecting the morphology of the cortex may lead to totally inaccurate estimates. Moreover, it is often crucial to comply with specific conditions that characterize the phenomenon at the boundaries of the spatial domain. As an example, in the already mentioned study of the velocity of blood flow, the domain of interest is the quasi-circular section of an artery and there are some physiological constraints at the boundaries of this domain: the blood flow is indeed confined within this domain and its velocity is zero at the arterial wall, the boundaries of the spatial domain, due to the friction between the blood particles and the arterial wall.
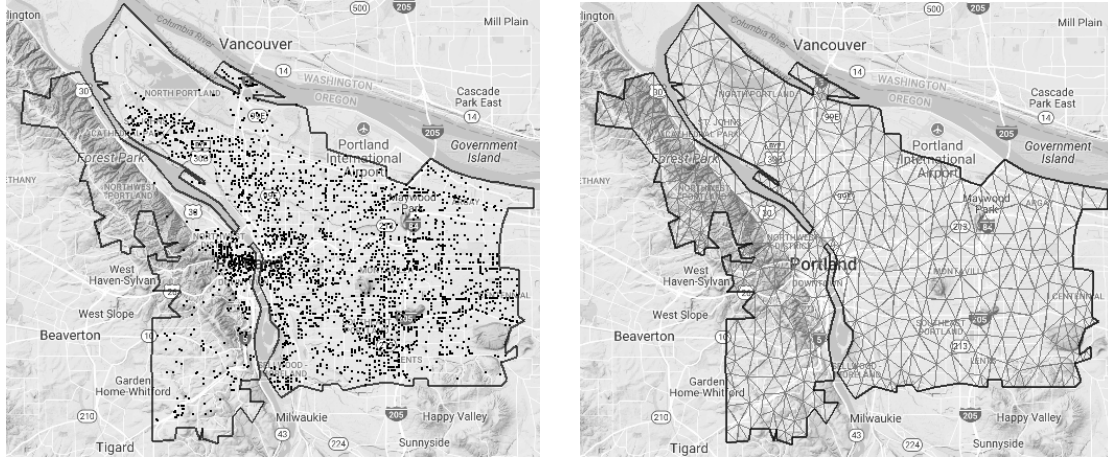


Figure 1: Left: Spatial distribution of crimes over the city of Portland, Oregon, USA: each dot indicates the location of one of more crimes (data from the Real-Time Crime Forecasting Challenge of the National Institute of Justice https://nij.ojp.gov/funding/real-time-crime-forecasting-challenge). The figure shows that criminality does not vary smoothly across the river. This is particularly apparent in the north of the municipality, where the high criminality in the East of the river is opposed to a low criminality in the West of the river; this sharp difference cannot be fully explained by population density or other census information. Right: triangulation of the domain.
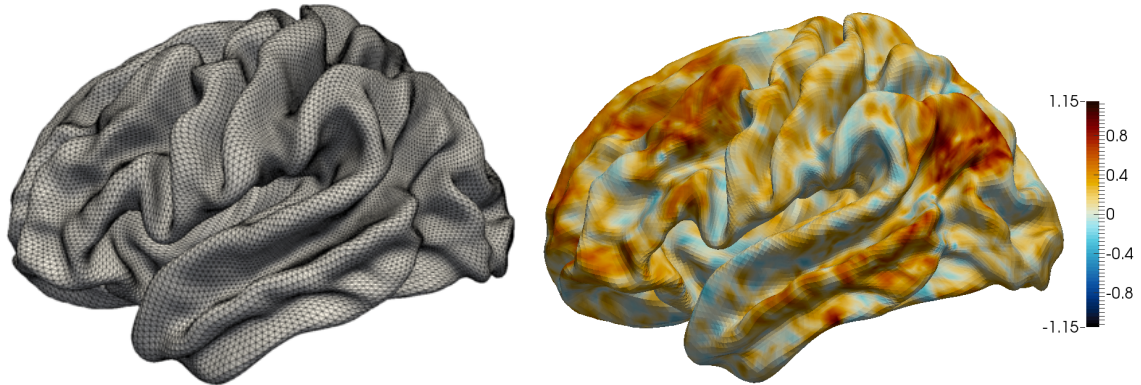


Figure 2: Left: Triangulated surface approximating the left hemisphere of the cerebral cortex of a template brain. Right: functional connectivity map obtained from a functional magnetic resonance imaging scan on a healthy subject (data from the Human Connectome Project Essen et al., 2012).

This work focuses entirely on a specific class of models we have recently been developing (in Sangalli et al., 2013; Azzimonti et al., 2014, 2015; Wilhelm and Sangalli, 2016; Ettinger et al., 2016; Wilhelm et al., 2016; Lila et al., 2016; Bernardi et al., 2017, 2018; Arnone et al., 2019), that we name

Spatial Regression with Partial Differential Equation regularization: SR-PDE. These are regression methods with regularization terms that involve a Partial Differential Equation (PDE). PDEs are extensively employed in most fields of sciences and engineering to model complex phenomena behaviors. Their inclusion as regularizing terms in a spatial regression model establishes a novel way of defining spatial variation, offering an important alternative to the classical paradigm in spatial data analysis. The differential regularization permits to model the spatial and spatio-temporal variation in a highly rich and flexible manner, that can naturally render varied forms of anisotropy and non-stationarity, and that can be driven by problem-specific knowledge. Moreover, SR-PDE can accurately and efficiently deal with data distributed over complex spatial domains, such as domains with strong concavities or holes, and non-planar domains with complicated geometries; in fact, it considers distances within the domain of interest, naturally complying with boundaries and non-Euclidean geometries. Furthermore, it is able to account for specific conditions at the boundaries of the domain of interest. SR-PDE also permits the definition of principal component analysis techniques, thus enabling population studies of complicated spatial signals.

All this is made possible by a synergistic interplay of approaches and ideas from statistics, applied mathematics and engineering. The estimation problem in SR-PDE is solved resorting to numerical analysis techniques, such as finite element analysis or isogeometric analysis (see,e.g., the textbooks Ciarlet, 2002; Cottrell and Bazilevs, 2009). These techniques also ensure the high computational efficiency of the methods.

SR-PDE methods are implemented in the R package `fdaPDE` (Lila et al., 2020), available from CRAN (R Core Team, 2020).

The remainder of the paper is organized as follows. Section 2 briefly offers the essential background literature. Section 3 introduces the general form of SR-PDE, discuss the modelling of space variation through the differential regularization (Section 3.1), the inclusion of boundary conditions (Section 3.2), and the characterization of SR-PDE infinite dimensional estimation problem (Section 3.3). Section 4 reviews the use of numerical techniques to solve the estimation problem. Section 5 derives some properties of the estimators. Section 6 outlines a few model extensions, including different sampling schemes (Section 6.1), generalized linear models (Section 6.2), models for spatio-temporal data (Section 6.3) and for data observed over non-planar domains (Section 6.4). Section 7 describes a first case study concerning the analysis of blood-flow velocity. Section 8 reviews a regularized functional principal component analysis method that relies on SR-PDE. Section 9 briefly presents a study of neuroimaing signals measuring cortical connectivity on a population of healthy subjects. Section 10 draws some concluding remarks.

# 2    Background literature

This article reviews progress to date on a very specific class of models, SR-PDE. This research line is across spatial data analysis and functional data analysis; moreover, the basic form of SR-PDE is a special case of a nonparametric/semiparametric model with quadratic differential regularization. This section gives some essential classical references of these important areas of statistics, setting the background for the proposed models. Moreover, it points to other recent proposals to handle data over complicated domains.

## 2.1    Spatial data analysis

Spatial data analysis is one the main branches of statistics and has played a leading role in the development of models and methods for the analysis of spatially (and spatio-temporal) distributed data, with a vast literature (see, e.g., the textbooks Cressie, 2015; Cressie and Wikle, 2011; Diggle and Ribeiro, 2007), and extensive software produced (see, e.g., Bivand et al., 2013). Most approaches to spatial data analysis are designed for data scattered over regular domains. See Section 2.4 for some proposals that try to overcome this limitation.

In the classical spatial data analysis approach, the unknown spatial field is assumed to be stochastic and the covariance of the stochastic field is used to model the second order spatial

dependence of the phenomenon under study. Instead, in the proposed SR-PDE, the unknown spatial field is assumed to be deterministic and the spatial structure of the phenomenon is modelled via the PDE in the regularizing term.

## 2.2  Functional data analysis

Functional data analysis is concerned with samples whose atoms are noisy measurements of functions (see, e.g., Ramsay and Silverman, 2005; Ferraty and Vieu, 2006; Kokoszka and Reimherr, 2017). The connection between the proposed class of models and functional data analysis is made explicit in Section 8, that reviews a functional principal component method for samples of functions over complicated two-dimensional domains, based on SR-PDE. More generally, the use of differential regularizations is widespread in functional data analysis. In particular, Ramsay and Hooker (2017) describes similar approaches to the one here proposed, but in simpler one-dimensional settings and using ordinary differential equations.

Moreover, there is a very active area of research across spatial data analysis and functional data analysis. Some classical spatial data analysis methods such as kriging have been extended to deal with functional data (see, e.g., the reviews and collections in Delicado et al., 2010; Mateu and Romano, 2017; Menafoglio and Secchi, 2017; Martínez-Hernández and Genton, 2020). SR-PDE contributes to the literature on spatial functional data analysis by proposing innovative approaches for the study of spatially distributed data, that are influenced by the literature on functional data analysis. Moreover, the method extension discussed in Section 6.3 explicitly targets spatially dependent functional data.

## 2.3  Nonparametric and semiparametric models with quadratic differential regularization

Various forms of nonparametric smoothers and of semiparametric models with quadratic differential regularization have been proposed in the literature, and their properties have been thoroughly studied in a well established literature (see, e.g., the textbooks by Green and Silverman, 1994; Eubank, 1999; Bickel et al., 1998; Härdle et al., 2004, and references therein). With respect to these models, SR-PDE features more general regularizing terms, which can involve linear second-order PDEs with space-varying second order, first order and zero order differential operators, and space-varying forcing terms, instead of the simple differential operators, constant over space, typically considered by classical smoothers and semiparametric models. Moreover, as emphasized in the Introduction, SR-PDE considers spatial domains with complex shapes, both planar and curved, and complies with general forms of boundary conditions. Such high flexibility comes at the price of a higher analytic complexity of the proposed models. The solution to the estimation problem in SR-PDE is not explicitly available, differently from classical smoothers and semiparametric models with differential regularization. Advanced numerical analysis techniques are needed to obtain approximate solutions.

## 2.4  Data over complex domains

The problem of handling data over non-trivial planar domains, appropriately taking into account the shape of the domain, has recently attracted an increasing interest, and some techniques have been proposed to tackle this issue. Regularized least-square smoothing methods, that can handle data on complex planar domains, include: FELsplines (Ramsay, 2002); soap film smoothing (Wood et al., 2008); bivariate splines over triangulations (see, e.g., Lai and Schumaker, 2007; Guillas and Lai, 2010; Ettinger et al., 2012; Lai and Wang, 2013); low-rank thin-plate spline approximations (Wang and Ranalli, 2007; Scott-Hayward et al., 2014); SR-PDE (see, e.g., Sangalli et al., 2013; Azzimonti et al., 2015; Bernardi et al., 2017), that in turn expands the methods introduced by Ramsay (2002). Within the kriging framework, some proposals to address data distributed over complex planar domains have been advanced by ? and Menafoglio et al. (2018, 2021). Moreover, the stochastic PDE approach introduced by Lindgren et al. (2011) can handle data over irregularly

shaped domains, and the same holds for the intrinsic Gaussian processes in Niu et al. (2019). Soap film smoothing (Wood et al., 2008), the stochastic PDE approach (Bakka et al., 2019) and the kriging technique in **?** comply with some boundary conditions. As detailed in Section 3.2, SR-PDE can handle general forms of boundary conditions (see, e.g., Azzimonti et al., 2014).

As for techniques to fit data on non-planar domains, there is a very active literature concerning random fields on spheres; see, e.g., Gneiting (2013); Castruccio and Stein (2013); Jeong and Jun (2015); Porcu et al. (2016, 2018) and references therein. Moreover, the methods proposed by Lindgren et al. (2011) and by Niu et al. (2019) can deal with more general forms of non-planar domains. Regularized least-square smoothing methods able to fit data scattered on spheres include Wahba (1981); Baramidze et al. (2006); Lai et al. (2009), while Duchamp and Stuetzle (2003) and SR-PDE (Ettinger et al., 2016; Lila et al., 2016; Wilhelm et al., 2016) work on general non-planar domains. Finally, other smoothing methods over general non-planar domains include nearest-neighborhood techniques (see, e.g., Hagler et al., 2006) and heat-kernel smoothing (see, e.g., Chung et al., 2005, 2017, and references therein).

# 3 Spatial regression with differential regularization

This section introduces SR-PDE in its most basic formulation, while subsequent sections present various extensions. Consider $n$ locations $\mathbf{p}_1, \ldots, \mathbf{p}_n$ over a spatial domain $\Omega \subset \mathbb{R}^2$, with boundary $\partial\Omega \in \mathcal{C}^2$. At location $\mathbf{p}_i$ we observe a variable of interest $z_i \in \mathbb{R}$, and possibly also a set of covariates $\mathbf{w}_i \in \mathbb{R}^q$. We consider the following semiparametric generalized additive model:

$$z_i = \mathbf{w}_i^t \boldsymbol{\beta} + f(\mathbf{p}_i) + \epsilon_i, \qquad i = 1, \ldots, n \tag{1}$$

where $\boldsymbol{\beta} \in \mathbb{R}^q$ is an unknown vector of regression coefficients, that describes the effect of the covariates on the mean of the variable of interest, $f : \mathcal{D} \to \mathbb{R}$ is an unknown deterministic field, that captures the spatial structure of the phenomenon under study, and $\epsilon_1, \ldots, \epsilon_n$ are uncorrelated errors, with zero mean and finite variance. In Sangalli et al. (2013) we propose to estimate the unknown $\boldsymbol{\beta}$ and $f$ by minimizing the regularized sum-of-square-error functional

$$\sum_{i=1}^{n} \left(z_i - \mathbf{w}_i^\top \boldsymbol{\beta} - f(\mathbf{p}_i)\right)^2 + \lambda \int_\Omega \left(\Delta f\right)^2 d\mathbf{p} \tag{2}$$

where $\lambda$ is a positive smoothing parameter and $\Delta$ denotes the Laplace operator. To define this partial differential operator, consider a generic function $f : \Omega \to \mathbb{R}$. Denote the gradient of $f$ by

$$\nabla f(\mathbf{p}) := \left(\frac{\partial f}{\partial p_1}(\mathbf{p}), \frac{\partial f}{\partial p_2}(\mathbf{p})\right)^\top$$

where $^\top$ is the transpose operator. Moreover, given a vector field $\mathbf{f} = \left(f_1(\mathbf{p}), f_2(\mathbf{p})\right)^\top : \Omega \to \mathbb{R}^2$, define the divergence of the vector field as

$$\operatorname{div} \mathbf{f}(\mathbf{p}) := \frac{\partial f_1}{\partial p_1}(\mathbf{p}) + \frac{\partial f_2}{\partial p_2}(\mathbf{p}).$$

Then, the Laplacian of $f$ is defined as

$$\Delta f(\mathbf{p}) := \operatorname{div} \nabla f(\mathbf{p}) = \frac{\partial^2 f}{\partial p_1^2}(\mathbf{p}) + \frac{\partial^2 f}{\partial p_2^2}(\mathbf{p}). \tag{3}$$

The Laplace operator provides a simple measure of the local curvature of the field $f$. Higher values of the smooothing parameter $\lambda$ result in smoother estimates of $f$, while lower values of $\lambda$ lead to more data-adapted estimates. It should be noticed that the Laplacian is invariant with respect to rigid transformations (rotations, translations and reflections) of the spatial coordinates; this

ensures that the resulting smoothness of the estimate is independent of the arbitrary orientation of the coordinate system.

The functional (2) is well defined for $\boldsymbol{\beta} \in \mathbb{R}^q$ and $f \in H^2(\Omega)$, where $H^2(\Omega)$ is the Sobolev space of functions $h : \Omega \to \mathbb{R}$ such that $h$ and its first and second derivatives are in $L^2(\Omega)$; see, e.g., Rudin (1991).

When covariates are not available, model (1) is replaced by

$$z_i = f(\mathbf{p}_i) + \epsilon_i, \qquad i = 1, \ldots, n$$

and $f$ can be estimated by minimizing (2), or another of the estimation functionals considered in the following sections, but omitting $-\mathbf{w}_i^\top \boldsymbol{\beta}$ in the least-square term. In the following we consider the more general model that includes covariates, briefly commenting on simplifications for the model without covariates.

## 3.1 Modelling problem-specific information via a partial differential equation

Assume that some problem-specific information is available, coming for instance from the physics, mechanics, chemistry or morphology of the phenomenon under study, and that this information can be formalized in terms of a PDE $Lf = u$. Specifically, on the base of the problem-specific information, we assume that the misfit $Lf - u$ is small, though we do not require it to be null. Then, it makes sense to estimate the unknown $\boldsymbol{\beta}$ and $f$ minimizing the functional

$$\sum_{i=1}^n \left( z_i - \mathbf{w}_i^\top \boldsymbol{\beta} - f(\mathbf{p}_i) \right)^2 + \lambda \int_\Omega \left( Lf - u \right)^2 d\mathbf{p} \tag{4}$$

that trades-off a data fidelity criterion, the least-square term, and a model-fidelity criterion, the misfit with respect to the PDE (see Azzimonti et al., 2015, 2014).
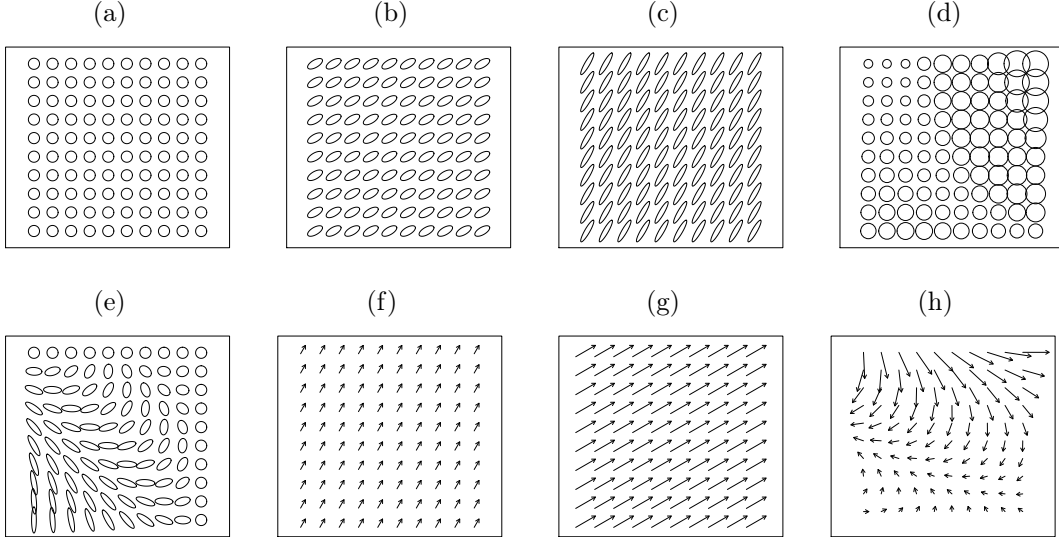


Figure 3: (a): Isotropic diffusion field. (b) and (c): anisotropic diffusion fields with different directions and intensities of the anisotropy. (d): non-stationary isotropic diffusion field. (e) non-stationary anisotropic diffusion field. (f) and (g): transport fields with different directions and intensities of the transport. (h): non-stationary transport field.

The regularizing term enables a very rich modelling of spatial variation. In particular, $L$ is a general linear differential operator, including, for instance, second order, first order and zero order

differential operators. Consider a symmetric and positive definite matrix $K = \{K_{ij}\} \in \mathbb{R}^{2 \times 2}$, named diffusion tensor, a vector $\mathbf{b} = \{b_j\} \in \mathbb{R}^2$, named transport vector, and a positive scalar $c \in \mathbb{R}^+$, named reaction term. The differential operator $L$ can include: a second order differential operator such as the divergence of the gradient, i.e.,

$$\text{div}(K\nabla f) = \frac{\partial}{\partial p_1}\left(K_{11}\frac{\partial f}{\partial p_1} + K_{12}\frac{\partial f}{\partial p_2}\right) + \frac{\partial}{\partial p_2}\left(K_{21}\frac{\partial f}{\partial p_1} + K_{22}\frac{\partial f}{\partial p_2}\right),$$

a first order differential operator such as the gradient, i.e.,

$$\mathbf{b} \cdot \nabla f = b_1 \frac{\partial f}{\partial p_1} + b_2 \frac{\partial f}{\partial p_2},$$

and a zero order differential operator such as $cf$. We thus consider general differential operators of the form

$$Lf = -\text{div}(K\nabla f) + \mathbf{b} \cdot \nabla f + cf. \tag{5}$$

Furthermore, $K$, $\mathbf{b}$ and $c$ can vary over space, i.e., $K = K(\mathbf{p})$, $\mathbf{b} = \mathbf{b}(\mathbf{p})$ and $c = c(\mathbf{p})$. The three operators composing (5) model various forms of anisotropy and non-stationarity. The diffusion term $-\text{div}(K\nabla f)$ produces a smoothing in all the directions. When $K$ is a multiple of the identity matrix $I$, this term induces an isotropic smoothing effect; otherwise it induces an anisotropic smoothing with a preferential direction identified by the first eigenvector of $K$, with a degree of anisotropy controlled by the ratio between the two eigenvalues of $k$. Figure 3 visualizes the diffusion term via ellipses whose axes are oriented according to the eigenvectors of $K$ and with lengths proportional to the corresponding eigenvalues: panel (a) visualizes an isotropic and stationary diffusion; panels (b) and (c) provide two examples of anisotropic diffusion, with different directions and intensities of the anisotropy; panel (d) shows a non-stationarity isotropic case and panel (e) a non-stationarity anisotropic case. The transport term $\mathbf{b} \cdot \nabla f$ produces a smoothing toward a unique direction, identified by $\mathbf{b}$. This is visualized in Figure 3: panels (f) and (g) display two transport fields with different directions and intensities; panel (h) presents a non-stationary transport field. The reaction term $cf$ induces shrinkage of the field to zero; this effect can as well be non-stationary.

The forcing term $u \in L^2(\Omega)$ can either be a null function $u = 0$ (so-called homogeneous case), or $u \neq 0$ (non-homogeneous case), even further increasing the flexibility in the modelling of spatial variation. In the following sections, for simplicity of exposition, we consider the homogeneous case, and only briefly comment on changes required by non-homogeneous forcing terms. We refer the reader to Azzimonti et al. (2014, 2015); Arnone et al. (2019) for details on handling non-homogeneous forcing terms.

For $K = I$, $\mathbf{b} = \mathbf{0}$, $c = 0$ and $u = 0$ we get the special case corresponding to (2), where the regularization with the Laplacian $\Delta f$ implies a stationary and isotropic smoothing.

It should be pointed out that we are not here interested in looking for the PDE solution that is closest to the data, as done for instance with an analogous approach by Xun et al. (2013). In fact, we do not make the assumption that the unknown field satisfies the PDE. Rather, we use the PDE as a powerful tool to model spatial variation, as described above.

## 3.2  Boundary conditions

Moreover, when the domain $\Omega$ has a boundary $\partial\Omega$, there can be problem-specific conditions that $f$ satisfies at $\partial\Omega$. SR-PDE handles different types of boundary conditions, including: Dirichlet conditions, that fix the value of $f$ at $\partial\Omega$; Neumann conditions, that fix the value of its normal derivative, hence governing the flow across $\partial\Omega$; Robin conditions, that are linear combinations of the above conditions. Moreover, different conditions may be imposed on different portions, $\Gamma_D$, $\Gamma_N$, $\Gamma_R$, of $\partial\Omega$, yielding a partition of the boundary. Denoting by $\boldsymbol{\nu}$ the outward unit normal vector to $\partial\Omega$, we can summarize the admissible boundary conditions as

$$\begin{cases} f = \gamma_D & \text{on } \Gamma_D, \\ K\nabla f \cdot \boldsymbol{\nu} = \gamma_N & \text{on } \Gamma_N, \\ K\nabla f \cdot \boldsymbol{\nu} + \chi f = \gamma_R & \text{on } \Gamma_R, \end{cases}$$
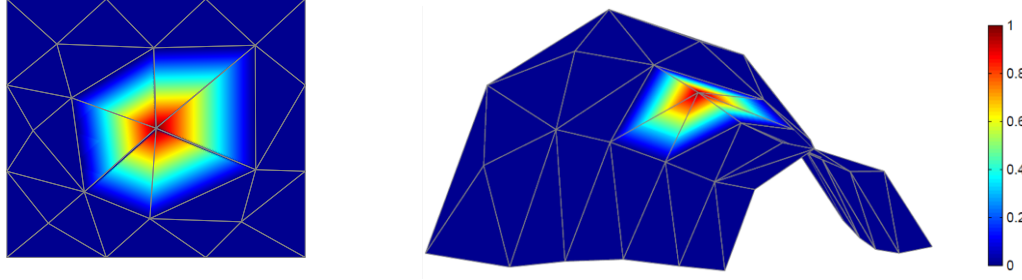
7

Figure 4: Linear finite element basis on planar triangulation (left) and on non-planar triangulation (right).

where the functions $\gamma_D$, $\gamma_N$, $\gamma_R$, and the portions of the boundary $\Gamma_D$, $\Gamma_N$, and $\Gamma_R$ must satisfy some regularity conditions to ensure that the estimation functional is well-defined (Azzimonti et al., 2014). The boundary conditions are said homogeneous when the functions $\gamma_D$, $\gamma_N$, $\gamma_R$ are constant zero functions. For simplicity of exposition, in the following sections we consider homogeneous Dirichlet or Neumann conditions; see Azzimonti et al. (2014) for details on how to handle other boundary conditions.

In several applications the ability to comply with specific boundary conditions is crucial to obtain meaningful estimates. An illustrative example in this respect is given in Section 7.

## 3.3 Characterization of the solution to the estimation problem

The functionals in (2) and (4) are well defined for $\boldsymbol{\beta} \in \mathbb{R}^q$ and $f \in H^2(\Omega)$. Moreover, when the domain is bounded, the use of appropriate boundary conditions guarantees the uniqueness of the solution (see, e.g., Sangalli et al., 2013; Azzimonti et al., 2014, for details). We denote by $V(\Omega)$ the subspace of $H^2(\Omega)$ with the chosen boundary conditions.

Set $\mathbf{z} := (z_1, \ldots, z_n)^\top$, the vector of observed data values, and, for any function $v : \Omega \to \mathbb{R}$, set $\mathbf{v}_n := (v(\mathbf{p}_1), \ldots, v(\mathbf{p}_n))^\top$, the vector of evaluations of $v$ at the $n$ spatial locations. Moreover, if covariates are present, let $W$ denote the $n \times q$ matrix having the $q$ covariates $\mathbf{w}_i^\top$ observed at $\mathbf{p}_i$ in the $i$th row, and assume $W$ has full rank. Moreover, set $Q := I - W(W^\top W)^{-1} W^\top$, the matrix that projects into the orthogonal complement of $\mathbb{R}^n$ with respect to the subspace of $\mathbb{R}^n$ spanned by the columns of $W$.

The following proposition characterizes the solution to the estimation problem, under homogeneous boundary conditions and forcing terms (see Azzimonti et al. (2014) for non-homogeneous boundary conditions and forcing terms).

**Proposition 1.** *There exists a unique pair of estimators* $(\hat{\boldsymbol{\beta}} \in \mathbb{R}^q, \hat{f} \in V(\Omega))$ *which minimize* (4). *Moreover,*

$$\hat{\boldsymbol{\beta}} = (W^\top W)^{-1} W^\top (\mathbf{z} - \hat{\mathbf{f}}_n) \tag{6}$$

*and* $\hat{f}$ *satisfies*

$$\mathbf{v}_n^\top Q \, \hat{\mathbf{f}}_n + \lambda \int_\Omega (Lv)(L\hat{f}) = \mathbf{v}_n^\top Q \, \mathbf{z}, \quad \forall v \in V(\Omega). \tag{7}$$

In the model without covariates, $\hat{f}$ satisfies a problem like (7), but where $Q$ does not appear (or, equivalently, it is replaced by the identity matrix).

# 4 Numerical solution of the estimation problem

The fourth order differential problem (7) cannot be solved analytically. We hence use advanced numerical discretization procedures. Here we briefly review the discretization based on a mixed finite element formulation. Wilhelm et al. (2016) describes instead an alternative discretization by isogeometric analysis with non-uniform rational B-splines.

The numerical discretization reduces the estimation problem to the solution of a linear system. The spatial domain $\Omega$ is represented by an appropriate mesh, and the functions over $\Omega$ are approximated by a finite system of bases defined on this mesh. This permits to consider spatial domains with complex shapes.

## 4.1 Construction of the mesh

Convenient meshes of the spatial domain are obtained by constrained Delaunay triangulation; this is a generalization of the Delaunay triangulation (see, e.g., Hjelle and Dæhlen, 2006) that enables the definition of the boundary of the domain, forcing the required segments into the triangulation. The triangulated domain is denoted by $\Omega_{\mathcal{T}}$, where $\mathcal{T}$ is the set of all the triangles. It is important to note that the mesh nodes can differ from the data locations.

The R package `fdaPDE` (Lila et al., 2020) provides functions to construct the mesh, such as `create.mesh.2D`, that creates the mesh starting from a set of points to be used as triangle vertices and/or a set of edges to be used as boundary. The package also contains functions to refine the mesh, such as `refine.mesh.2D`, whose options allows to specify the maximal allowed triangle area and/or the minimal allowed triangle angle. The right panel of Figure 1 shows a mesh of the municipality of Portland. The triangulation has been obtained starting from the boundary and refining it according to criteria of maximal allowed triangle area and minimal allowed triangle angle. The mesh is able to very well represent this complicated domain, accurately rendering the Willamette river, that cuts through the city, and other detailed features. The package `fdaPDE` also handles non-planar triangular meshes, such as the one in the left panel of Figure 2, that represents the left hemisphere of the cerebral cortex of a template brain.

In general, the mesh should be fine enough to capture the features in the signal. For medium sample sizes, a regular mesh having a number of nodes higher than the number of data, but of the same order of magnitude, typically works efficiently. Triangles having too acute angles should be avoided (setting for instance a minimum angle of 25 degrees in the refinement function), as they may be associated with numerical instability. For data displaying highly localized features, it may be convenient to consider data-driven meshes, that may be constructed starting from (a subsample of) the data locations. Such data-driven meshes permit to capture strongly localized features, while being parsimonious (i.e., using a limited number of mesh nodes) and thus requiring a lower computational cost.

## 4.2 Finite elements

Finite elements provide bases for globally continuous, piecewise polynomial functions; see, e.g., Ciarlet (2002); Quarteroni (2017) for an introduction to finite elements. The finite element space $V_{\mathcal{T}}^r(\Omega)$, with $r \in \{1, 2, \ldots\}$, is defined as the space of continuous functions over $\Omega_{\mathcal{T}}$ that are polynomials of degree $r$ once restricted to any triangle in $\mathcal{T}$. It is convenient to introduce the nodes of the triangulation, $\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_{N_{\mathcal{T}}}$. In the case of linear finite elements, the nodes are given by the vertices of the triangles in $\mathcal{T}$. In the case of quadratic finite elements, the nodes also include the middle points of the triangle edges. For higher orders $r$, the nodes include also other specific points of the triangles. It is then possible to define a set of $N_{\mathcal{T}}$ basis functions $\psi_1, \ldots, \psi_{N_{\mathcal{T}}}$, associated with these nodes, and spanning the finite element space. In particular, for each node $\boldsymbol{\xi}_j, j \in \{1, \ldots, N_{\mathcal{T}}\}$, the basis $\psi_j$ is defined as a locally supported piecewise polynomial function of the desired order $r$, taking value 1 at the node $\boldsymbol{\xi}_j$, and value 0 at all other nodes, i.e., $\psi_j(\boldsymbol{\xi}_i) = \delta_{ji}$, where $\delta_{ji} = 1$ if $j = i$ and $\delta_{ji} = 0$ if $j \neq i$. Figure 4 displays linear finite element bases over a planar and over a non-planar triangulation.

Let $\boldsymbol{\psi} = (\psi_1, \ldots, \psi_{N_\mathcal{T}})^\top$ be the $N_\mathcal{T}$-vector of the basis functions. Then any function $v \in V_\mathcal{T}^r(\Omega)$ can be represented as an expansion on this basis system, i.e., $v(\mathbf{p}) = \mathbf{v}^\top \boldsymbol{\psi}(\mathbf{p})$, where $\mathbf{v} = (v_1, \ldots, v_{N_\mathcal{T}})$ is the vector of coefficients of the basis expansion. It turns out that $\mathbf{v} = (v(\boldsymbol{\xi}_1), \ldots, v(\boldsymbol{\xi}_{N_\mathcal{T}}))^\top$, i.e., the vector of coefficients of the basis expansion is given by the vector of evaluations of $v$ at the $N_\mathcal{T}$ nodes of the mesh. In fact, $v(\boldsymbol{\xi}_i) = \sum_{j=1}^{N_\mathcal{T}} v_j \psi_j(\boldsymbol{\xi}_i) = \sum_{j=1}^{N_\mathcal{T}} v_j \delta_{ij} = v_i$. This property is computationally very convenient.

Homogeneous boundary conditions on $\Gamma_D$ can be simply enforced discarding the finite element bases associated with nodes in $\Gamma_D$.

The function `create.FEM.basis` of `fdaPDE` constructs a finite element basis of the desired order (linear and quadrating finite elements are currently implemented), requiring as imput the mesh.

## 4.3 Solution of the estimation problem using finite elements

Denote by $\Psi$ the $n \times N_\mathcal{T}$ matrix having as entries the evaluations of the $N_\mathcal{T}$ basis functions at the $n$ data locations:

$$\Psi = \begin{bmatrix} \psi_1(\mathbf{p}_1) & \ldots & \psi_{N_\mathcal{T}}(\mathbf{p}_1) \\ \vdots & \vdots & \vdots \\ \psi_1(\mathbf{p}_n) & \ldots & \psi_{N_\mathcal{T}}(\mathbf{p}_n) \end{bmatrix}.$$

Note that $\mathbf{v}_n = (v(\mathbf{p}_1), \ldots, v(\mathbf{p}_n))^\top$, the vector of evaluations of $v$ at the $n$ spatial locations, and $\mathbf{v} = (v(\boldsymbol{\xi}_1), \ldots, v(\boldsymbol{\xi}_{N_\mathcal{T}}))^\top$, the vector of evaluations of $v$ at the $N_\mathcal{T}$ mesh nodes, are related by $\mathbf{v}_n = \Psi \mathbf{v}$. Moreover, let $R_0$ and $R_1$ be the following $N_\mathcal{T} \times N_\mathcal{T}$ matrices:

$$R_0 = \int_{\Omega_\mathcal{T}} \boldsymbol{\psi} \boldsymbol{\psi}^\top$$

$$R_1 = \int_{\Omega_\mathcal{T}} \left( \nabla \boldsymbol{\psi}^\top \mathbf{K} \nabla \boldsymbol{\psi} + \nabla \boldsymbol{\psi}^\top \mathbf{b} \boldsymbol{\psi}^\top + c \boldsymbol{\psi} \boldsymbol{\psi}^\top \right)$$

Proposition 2 states that, with the introduction of the finite element space, the estimation problem becomes a linear system.

**Proposition 2.** *There exists a unique pair of estimators $(\hat{\boldsymbol{\beta}} \in \mathbb{R}^q, \hat{f} \in V_\mathcal{T}(\Omega))$ which solve the discrete counterpart of the estimation problem. Moreover,*

$$\hat{\boldsymbol{\beta}} = (W^\top W)^{-1} W^\top (\mathbf{z} - \hat{\mathbf{f}}_n)$$

*and $\hat{f} = \hat{\mathbf{f}}^\top \boldsymbol{\psi}$, with $\hat{\mathbf{f}}$ satisfying*

$$\begin{bmatrix} -\Psi^\top Q \Psi & \lambda R_1^\top \\ \lambda R_1 & \lambda R_0 \end{bmatrix} \begin{bmatrix} \hat{\mathbf{f}} \\ \mathbf{g} \end{bmatrix} = \begin{bmatrix} -\Psi^\top Q \mathbf{z} \\ \mathbf{0} \end{bmatrix}. \tag{8}$$

The system (8) is of order $2N_\mathcal{T}$, so it can be very large. On the other hand, it is highly sparse: $R_0$ and $R_1$ are in fact highly sparse matrices because most of the cross-products of the basis functions and of their partial derivatives are null, due to local support of the bases. As an example, the triangulation of the municipality of Portland in Figure 1 has 492 nodes and only about 2% of the entries of $R_0$ and $R_1$ are non-zero. For this reason, the solution of (8) is fast.

From Proposition 2, we get that

$$\hat{f} = \hat{\mathbf{f}}^\top \boldsymbol{\psi} \quad \text{with} \quad \hat{\mathbf{f}} = (\Psi^\top Q \Psi + \lambda P)^{-1} \Psi^\top Q \mathbf{z} \tag{9}$$

where the positive definite matrix $P := R_1^\top R_0^{-1} R_1$ discretizes the regularizing term in (4). This approximation of the regularization only involves first order derivatives. In fact, in the mixed finite element approach here considered, the fourth order problem (7) is rewritten as a system of two coupled second order problems; integrating by parts against test functions, these second order problems are hence reformulated in a version that only involves first order derivatives. See, e.g., Azzimonti et al. (2015) for details.

For non-homogeneous forcing terms $u \in L^2(\Omega)$, $u \neq 0$, the vector $\mathbf{0}$ in the right hand side of (8) is replaced by $\lambda \mathbf{u}$, where $\mathbf{u}$ is the discretization of the forcing term $u$, i.e., $\mathbf{u} = (u(\xi_1), \dots, u(\xi_N))^\top$. For the model without covariates, $\hat{f}$ is as in (8) and (9) with $Q = I$.

SR-PDE estimates can be obtained by the function `smooth.FEM` of the package `fdaPDE` (Lila et al., 2020). The option `PDE.parameters` of this function permits to specify the parameters $K$, $\mathbf{b}$, $c$ and $u$ of the PDE in the regularizing term; its default value corresponds to the regularization of the Laplace operator.

# 5 Properties of the estimators

The estimators $\hat{\boldsymbol{\beta}}$ and $\hat{f}$ are linear in the observed data values $\mathbf{z}$. $S$ denote the $n \times n$ matrix $S := \Psi(\Psi^\top Q \Psi + \lambda P)^{-1} \Psi^\top Q$. Then, $\hat{\boldsymbol{\beta}} = (W^\top W)^{-1} W^\top \{I - S\} \mathbf{z}$ and $\hat{\mathbf{f}}_n = S \mathbf{z}$. Assuming that the random errors $\epsilon_1, \dots, \epsilon_n$ in model (1) are uncorrelated, with zero mean and finite constant variance $\sigma^2$, we can readily derive

$$\mathbb{E}[\hat{\boldsymbol{\beta}}] = \boldsymbol{\beta} + (W^\top W)^{-1} W^\top (I - S) \mathbf{f}_n \tag{10}$$

$$Var(\hat{\boldsymbol{\beta}}) = \sigma^2 (W^\top W)^{-1} + \sigma^2 (W^\top W)^{-1} W^\top \{S\,S^\top\} W (W^\top W)^{-1}$$

and

$$\mathbb{E}[\hat{\mathbf{f}}_n] = S\,\mathbf{f}_n \tag{11}$$

$$Var(\hat{\mathbf{f}}_n) = \sigma^2 S\,S^\top.$$

Moreover, consider the field estimator $\hat{f}$ at the generic location $\mathbf{p} \in \Omega$:

$$\hat{f}(\mathbf{p}) = \boldsymbol{\psi}(\mathbf{p})^t (\Psi^t Q \Psi + \lambda P)^{-1} \Psi^t Q \mathbf{z}.$$

Its mean and variance/covariance are given by

$$E[\hat{f}(\mathbf{p})] = \boldsymbol{\psi}(\mathbf{p})^t (\Psi^t Q \Psi + \lambda P)^{-1} \Psi^t Q \mathbf{f}_n \tag{12}$$

$$Var[\hat{f}(\mathbf{p})] = \sigma^2 \boldsymbol{\psi}(\mathbf{p})^t (\Psi^t Q \Psi + \lambda P)^{-1} \Psi^t Q \Psi (\Psi^t Q \Psi + \lambda P)^{-1} \boldsymbol{\psi}(\mathbf{p})$$

$$Cov[\hat{f}(\mathbf{p}_1), \hat{f}(\mathbf{p}_2)] = \sigma^2 \boldsymbol{\psi}(\mathbf{p}_1)^t (\Psi^t Q \Psi + \lambda P)^{-1} \Psi^t Q \Psi (\Psi^t Q \Psi + P)^{-1} \boldsymbol{\psi}(\mathbf{p}_2).$$

for any $\mathbf{p}, \mathbf{p}_1, \mathbf{p}_2 \in \Omega$. As highlighted by the expressions above, the first order structure of the field estimator, the mean of $\hat{f}$, as well as its second order structure, the covariance of $\hat{f}$, depend on the regularization being considered.

A new observation at $\mathbf{p}_{n+1}$, having covariates $\mathbf{w}_{n+1}$, can be predicted by

$$\hat{z}_{n+1} = \mathbf{w}_{n+1}^\top \hat{\boldsymbol{\beta}} + \hat{f}(\mathbf{p}_{n+1}) = \mathbf{w}_{n+1}^\top \hat{\boldsymbol{\beta}} + \hat{\mathbf{f}}^\top \boldsymbol{\psi}(\mathbf{p}_{n+1}),$$

and the mean and variance of this estimator can be analogously derived.

The vector of fitted values at the $n$ data locations $\hat{\mathbf{z}} = W\hat{\boldsymbol{\beta}} + \hat{\mathbf{f}}_n = (I - Q + Q\,S)\,\mathbf{z}$ is also linear in the observations $\mathbf{z}$. The trace of the linear operator that takes from the observed to the fitted values, $tr(I - Q + Q\,S) = q + tr(S)$, can be taken as equivalent degrees of freedom of the model (see, e.g., Buja et al., 1989). The equivalent degrees of freedom are thus given by the sum of $q$, the number of covariates, and $tr(S)$, the equivalent degrees of freedom of the the field estimator. An estimator of $\sigma^2$ is given by

$$\hat{\sigma}^2 = \frac{1}{n - (q + tr(S))} (\mathbf{z} - \hat{\mathbf{z}})^\top (\mathbf{z} - \hat{\mathbf{z}}).$$

It is possible to automatically select the smoothing parameter $\lambda$ using for instance k-fold cross validation or Generalized-Cross-Validation (Craven and Wahba, 7879):

$$GCV(\lambda) = \frac{1}{n\big(1 - (q + tr(S))/n\big)^2} \left(\mathbf{z} - \hat{\mathbf{z}}\right)^\top \left(\mathbf{z} - \hat{\mathbf{z}}\right).$$

Expressions (10), (11) and (12) point out that the estimators are biased. Specifically, $\hat{\boldsymbol{\beta}}$ is biased as a result of the bias affecting $\hat{f}$. The bias of $\hat{f}$ in turn has two sources: the discretization and the regularization. The bias due to the discretization is common to any nonparametric estimator that employs a basis expansion. The bias due to the regularization is common to any penalized regression estimator, and affects also the infinite-dimensional estimators in Proposition 1, not only their discrete counterparts: unless the true function $f$ satisfies exactly the PDE in the regularizing term (hence annihilating the penalty term), this term induces in fact a bias. Simulations show that this bias might be appreciable in small sample scenarios, especially under particularly irregular sampling designs. Azzimonti et al. (2014), considering for simplicity an SR-PDE model without covariates, investigates the convergence of the bias of the finite element estimator, with respect to the size of the triangulation. Arnone et al. (2020), always considering a model without covariates, and assuming other simplifying hypothesis, derives some preliminary results on the consistency of the methods.

SR-PDE has been thoroughly tested under various simulated settings (with and without covariates, with true fields generated parametrically or randomly, with uncorrelated and correlated noise, with various signal-to-noise ratios, sampling sizes and sampling designs), and compared to kriging and to other regularized regression estimators such as those based on thin-plate splines or soap film smoothing (see, e.g., Sangalli et al., 2013; Azzimonti et al., 2014; Bernardi et al., 2017; Ettinger et al., 2016; Bernardi et al., 2018; Arnone et al., 2019). In all the simulation studies, the estimates provided by SR-PDE, in correspondence of the smoothing parameter automatically selected by GCV or by k-fold cross validation, have always shown comparative advantages with respect to the alternatives, especially in the cases of irregularly shaped domains and of fields exhibiting complicated localized features.

# 6  Model extensions

## 6.1  Areal data

Instead of so-called geostatistical data, i.e., data observed at point-wise locations $\mathbf{p}_1, \ldots, \mathbf{p}_n$, we can consider data referred to areal sub-domains of $\Omega$. Specifically, let $D_1, \ldots, D_n$ be $n$ disjoint spatial sub-domains of $\Omega$. In such case, the model must involve areal evaluations of the unknown field $f$. Depending on the data and problem being considered, these areal evaluations can for instance be either mean evaluations or integral evaluations of $f$. For instance, we may assume the model

$$z_i = \mathbf{w}_i^\top \boldsymbol{\beta} + \frac{1}{|D_i|} \int_{D_i} f + \epsilon_i \qquad i = 1, \ldots, n$$

where $|D_i|$ denotes the area of $D_i$, and $\epsilon_1, \ldots, \epsilon_n$ are uncorrelated zero-mean errors with $Var(\epsilon_i) \propto \frac{1}{|D_i|}$. In such case, $\boldsymbol{\beta}$ and $f$ are estimated minimizing the regularized weighted sum-of-square-error functional

$$\sum_{i=1}^{n} |D_i| \left( z_i - \mathbf{w}_i^\top \boldsymbol{\beta} - \frac{1}{|D_i|} \int_{D_i} f \right)^2 + \lambda \int_{\Omega} \left( Lf - u \right)^2$$

See Azzimonti et al. (2015) for details. Areal data can be handled in the function `smooth.FEM` of `fdaPDE`, specifying the option `incidence.matrix` to identify $D_1, \ldots, D_n$ over the triangulated domain.

## 6.2 Generalized linear model

The sum of square error in (2) can be interpreted as a Gaussian (negative) log-likelihood. Indeed, the model in Section 3 can be extended to any distribution in the exponential family, via a Generalized Linear Model framework (see, e.g., Hastie and Tibshirani, 1990; McCullagh and Nelder, 1989). Let $Z_1, \ldots, Z_n$ be independent random variable having a distribution that belongs to the exponential family. In Wilhelm and Sangalli (2016) we model the expected value of $Z_i$, at location $\mathbf{p}_i$ and with covariates $\mathbf{w}_i$ (whenever available), by

$$g\big(\mathbb{E}(Z_i)\big) = \theta_i(\boldsymbol{\beta}, f) = \mathbf{w}_i^\top \boldsymbol{\beta} + f(\mathbf{p}_i)$$

where $g$ is the canonical link function associated with the considered distribution. We thus estimate $\boldsymbol{\beta}$ and $f$ maximizing the penalized log-likelihood

$$\sum_{i=1}^n l(z_i; \theta_i(\boldsymbol{\beta}, f)) - \lambda \int_\Omega \Big(Lf - u\Big)^2 d\mathbf{p} \tag{13}$$

where $l(; \theta)$ is the log-likelihood for the considered distribution. This of course coincides with the minimization of the regularized least-squares in (2) when the considered distribution is Gaussian. This generalization significantly broadens the possible applications of SR-PDE, since most of the (continuous and discrete) well-known distributions belong to the exponential family. The functional in (13) is then optimized through iterative penalized least-squares. We refer the interested reader to Wilhelm and Sangalli (2016) for details. The generalized linear version of SR-PDE can be implemented specifying the option `family` of the function `smooth.FEM` of `fdaPDE`.

## 6.3 Spatio temporal models

SR-PDE can also handle spatio-temporal data and spatially dependent functional data (Delicado et al., 2010; Mateu and Romano, 2017; Menafoglio and Secchi, 2017). To this aim, let $T$ denote the temporal domain of interest and consider the field $f$ defined over $\Omega \times T$. Different sampling designs may be considered, with point-wise or areal observations in space, and point-wise or interval observations in time, leading to different least-square terms (or, equivalently, different likelihood terms in a generalized linear framework). See Arnone et al. (2019) for details. As for the regularization, we can consider two separate terms

$$\lambda_S \int_T \int_\Omega \Big(Lf - u\Big)^2 d\mathbf{p}\,dt + \lambda_T \int_\Omega \int_T \Big(\frac{\partial^2 f}{\partial t^2}\Big)^2 dt\,d\mathbf{p}; \tag{14}$$

the first penalty in (14) regularizes the field in space, likewise in (4), while the second penalty regularizes the field in time, considering the classical regularization of second derivative. In the simple case where $L$ is the Laplace operator, this model is considered by Bernardi et al. (2017). Analogous models are also considered by Marra et al. (2012) and Augustin et al. (2013). Alternatively, as detailed in Arnone et al. (2019), the regularization can be based on the misfit with respect to a time-dependent PDE, jointly modelling the spatio-temporal behavior,

$$\lambda \int_T \int_\Omega \Big(\frac{\partial f}{\partial t} + Lf - u\Big)^2 d\mathbf{p}\,dt. \tag{15}$$

See Arnone et al. (2019) for details. An application of this model is shown in Section 7.

Depending on whether the regularization in (14) or the one in (15) is considered, convenient discretizations of the estimation problem involve finite elements in space, and either splines or finite differences in time. Appropriately stacking the data observed at the $nm$ spatio-temporal locations in the vector $\mathbf{z}$, and correspondingly defining the vector of evaluations of $f$ at the same $nm$ spatio-temporal locations $\mathbf{f}_{nm}$, the covariate matrix $W$ and the matrix $Q = I - W(W^\top W)^{-1} W^\top$, it is possible to obtain the estimators

$$\begin{aligned}
\hat{\boldsymbol{\beta}} &= (W^\top W)^{-1} W^\top (\mathbf{z} - \hat{\mathbf{f}}_{nm}) \\
\hat{\mathbf{f}} &= (\widetilde{\Psi}^\top Q \widetilde{\Psi} + \widetilde{P})^{-1} \widetilde{\Psi}^\top Q \mathbf{z}
\end{aligned}$$

where $\widetilde{\Psi}$ contains the evaluations of the spatial and temporal basis functions at the $nm$ spatio-temporal locations, $\widetilde{P}$ is the discretization of the considered regularizing term, either (14) or (15), also including the corresponding smoothing parameters, and $\hat{\mathbf{f}}$ is the vector of basis expansion coefficients that returns the estimate of the spatio-temporal field. Hence, the estimators have the same form as in Section 3, and it is possible to derive their properties along the same lines as in Section 5. See Bernardi et al. (2017) and Arnone et al. (2019) for details.

Both methods are implemented in the function `smooth.FEM.time` of fdaPDE. This function has similar options as `smooth.FEM`. For instance, the option `PDE.parameters` permits to specify the parameters of the PDE in the regularizing term; the option `incidence.matrix` allows to handle areal data in space.

## 6.4   Non-planar domains

As mentioned in the Introduction, SR-PDE can as well deal with data whose locations lie on non-planar bi-dimensional domains. Specifically, let $\mathcal{M}$ be a general bi-dimensional Riemmanian manifold, and let $\mathbf{p}_1, \ldots, \mathbf{p}_n \in \mathcal{M}$ be the $n$ data locations where the variables $z_1, \ldots, z_n$, and, whenever available, also the covariates $\mathbf{w}_1, \ldots, \mathbf{w}_n$, are observed. The field $f$ is now defined on the manifold domain, $f : \mathcal{M} \to \mathbb{R}$. In this case, the differential operator in the regularizing term must be appropriately referred to the manifold domain $\mathcal{M}$. In particular, in Ettinger et al. (2016); Lila et al. (2016); Wilhelm et al. (2016); Dassi et al. (2015), we consider the Laplace-Beltrami operator, $\Delta_{\mathcal{M}} f$, a generalization of the standard Laplacian to functions $f$ defined on a non-planar domain $\mathcal{M}$. To define $\Delta_{\mathcal{M}} f$ it is necessary to compute the gradient operator $\nabla_{\mathcal{M}}$ and the divergence operator $\text{div}_{\mathcal{M}}$ over $\mathcal{M}$ (see, e.g., Jost, 2017). Then, the Laplace-Beltrami operator is defined as $\Delta_{\mathcal{M}} f = \text{div}_{\mathcal{M}} \nabla_{\mathcal{M}} f$. Alike to the standard Laplacian, $\Delta_{\mathcal{M}} f$ yields a measure of the local curvature of the function $f$, defined on the non-planar domain $\mathcal{M}$. This operator reduces to the Laplace operator in (3) if $\mathcal{M} \subset \mathbb{R}^2$. We hence assume the same semiparametric model in (1) and estimate the unknown $\boldsymbol{\beta}$ and $f$ by minimizing the regularized sum-of-square-error functional as in (2), with the standard Laplacian replaced by the Laplace-Beltrami, i.e.,

$$\sum_{i=1}^{n} \left( z_i - \mathbf{w}_i^\top \boldsymbol{\beta} - f(\mathbf{p}_i) \right)^2 + \lambda \int_{\mathcal{M}} \left( \Delta_{\mathcal{M}} f \right)^2 \, d\mathbf{p}.$$

The Laplace-Beltrami is also invariant with respect to rigid transformations of the coordinates of the non-planar domain. Hence, the use of $\Delta_{\mathcal{M}} f$ in the penalty guarantees that the concept of smoothness is independent of the orientation of $\mathcal{M}$ and of the coordinate system.

Both closed bi-dimensional domains as well as bounded bi-dimensional domains can be considered (for bounded domains, appropriate boundary conditions must be considered to guarantee uniqueness of the solution). The discretization of the estimation problem and the properties of the estimators follow as in the case of planar domains. Lila et al. (2016) describes the discretization via finite elements defined over non-planar triangulations. The left panel of Figure 2 displays a non-planar triangulation and the right panel of Figure 4 provides an example of linear finite element basis on a non-planar triangulation. The function `smooth.FEM` of fdaPDE can handle data over non-planar triangulations. An alternative discretization by isogeometric analysis with non-uniform rational B-splines is given in Wilhelm et al. (2016).

# 7   Illustrative case study: analysis of blood-flow velocity field from eco-color doppler data

We consider a problem that has arisen within the research project Mathematics for Carotid Endarterectomy @ MOX, concerning the analysis of the velocity field of blood flow in sections of carotid arteries, with data obtained from Magnetic Resonance Imaging (MRI) and Echo-Color Doppler (ECD) (see Azzimonti et al., 2015; Arnone et al., 2019, for details). The project ex-

plored the role of vessel morphology and of blood fluid dynamics on the onset and progression of atherosclerotic plaques.
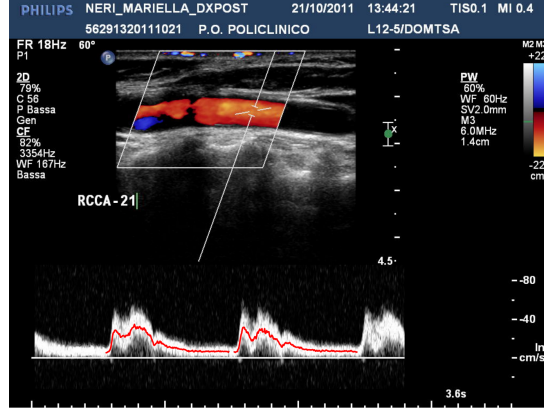


Figure 5: Echo-color doppler acquisition of blood flow velocity over beam 3; see Figures 6 and 8.
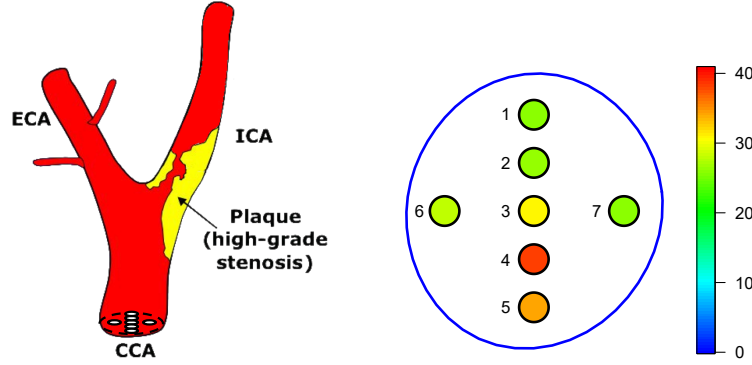


Figure 6: Left: ECD measuraments are taken at a cross-section of the Common Carotid Artery (CCA), 2 cm before the bifurcation of the CCA in Internal Carotid Artery (ICA) and External Carotid Artery (ECA). The ECD signal is acquired in 7 beams, located in the cross shaped pattern shown at the cross-section. Right: Cross-section of the common carotid artery, as reconstructed from MRI data, with the seven beams where the ECD signal is taken. The color correspond to the mean velocity over the beam at systolic peak. The figure also highlights the physiological boundary conditions: the velocity of blood-flow is zero at the arterial wall, due to the friction between the wall and the blood particles.

Figure 5 displays one ECD image from the study. The upper part of this image shows the ultrasound scan of the vessel; the small gray box indicates the beam where the velocity of blood particles is acquired. The lower part of the same image displays the measured blood flow velocity over the time lapse of about three heart-beats. The solid red line superimposed to this signal shows the mean velocity.

The study starts from ECD measurements of the velocity of the blood flow, at a cross-section of the common carotid artery located 2 cm before the bifurcation of the carotid, for patients having a high-grade stenosis ($> 70\%$) at the carotid bifurcation. The left panel of Figure 6 displays a schematic carotid artery: the lower part of this draw shows the cross-section where the ECD measurements are taken; specifically, the ECD signal is acquired over 7 beams displayed in a cross-shaped pattern. The right panel of the same figure shows a zoom of this cross-section: the quasi-circular patient-specific profile of the section is reconstructed from MRI data; the seven beams where the ECD signal is measured are colored according to the mean velocity of blood particles at systolic peak. Figure 8 shows the mean velocity over the seven beams during one

15

heart-beat. The seven temporal profiles have different shapes, some displaying two peaks during the systolic phase, with the highest velocities attained in the second peak, and others displaying only one peak, earlier in time.
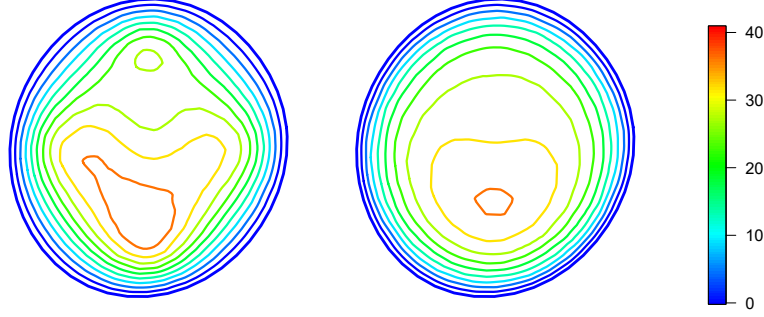


Figure 7: Left: estimate of the blood-flow velocity field at systolic peak, obtained by SR-PDE, imposing the appropriate boundary conditions (i.e.: zero velocity at the arterial wall), but using $L = \Delta$ (stationary isotropic estimate in space). Right: estimate of the blood-flow velocity field at systolic peak, obtained by SR-PDE, imposing the appropriate boundary conditions and using the differential operator $L$ suggested by problem-specific information, with the diffusion and transport fields shown in Figure 9 (non-stationary anisotropic estimate in space).
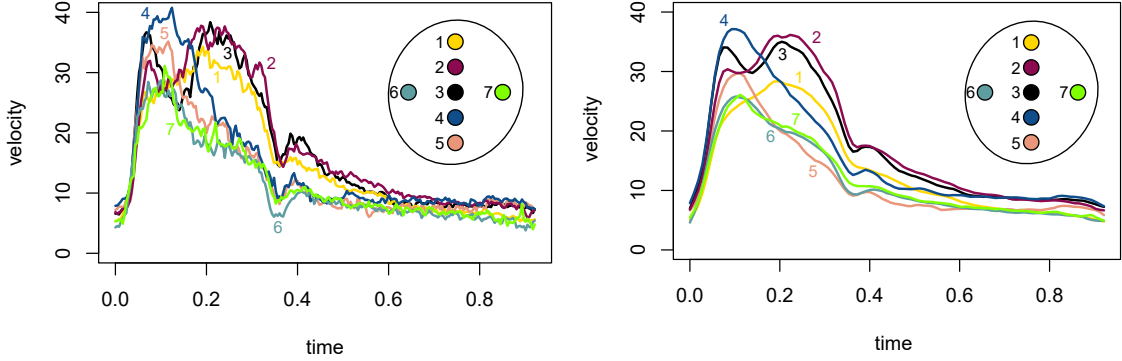


Figure 8: Left. Mean blood-flow velocity recorded over the seven acquisition beams. Different numbers and colors identify the signal over each beam. Right. Temporal profiles of the estimated mean blood-flow velocity, evaluated at the centers of the beams.

The first goal of the project consisted in accurately estimating, for each of the patients involved in the study, the time-dependent blood-flow velocity field over the whole carotid cross-section, starting from the ECD measurements, and from the reconstructions of patient-specific artery sections obtained from MRI data. This application though features some peculiarities, that hinder the applicability techniques for spatio-temporal data and for spatially dependent functional data. Firstly, in this problem the domain shape, the carotid cross section, strongly influences the phenomenon under study: the blood flow is in fact confined within the artery, whose shape heavily reflects on the blood flow velocity field. Secondly, there are specific conditions at the boundaries of this domain: in fact, the velocity of the blood flow is is zero at the arterial wall, due to the friction between the wall and the blood particles. Thirdly, the irregular pattern of the observations, shown in the right panel of Figure 6, misleads stationary and isotropic methods, that yield non-physiological estimates. Figure 7 illustrates this issue, displaying the estimate of the blood-flow velocity field at systolic peak that is obtained by SR-PDE, imposing the appropriate boundary condition and using a regularizing term that involves the stationary and isotropic differential op-
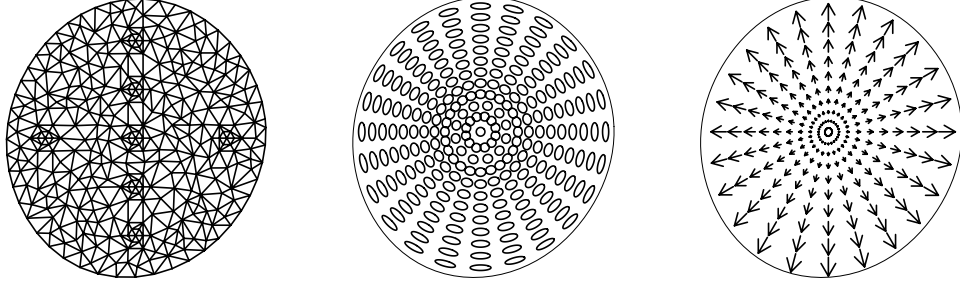
Figure 9: Left: Triangulation of the patient-specific cross-section of the carotid. Center: non-stationary anisotropic diffusion tensor field $K$, that induces a smoothing along concentric circles. Right: non-stationary transport field $\mathbf{b}$, that induces a smoothing from the center of the artery section to the arterial wall.
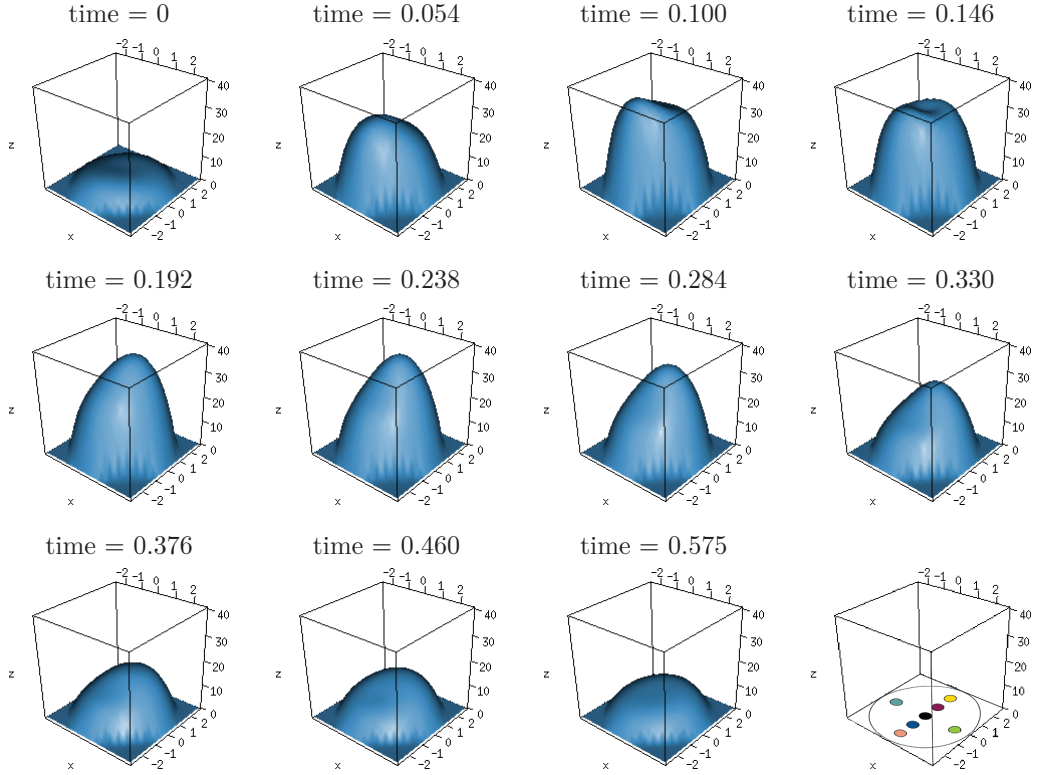


Figure 10: Estimate of the velocity field of the blood flow obtained by SR-PDE imposing the appropriate boundary conditions and using the differential operator $L$ suggested by problem-specific information, with the diffusion and transport fields shown in Figure 9. The bottom right panel displays the orientation of the artery cross-section.

erator $L = \Delta$: penalizing the local curvature of the field pull the solution toward a plane where data are missing, returning non-physiological estimates with rhomboidal isolines.

On the other hand, we can here take advantage of an advanced problem-specific information about the phenomenon under study, concerning fluid dynamics and hemodynamics (see, e.g., Formaggia et al., 2009). This information can be appropriately formalized into a PDE, along with the physiological boundary conditions. Specifically, we can consider the parabolic PDE $\partial f / \partial t + L f = 0$, where $L$ includes a nonstationary anisotropic diffusion tensor that induces a smoothing along concentric circles (Figure 9, center) and a nonstationary transport field that induces a smoothing in the radial direction, from the center of the artery to the arterial wall (Figure 9, right); the reaction term and the forcing term are instead not required by this application (see Azzimonti et al., 2015; Arnone et al., 2019, for the details). The right panel of Figure 7 displays the estimate of the blood-flow velocity field at systolic peak, obtained by SR-PDE, imposing the appropriate boundary condition and with a regularizing term that involves the non-stationary anisotropic differential operator $L$ illustrated above. This estimate efficiently incorporates the problem-specific information and yields a realistic estimate, that displays physiological isolines, unaffected by the cross-shaped pattern of the observations. Figure 10 shows the estimate at some temporal instants. The velocity field strongly varies in shape during the heartbeat, exhibiting asymmetries and eccentricities. These features are of particular interest to the medical doctors, that investigate their relationship with atherosclerosis. The left panel in Figure 8 shows the temporal profiles of the estimated velocity, evaluated at the central point of each beam. A visual comparison with the original data in the right panel of the same figure highlights that the estimate captures very well the main features of the ECD signals.

# 8 Population studies: a principal component analysis method based on SR-PDE

In this section we consider population studies. Consider a sample of $m$ realizations of the field, $\mathbf{z}_1, \ldots, \mathbf{z}_m$, corresponding to $m$ statistical units, where $\mathbf{z}_j := (z_{j1}, \ldots, z_{jn})^\top$ and $z_{ji}$ is the value of the variable of interest observed for the $j$-th statistical unit at location $\mathbf{p}_i$. For simplicity of exposition, we assume that the $m$ realizations of the signal are observed at the same locations $\mathbf{p}_1, \ldots, \mathbf{p}_n$; later we shall remove such assumption. We are now interested in exploring the variability in the sample. Zhou and Pan (2014) proposes a method for Principal Component Analysis (PCA) of data spatially distributed over planar domains with irregular shape; the authors in particular consider a mixed effect model where the principal components are represented by splines over triangulations. Here we review the method described in Lila et al. (2016), that is instead based on a regularized low rank approximation of PCA, and leverages directly on SR-PDE. The method is implemented in the function `FPCA.FEM` of fdaPDE; it can tackle signals over complicated domains, as those considered in the previous sections.

To this aim, we formulate PCA in a functional data analysis framework. Denote by $\mathcal{D}$ the spatial domain of interest, either planar or curved. Consider the space $L^2(\mathcal{D}) = \{f : \mathcal{D} \to \mathbb{R} \text{ s.t. } \int_{\mathcal{D}} f^2 < \infty\}$, of square integrable functions on $\mathcal{D}$, , equipped with its standard inner product $\langle f, g \rangle = \int_{\mathcal{D}} f\, g$ and norm $\|f\|^2 = \int_{\mathcal{D}} f^2$, where $f, g \in L^2(\mathcal{D})$. Consider the random field $Z(\cdot)$ taking values in $L^2(\mathcal{D})$, with mean $\mu(\cdot) = \mathbb{E}[Z(\cdot)]$ and finite second moment, i.e., $\int_{\mathcal{D}} \mathbb{E}(Z^2) d\mathbf{p} < \infty$; moreover assume that its covariance $\Sigma(\mathbf{p}, \mathbf{q}) = \mathbb{E}[(Z(\mathbf{p}) - \mu(\mathbf{p}))(Z(\mathbf{q}) - \mu(\mathbf{q}))]$ is square-integrable. Mercer's Lemma (Riesz and Sz.-Nagy, 1990) ensures that there exist a non-increasing sequence of eigenvalues $\{\zeta_j\}_{j=1,2,\ldots}$ of $\Sigma$, and an orthonormal sequence of corresponding eigenfunctions $\{\mathsf{f}_j\}_{j=1,2,\ldots}$ such that

$$\int_{\mathcal{D}} \Sigma(\mathbf{p}, \mathbf{q}) \mathsf{f}_j(\mathbf{p}) d\mathbf{p} = \zeta_j \mathsf{f}_j(\mathbf{q}) \qquad \forall \mathbf{q} \in \mathcal{D} \tag{16}$$

and that $\Sigma(\mathbf{p}, \mathbf{q}) = \sum_{j=1}^{\infty} \zeta_j \mathsf{f}_j(\mathbf{p}) \mathsf{f}_j(\mathbf{q})$ for each $\mathbf{p}, \mathbf{q} \in \mathcal{D}$. Hence, the random field $Z$ can be expressed as $Z(\cdot) = \mu(\cdot) + \sum_{j=1}^{\infty} u_j \mathsf{f}_j(\cdot)$, where $u_j = \langle Z - \mu, \mathsf{f}_j \rangle$ are uncorrelated random variables, named scores. This is the so-called Karhunen-Loève expansion of $Z$.

The orthonormal functions $f_1, f_2, \ldots \in L^2(\mathcal{D})$ are named principal components of the random variable $Z$. These functions identify the strongest modes of variations of $Z$. Specifically, $f_1$ solves the following maximization problem

$$f_1 = \underset{f : \|f\| = 1}{\mathrm{argmax}} \quad Var \langle Z, f \rangle$$

while $f_d$, for $d > 1$, solves an analogous maximization problem, but with the added constraint of being orthogonal to previous principal components, i.e.,

$$f_d = \underset{\substack{f : \|f\| = 1 \\ \langle f, f_j \rangle = 0\ \forall j < d}}{\mathrm{argmax}} \quad Var \langle Z, f \rangle.$$

Now, suppose $m$ realizations $z_1(\cdot), \ldots, z_m(\cdot)$ of $Z(\cdot)$ were available, observed for any $\mathbf{p} \in \mathcal{D}$ and without error. Then we could compute the sample mean $\hat{\mu}(\mathbf{p}) = \frac{1}{m} \sum_{j=1}^m z_j(\mathbf{p})$ and sample covariance $\hat{\Sigma}(\mathbf{p}, \mathbf{q}) = \frac{1}{m} \sum_{j=1}^m [(z_j(\mathbf{p}) - \hat{\mu}(\mathbf{p}))(z_j(\mathbf{q}) - \hat{\mu}(\mathbf{q}))]$, for any $\mathbf{p}, \mathbf{q} \in \mathcal{D}$, and we could find the estimates $\hat{f}_1, \hat{f}_2, \ldots, \hat{f}_d$ of the first $d$ principal components solving (numerically) the eigenvalue problem (16), but with the true covariance replaced by the sample covariance, i.e.

$$\int_{\mathcal{D}} \hat{\Sigma}(\mathbf{p}, \mathbf{q}) \hat{f}_j(\mathbf{p}) d\mathbf{p} = \hat{\zeta}_j \hat{f}_j(\mathbf{q}) \qquad \forall \mathbf{q} \in \mathcal{D}. \tag{17}$$

Unfortunately, it is impossible to have realizations of the random field $Z$ observed for any $\mathbf{p} \in \mathcal{D}$ and without noise; only discrete and noisy realizations of $Z$ can be observed. The classical approach in functional data analysis is thus the following: first, smooth estimates of each single realization of the signal, $\hat{z}_1(\cdot), \ldots, \hat{z}_m(\cdot)$, are obtained, through some appropriate smoothing procedure; then, these estimates are used to compute the sample mean $\hat{\mu}$ and sample covariance $\hat{\Sigma}$, and hence the estimates $\hat{f}_1, \hat{f}_2, \ldots, \hat{f}_d$ of the first $d$ principal components are obtained solving numerically problem (17). This is the so-called pre-smoothing approach.

Instead, in Lila et al. (2016) we propose an alternative method for the estimation of the principal components, inspired by approaches based on regularized functional PCA (fPCA), developed for the case of functional data over one-dimensional domains by, e.g., Rice and Silverman (1991); Silverman (1996); Huang et al. (2008). The proposed approach relies on another characterization of principal components, known as best $\kappa$ basis approximation property. Namely, for any $\kappa = 1, 2, \ldots$, the first $\kappa$ principal components satisfy

$$(f_j)_{j=1}^\kappa =$$

$$\underset{(\{f_j\}_{j=1}^\kappa : \langle f_j, f_l \rangle = \delta_{jl})}{\mathrm{argmin}} \mathbb{E} \int_{\mathcal{D}} \left\{ Z - \mu - \sum_{j=1}^\kappa \langle Z, f_j \rangle f_j \right\}^2 \tag{18}$$

If we consider $\kappa = 1$ and we take data already centered around the mean, the empirical version of the minimization functional in (18) is given by

$$\frac{1}{n} \frac{1}{m} \sum_{j=1}^m \sum_{i=1}^n \left( z_{ji} - u_j f(\mathbf{p}_i) \right)^2 \tag{19}$$

that involves row data, without requiring any presmoothing. We can thus estimate the first principal component and associated score minimizing a regularized sum-of-square-error functional, involving the least-square criterion (19) and a regularizing term that encourages smoothness of $f$. This regularizing term can for instance involve $\Delta_{\mathcal{D}} f$, where $\Delta_{\mathcal{D}}$ denotes the Laplace or Laplace-Beltrami operator, depending on whether $\mathcal{D}$ is a planar or non-planar domain. This suggests to estimate the first principal component function $f$ and the associated score vector $\mathbf{u} = (u_1, \ldots, u_m)^\top$ by minimizing

$$\sum_{j=1}^m \sum_{i=1}^n \left( z_{ji} - u_j f(\mathbf{p}_i) \right)^2 + \lambda \int_{\mathcal{D}} (\Delta_{\mathcal{D}} f)^2 \tag{20}$$

over $\mathbf{u} \in \mathbb{R}^m$ such that $\mathbf{u}^\top \mathbf{u} = 1$, and $f \in H^2(\mathcal{D})$, with boundary conditions in case $\mathcal{D}$ has boundaries. In particular, setting the constraint $\mathbf{u}^\top \mathbf{u} = 1$ makes the representation unique, without the need to impose that $f$ has unitary norm. Subsequent principal component functions are extracted sequentially by removing previous principal components from the data matrix $\mathbf{Z}$, whose $j$-th row is given by $\mathbf{z}_j$.

The minimization of (20) can be performed via an iterative algorithm that alternates the estimation of the scores $\mathbf{u}$ given the principal component f, and the estimation of the principal component f given the scores $\mathbf{u}$. Specifically, the two steps of the algorithm are as follows:

Step 1: $\mathbf{u}$ given f. The unitary norm vector $\mathbf{u}$ that minimizes the functional in (20) for a fixed $f$ is given by

$$\mathbf{u} = \frac{\mathbf{Z}\mathbf{f}_n}{\|\mathbf{Z}\mathbf{f}_n\|}$$

where $\mathbf{f}_n = \big(\mathsf{f}(\mathbf{p}_1)\ldots,\mathsf{f}(\mathbf{p}_n)\big)^\top$ constitutes the loading vector. The expression for the scores is thus analogous to the one obtained in the multivariate setting.

Step 2: f given $\mathbf{u}$. For a fixed $\mathbf{u}$ with unitary norm, let $y_i$ denote the $i$-th entry of the vector $\mathbf{y} = \mathbf{Z}^\top \mathbf{u}$. Then, finding $f$ that minimizes (20), for the fixed $\mathbf{u}$, is equivalent to finding $f$ that minimizes

$$\sum_{i=1}^{n} \big(y_i - f(\mathbf{p}_i)\big)^2 + \lambda \int_{\mathcal{D}} (\Delta_{\mathcal{D}} f)^2.$$

This is the tipical SR-PDE problem (without covariates), that has been described in detail in the previous sections.

The principal components and scores can then be rescaled to obtain principal components with unitary norms.

Missing data can as well be considered. Specifically, assume that the j-th statistical unit $\mathbf{z}_j := (z_{j1}, \ldots, z_{jn_j})^\top$, is observed at locations $\mathbf{p}_{j1}, \ldots, \mathbf{p}_{jn_j}$. Then, the least-square term in (20) can be replaced by

$$\sum_{j=1}^{m} \sum_{i=1}^{n_j} \big(z_{ji} - u_j f(\mathbf{p}_{ji})\big)^2$$

while the regularizing term is left unchanged. Similar sparse designs can be considered for the spatio-temporal problems in Section 6.3.

# 9 Illustrative case study: analysis of neural connectivity on the cortical surface

This section illustrates a case study concerning the analysis of high-dimensional neuroimaging signals associated with neural activity in the cerebral cortex. The main goal is to explore the main neural connectivity patterns in a population of healthy subjects. The analyzed data come from the Human Connectome Project (Essen et al., 2012), and are obtained from resting-state functional Magnetic Resonance Imaging (fMRI) on 491 subjects. The fMRI signal measures neural activity through changes in the concentration of deoxy-hemoglobin, associated with energy use by neural cells. As standard practice in the neuroimaging community, the signal of each individual is mapped to a common template of cortex, to enable multi-subject analyses. Figure 2 shows a triangular mesh representing the cortical surface of a template brain, where the signals of all subjects are projected. The figure highlights the highly convoluted morphology of the cortex. Still today, the most part of neuroimaging studies are carried out either neglecting the spatial dependence in the signal, or employing standard methods that rely on the Euclidean distance, hence disregarding the complicated morphology of the cortex. As mentioned in the Introduction, this may nonetheless lead to inaccurate estimates. There is nowadays a growing awareness of the need to include the complex brain morphology, to advance our still limited knowledge about brain functioning (see, e.g., Glasser et al., 2013, and references therein). This has generated a strong momentum in the

international community for the development of methods able to accurately analyze data arising from these complex imaging scans. Lila et al. (2016) offers in this respect a first technique for population studies of signals referred to bi-dimensional manifold domains.
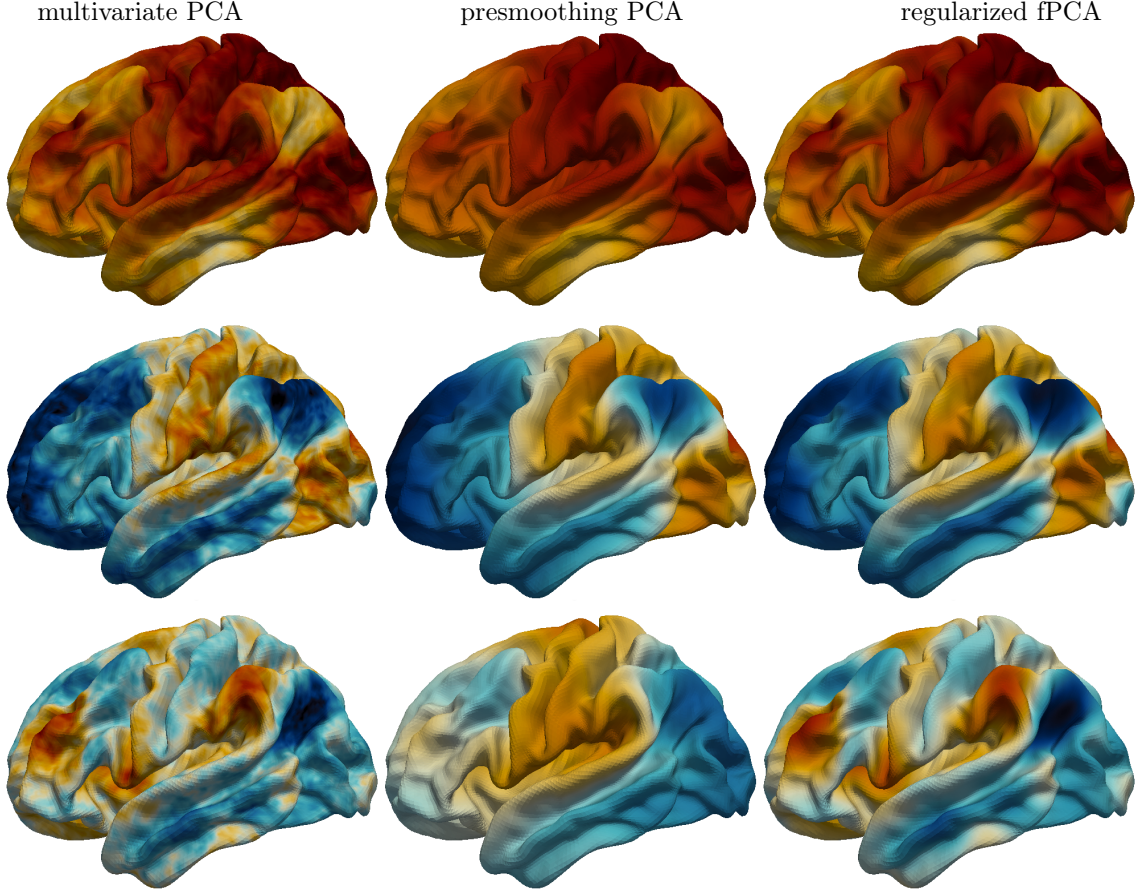
| multivariate PCA | presmoothing PCA | regularized fPCA |
|---|---|---|



Figure 11: Estimates of the first (top), second (center) and third (bottom) principal components, obtained by multivariate PCA (left), pre-smoothing followed by multivariate PCA (center), and the proposed fPCA based on SR-PDE (right).

The analysis is focused on functional connectivity maps, computed from fMRI data. In particular, we select a region of interest on the cortex template. This is the area whose behaviour, as compared to the rest of the cortical surface, is of interest for the investigator. In the current study we select a region of interest in the precuneus, an area of the cerebral cortex devoted to higher mental processes. For each subject we then compute the functional connectivity map with respect to the region of interest (see, e.g., Lila et al., 2016); this map highlights the areas of the cortex that are more highly connected to the selected region of interest. A snapshot of the functional connectivity map for the chosen region of interest, for one subject in the study, is shown in the right panel of Figure 2.

We now wish to explore the main modes of variation of these functional connectivity maps among the different subjects. The last colum of Figure 11 shows the first three principal components estimated by the proposed regularized fPCA, and compares the estimates to those obtained by standard multivariate PCA on the data matrix $\mathbf{Z}$, and to those obtained by the classical pre-smoothing approach, respectively shown in the first column and second column of the same figure. The smoothing parameters in the proposed regularized fPCA and in the presmoothing approach are chosen via cross-validation. As highlighted by Figure 11, the three principal components estimated by multivariate PCA display a strong local variability. This is due to the fact that the sample

size is too small with respect to the very high dimensionality of the data; moreover, multivariate PCA completely disregard the morphology of the domain, since spatial information is ignored by this approach. Both the presmoothing approach and the proposed regularized fPCA instead return smooth principal components. On the other hand, the presmoothing approach, smooth out sharpe changes in the modes of variations, returning estimates that miss some localized features, apparent in the estimates provided by multivariate PCA and by regularized fPCA. As an example, the third principal components estimated by multivariate PCA and by regularized fPCA display corresponding localized areas with high values (displayed in red) or low values (displayed in blue), that are instead missing in the corresponding estimate yielded by presmoothing PCA. By contrary, the presmoothing approach appears to introduce some artifacts: for instance the third principal component estimated by presmoothing PCA displays high values in the upper part of the plot, that do not find matches in the multivariate PCA and regularized fPCA estimates. In summary, the proposed regularized fPCA is able to combine the desired smoothness with the capacity to represent highly localized features of the modes of variation.

## 10    Discussion

Although the study of the theoretical properties of SR-PDE is still ongoing, and much remains to be done in this respect, this methodology has already shown to be very promising, and capable to handle complex applied problems. Moreover, its flexible framework permits numerous models extensions, for instance considering higher dimensional domains (such as three-dimensional spatial domains with complicated shapes) and density estimation problems (Ferraccioli et al., 2021).

As highlighted in this article, SR-PDE is based on a very rich blend of approaches from pure and applied mathematics, and from engineering. This endows the methods with remarkable advantages with respect to the available techniques and makes them able to accurately deal with data structures for which the classical methods are unfit. Moreover, leveraging on advanced numerical analysis techniques ensures that the methods are computational highly efficient. We are sure that SR-PDE will gain popularity, proving highly valuable in a number of applied domains.

## References

Arnone, E., Azzimonti, L., Nobile, F., and Sangalli, L. M. (2019). Modeling spatially dependent functional data via regression with differential regularization. *J. Multivariate Anal.*, 170:275–295.

Arnone, E., Kneip, A., Nobile, F., and Sangalli, L. M. (2020). Some first results on the consistency of spatial regression with partial differential equation regularization. *Statistica Sinica*, page DOI: 10.5705/ss.202019.0346.

Augustin, N. H., Trenkel, V. M., Wood, S. N., and Lorance, P. (2013). Space-time modelling of blue ling for fisheries stock management. *Environmetrics*, 24(2):109–119.

Azzimonti, L., Nobile, F., Sangalli, L. M., and Secchi, P. (2014). Mixed finite elements for spatial regression with PDE penalization. *SIAM/ASA J. Uncertain. Quantif.*, 2(1):305–335.

Azzimonti, L., Sangalli, L. M., Secchi, P., Domanin, M., and Nobile, F. (2015). Blood flow velocity field estimation via spatial regression with PDE penalization. *J. Amer. Statist. Assoc.*, 110(511):1057–1071.

Bakka, H., Vanhatalo, J., Illian, J. B., Simpson, D., and Rue, H. v. (2019). Non-stationary Gaussian models with physical barriers. *Spat. Stat.*, 29:268–288.

Baramidze, V., Lai, M. J., and Shum, C. K. (2006). Spherical splines for data interpolation and fitting. *SIAM J. Sci. Comput.*, 28(1):241–259.

Bernardi, M. S., Carey, M., Ramsay, J. O., and Sangalli, L. M. (2018). Modeling spatial anisotropy via regression with partial differential regularization. *J. Multivariate Anal.*, 167:15–30.

Bernardi, M. S., Sangalli, L. M., Mazza, G., and Ramsay, J. O. (2017). 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*, 31(1):23–38.

Bickel, P. J., Klaassen, C. A. J., Ritov, Y., and Wellner, J. A. (1998). *Efficient and adaptive estimation for semiparametric models*. Springer-Verlag, New York. Reprint of the 1993 original.

Bivand, R. S., Pebesma, E., and Gómez-Rubio, V. (2013). *Applied spatial data analysis with R*. Use R! Springer, New York, second edition.

Buja, A., Hastie, T., and Tibshirani, R. (1989). Linear smoothers and additive models. *Ann. Statist.*, 17(2):453–555.

Castruccio, S. and Stein, M. L. (2013). Global space-time models for climate ensembles. *Ann. Appl. Stat.*, 7(3):1593–1611.

Chung, M., Hanson, J., and Pollak, S. (2017). Statistical analysis on brain surfaces. In *Handbook of Modern Statistical Methods: Neuroimaging Data Analysis*, Chapman & Hall/CRC Handbooks of Modern Statistical Methods, pages 233–262. CRC Press, Boca Raton, FL.

Chung, M. K., Robbins, S. M., Dalton, K. M., Davidson, R. J., Alexander, A. L., and Evans, A. C. (2005). Cortical thickness analysis in autism with heat kernel smoothing. *NeuroImage*, 25:1256–1265.

Ciarlet, P. G. (2002). *The finite element method for elliptic problems*, volume 40 of *Classics in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA. Reprint of the 1978 original [North-Holland, Amsterdam; MR0520174 (58 #25001)].

Cottrell, J. A. abd Hughes, T. J. R. and Bazilevs, Y. (2009). *Isogeometric Analysis. Towards integration of CAD and FEA*. Wiley. John Wiley & Sons, Inc., New York.

Craven, P. and Wahba, G. (1978/79). Smoothing noisy data with spline functions. Estimating the correct degree of smoothing by the method of generalized cross-validation. *Numer. Math.*, 31(4):377–403.

Cressie, N. and Wikle, C. K. (2011). *Statistics for spatio-temporal data*. Wiley Series in Probability and Statistics. John Wiley & Sons, Inc., Hoboken, NJ.

Cressie, N. A. C. (2015). *Statistics for spatial data*. Wiley Classics Library. John Wiley & Sons, Inc., New York, revised edition. Paperback edition of the 1993 edition [ MR1239641].

Dassi, F., Ettinger, B., Perotto, S., and Sangalli, L. M. (2015). A mesh simplification strategy for a spatial regression analysis over the cortical surface of the brain. *Appl. Numer. Math.*, 90:111–131.

Delicado, P., Giraldo, R., Comas, C., and Mateu, J. (2010). Statistics for spatial functional data: some recent contributions. *Environmetrics*, 21(3-4):224–239.

Diggle, P. J. and Ribeiro, Jr., P. J. (2007). *Model-based geostatistics*. Springer Series in Statistics. Springer, New York.

Duchamp, T. and Stuetzle, W. (2003). Spline smoothing on surfaces. *J. Comput. Graph. Statist.*, 12(2):354–381.

Essen, D. V., Ugurbil, K., Auerbach, E., Barch, D., Behrens, T., Bucholz, R., Chang, A., Chen, L., Corbetta, M., Curtiss, S., Penna, S. D., Feinberg, D., Glasser, M., Harel, N., Heath, A., Larson-Prior, L., Marcus, D., Michalareas, G., Moeller, S., Oostenveld, R., Petersen, S., Prior, F., Schlaggar, B., Smith, S., Snyder, A., Xu, J., and Yacoub, E. (2012). The human connectome project: A data acquisition perspective. *NeuroImage*, 62(4):2222 – 2231.

Ettinger, B., Guillas, S., and Lai, M.-J. (2012). Bivariate splines for ozone concentration forecasting. *Environmetrics*, 23(4):317–328.

Ettinger, B., Perotto, S., and Sangalli, L. M. (2016). Spatial regression models over two-dimensional manifolds. *Biometrika*, 103(1):71–88.

Eubank, R. L. (1999). *Nonparametric regression and spline smoothing*, volume 157 of *Statistics: Textbooks and Monographs*. Marcel Dekker, Inc., New York, second edition.

Ferraccioli, F., Arnone, E., Finos, L., Ramsay, J. O., and Sangalli, L. M. (2021). Nonparametric density estimation over complicated domains. *J. R. Stat. Soc. Ser. B. Stat. Methodol.*, page DOI: 10.1111/rssb.12415.

Ferraty, F. and Vieu, P. (2006). *Nonparametric functional data analysis*. Springer Series in Statistics. Springer, New York. Theory and practice.

Formaggia, L., Quarteroni, A., and Veneziani, A. (2009). *Cardiovascular mathematics: modeling and simulation of the circulatory system*. Springer.

Glasser, M. F., Sotiropoulos, S. N., Wilson, J. A., Coalson, T. S., Fischl, B., Andersson, J. L., Xu, J., Jbabdi, S., Webster, M., Polimeni, J. R., Essen, D. C. V., and Jenkinson, M. (2013). The minimal preprocessing pipelines for the human connectome project. *NeuroImage*, 80(0):105 – 124.

Gneiting, T. (2013). Strictly and non-strictly positive definite functions on spheres. *Bernoulli*, 19:1087–1500.

Green, P. J. and Silverman, B. W. (1994). *Nonparametric regression and generalized linear models*, volume 58 of *Monographs on Statistics and Applied Probability*. Chapman & Hall, London. A roughness penalty approach.

Guillas, S. and Lai, M.-J. (2010). Bivariate splines for spatial functional regression models. *Journal of Nonparametric Statistics*, 22(4):477–497.

Hagler, Jr., D. J., Saygin, A. P., and Sereno, M. I. (2006). Smoothing and cluster thresholding for cortical surface-based group analysis of fMRI data. *NeuroImage*, 33:1093–1103.

Härdle, W., Müller, M., Sperlich, S., and Werwatz, A. (2004). *Nonparametric and semiparametric models*. Springer Series in Statistics. Springer-Verlag, New York.

Hastie, T. J. and Tibshirani, R. J. (1990). *Generalized additive models*, volume 43 of *Monographs on Statistics and Applied Probability*. Chapman and Hall, Ltd., London.

Hjelle, Ø. and Dæhlen, M. (2006). *Triangulations and applications*. Springer Science & Business Media.

Huang, J. Z., Shen, H., and Buja, A. (2008). Functional principal components analysis via penalized rank one approximation. *Electron. J. Stat.*, 2:678–695.

Jeong, J. and Jun, M. (2015). A class of Matérn-like covariance functions for smooth processes on a sphere. *Spat. Stat.*, 11:1–18.

Jost, J. (2017). *Riemannian geometry and geometric analysis*. Universitext. Springer, Cham, seventh edition.

Kokoszka, P. and Reimherr, M. (2017). *Introduction to Functional Data Analysis*. Chapman & Hall/CRC Texts in Statistical Science. Chapman & Hall/CRC.

Lai, M.-J. and Schumaker, L. (2007). *Spline functions on triangulations*, volume 110. Cambridge University Press.

Lai, M.-J., Shum, C. K., Baramidze, V., and Wenston, P. (2009). Triangulated spherical splines for geopotential reconstruction. *J. Geodesy*, 83(4):695–708.

Lai, M.-J. and Wang, L. (2013). Bivariate penalized splines for regression. *Statistica Sinica*, 23.

Lila, E., Aston, J. A. D., and Sangalli, L. M. (2016). Smooth Principal Component Analysis over two-dimensional manifolds with an application to neuroimaging. *Ann. Appl. Stat.*, 10(4):1854–1879.

Lila, E., Sangalli, L. M., Arnone, E., Ramsay, J., and Formaggia, L. (2020). *fdaPDE: Statistical Analysis of Functional and Spatial Data, Based on Regression with PDE Regularization*. R package version 1.0-9.

Lindgren, F., Rue, H., and Lindström, J. (2011). An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 73(4):423–498. With discussion and a reply by the authors.

Marra, G., Miller, D. L., and Zanin, L. (2012). Modelling the spatiotemporal distribution of the incidence of resident foreign population. *Stat. Neerl.*, 66(2):133–160.

Martínez-Hernández, I. and Genton, M. G. (2020). Recent developments in complex and spatially correlated functional data. *Braz. J. Probab. Stat.*, 34(2):204–229.

Mateu, J. and Romano, E. (2017). Advances in spatial functional statistics. *Stochastic Environmental Research and Risk Assessment*, 31(1):1–6.

McCullagh, P. and Nelder, J. A. (1989). *Generalized linear models*. Monographs on Statistics and Applied Probability. Chapman & Hall, London. Second edition [of MR0727836].

Menafoglio, A., Gaetani, G., and Secchi, P. (2018). Random domain decompositions for object-oriented kriging over complex domains. *Stochastic Environmental Research and Risk Assessment*, 32(12):3421–3437.

Menafoglio, A., Pigoli, D., and Secchi, P. (2021). Kriging riemannian data via random domain decompositions. *Journal of Computational and Graphical Statistics*, page Doi: 10.1080/10618600.2020.1853548.

Menafoglio, A. and Secchi, P. (2017). Statistical analysis of complex and spatially dependent data: a review of Object Oriented Spatial Statistics. *European J. Oper. Res.*, 258(2):401–410.

Niu, M., Cheung, P., Lin, L., Dai, Z., Lawrence, N., and Dunson, D. (2019). Intrinsic Gaussian processes on complex constrained domains. *J. R. Stat. Soc. Ser. B. Stat. Methodol.*, 81(3):603–627.

Porcu, E., Alegria, A., and Furrer, R. (2018). Modelling temporally evolving and spatially globally dependent data. *Int. Stat. Rev.*, 86(2):344–377.

Porcu, E., Bevilacqua, M., and Genton, M. G. (2016). Spatio-temporal covariance and cross-covariance functions of the great circle distance on a sphere. *J. Amer. Statist. Assoc.*, 111(514):888–898.

Quarteroni, A. (2017). *Numerical models for differential problems*, volume 16 of *MS&A. Modeling, Simulation and Applications*. Springer, Cham. Third edition of [ MR3183828].

R Core Team (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Ramsay, J. O. and Hooker, G. (2017). *Dynamic data analysis*. Springer Series in Statistics. Springer, New York. Modeling data with differential equations.

Ramsay, J. O. and Silverman, B. W. (2005). *Functional data analysis*. Springer Series in Statistics. Springer, New York, second edition.

Ramsay, T. (2002). Spline smoothing over difficult regions. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 64(2):307–319.

Rice, J. A. and Silverman, B. W. (1991). Estimating the mean and covariance structure nonparametrically when the data are curves. *J. Roy. Statist. Soc. Ser. B*, 53(1):233–243.

Riesz, F. and Sz.-Nagy, B. (1990). *Functional analysis*. Dover Books on Advanced Mathematics. Dover Publications, Inc., New York. Translated from the second French edition by Leo F. Boron, Reprint of the 1955 original.

Rudin, W. (1991). *Functional analysis*. International Series in Pure and Applied Mathematics. McGraw-Hill, Inc., New York, second edition.

Sangalli, L. M., Ramsay, J. O., and Ramsay, T. O. (2013). Spatial spline regression models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.*, 75(4):681–703.

Scott-Hayward, L., MacKenzie, M., Donovan, C., Walker, C., and Ashe, E. (2014). Complex region spatial smoother (cress). *Journal of Computational and Graphical Statistics*, 23(2):340–360.

Silverman, B. W. (1996). Smoothed functional principal components analysis by choice of norm. *Ann. Statist.*, 24(1):1–24.

Wahba, G. (1981). Spline interpolation and smoothing on the sphere. *SIAM J. Sci. Statist. Comput.*, 2(1):5–16.

Wang, H. and Ranalli, M. (2007). Low-rank smoothing splines on complicated domains. *Biometrics*, 63(1):209–217.

Wilhelm, M., Dedè, L., Sangalli, L. M., and Wilhelm, P. (2016). IGS: an IsoGeometric approach for smoothing on surfaces. *Comput. Methods Appl. Mech. Engrg.*, 302:70–89.

Wilhelm, M. and Sangalli, L. M. (2016). Generalized spatial regression with differential regularization. *J. Stat. Comput. Simul.*, 86(13):2497–2518.

Wood, S., Bravington, M., and Hedley, S. (2008). Soap film smoothing. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70:931–955.

Xun, X., Cao, J., Mallick, B., Maity, A., and Carroll, R. J. (2013). Parameter estimation of partial differential equation models. *J. Amer. Statist. Assoc.*, 108(503):1009–1020.

Zhou, L. and Pan, H. (2014). Principal component analysis of two-dimensional functional data. *J. Comput. Graph. Statist.*, 23(3):779–801.

# MOX Technical Reports, last issues

**31/2021**    Ferraccioli, F.; Arnone, E.; Finos, L.; Ramsay, J.O.; Sangalli, L.M.
*Nonparametric density estimation over complicated domains*

**29/2021**    Fumagalli, I.; Vitullo, P.; Scrofani, R.; Vergara, C.
*Image-based computational hemodynamics analysis of systolic obstruction in hypertrophic cardiomyopathy*

**30/2021**    Fumagalli, I.
*A reduced 3D-0D FSI model of the aortic valve including leaflets curvature*

**28/2021**    Ferro, N.; Perotto, S.; Bianchi, D.; Ferrante, R.; Mannisi, M.
*Design of cellular materials for multiscale topology optimization: application to patient-specific orthopedic devices*

**27/2021**    Scimone, R.;  Menafoglio,  A.; Sangalli, L.M.; Secchi, P.
*A look at the spatio-temporal mortality patterns in Italy during the COVID-19 pandemic through the lens of mortality densities*

**26/2021**    Vigano, L.; Sollini, M.; Ieva, F.; Fiz, F.; Torzilli, G.
*Chemotherapy-Associated Liver Injuries: Unmet Needs and New Insights for Surgical Oncologists*

**25/2021**    Tenderini, R.; Pagani, S.; Quarteroni, A.; Deparis S.
*PDE-aware deep learning for inverse problems in cardiac electrophysiology*

**24/2021**    Regazzoni, F.; Chapelle, D.; Moireau, P.
*Combining Data Assimilation and Machine Learning to build data-driven models for unknown long time dynamics - Applications in cardiovascular modeling*

**23/2021**    Scimone, R.; Taormina, T.; Colosimo, B. M.; Grasso, M.; Menafoglio,  A.; Secchi, P.
*Statistical modeling and monitoring of geometrical deviations in complex shapes with application to Additive Manufacturing*

**22/2021**    Domanin, M.; Bennati, L.; Vergara, C.; Bissacco, D.; Malloggi, C.; Silani, V.; Parati, G.; Trima
*Fluid structure interaction analysis to stratify the behavior of different atheromatous carotid plaques*