MOX-Report No. 48/2018

Modeling spatially dependent functional data via regression with di fferential regularization

Arnone, E.; Azzimonti, L.; Nobile, F.; Sangalli, L.M.

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

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

Modeling spatially dependent functional data via regression with differential regularization

Eleonora Arnone[1], Laura Azzimonti[2], Fabio Nobile[3], Laura Maria Sangalli[1]

September 13, 2018

[1] MOX – Dipartimento di Matematica, Politecnico di Milano, Piazza Leonardo da Vinci 32, 20133 Milano, Italy laura.sangalli@polimi.it
[2] IDSIA – Department of Innovative Technologies, Università della Svizzera Italiana, Galleria 1, via Cantonale, Switzerland
[3] MATHICSE-CSQI, École polytechnique fédérale de Lausanne, route Cantonale, 1015 Lausanne, Switzerland

Keywords: Finite elements, Partial differential equation, Penalized regression, Smoothing.

Abstract

We propose a method for modeling spatially dependent functional data, based on regression with differential regularization. The regularizing term enables to include problem-specific information about the spatio-temporal variation of the phenomenon under study, formalized in terms of a time-dependent partial differential equation. The method is implemented using a discretization based on finite elements in space and finite differences in time. This non-tensor product basis allows to handle efficiently data distributed over complex domains and where the shape of the domain influences the phenomenon's behavior. Moreover, the method can comply with specific conditions at the boundary of the domain of interest. Simulation studies compare the proposed model to available techniques for spatio-temporal data. The method is also illustrated via an application to the study of blood-flow velocity field in a carotid artery affected by atherosclerosis, starting from echo-color doppler and magnetic resonance imaging data.

1 Introduction

We consider the problem of modeling functional data with complex dependencies, such as spatially dependent curve data or time dependent surface

data. We are interested in particular in situations where problem specific information about the phenomenon under study is available, and this information can be formalized in terms of a time-dependent partial differential equation (PDE) that jointly models the spatio-temporal variation of the phenomenon.

Our contribution is at the crossroad between spatial data analysis [13, 14, 16] and functional data analysis; see, e.g., [39, 18, 23, 25]. This area has recently attracted a lot of interest. Many approaches and methods typical of functional data analysis have been generalized to handle spatial and spatio-temporal data; reciprocally, some techniques from spatial data analysis have been extended to handle spatially-dependent functional data. See, e.g., the reviews in [30] and [33]. Many authors have considered generalizations of kriging to functional data: ordinary kriging for functional data is for instance considered in [15, 34, 20], universal kriging in [9, 31, 32], kriging with external drift in [24], and cokriging in [21]. Other authors have proposed smoothing methods [3, 29, 7, 1] using roughness penalties that account separately for the regularity of the field in space and in time, in tensor product approaches.

Here we extend spatial regression with differential regularization [40, 41, 5] to spatially dependent functional data. Analogously to [5], the regularizing term involves a PDE that models the phenomenon under study. In contrast to [5] though, which only handles data in space, we deal here with spatio-temporal data, and consider a time-dependent PDE that jointly models the spatio-temporal dependence in the data, on the base of problem-specific information about the phenomenon under study. The use of a unique regularizing term that jointly regularizes the field in space and time also distinguishes the proposed model from the methods based on two regularizing terms described in [3, 29, 7, 1].

Specifically, in our proposal, the regularization involves the misfit of a time-dependent PDE $\partial f/\partial t + Lf = u$, where $\partial f/\partial t$ is the time derivative of the spatio-temporal function f, and L is a differential operator in space. The problem specific information is also formalized in terms of conditions that the estimated field must satisfy at the boundaries of the domain of interest, with a very flexible modeling of the behavior at the boundaries of the spatio-temporal field. We consider various sampling designs, including geo-statistical and areal/interval data. We prove that the corresponding estimation problems are well posed, i.e., that the estimators exist and are uniquely defined; this analysis has never been carried out for the techniques described in [3, 29, 7, 1]. Moreover we show that the estimation problems can be discretized in space by means of the finite element method, similarly to [40, 41, 7], and in time by means of the finite difference method. The finite element basis used in space allows to handle efficiently data distributed over irregularly shaped domains. This is crucial when the shape of the domain influences the phenomenon under study, as in the applied problem that has stimulated this research.

1.1 Motivating applied problem

This research was motivated by the study of blood flow velocity field in a section of a carotid artery, starting from Echo-Color Doppler (ECD) data. This problem arose within the research project MACAREN@MOX: MAthematics for CARotid ENdarterectomy @ MOX. The project involves medical doctors in cardiac surgery (from Ca' Granda Ospedale Maggiore Policlinico, in Milano, Italy), statisticians (from MOX Laboratory for Modeling and Scientific Computing, Department of Mathematics, Politecnico di Milano, Italy), and numerical analysts (from MOX and from the Mathematics Institute of Computational Science and Engineering, École polytechnique fédérale de Lausanne, Switzerland), with the intent of investigating the pathogenesis of atherosclerosis in human carotids. In particular, the project aims at exploring the role of blood fluid dynamics and vessel morphology on the possible onset and development of atherosclerotic plaques.

The study is based on ECD measurements of the blood flow velocity at a cross-section of the common carotid artery, 2 cm prior to the carotid bifurcation, for patients affected by high-grade stenosis ($> 70\%$) at the carotid bifurcation. ECD employs ultrasound waves to measure the velocity of blood particles, in a given acquisition beam. Thanks to its low cost, short acquisition time, and non-invasivity (ECD does not require contrast media), ECD is the first and most commonly used test to diagnose carotid artery diseases, such as ischemic stroke, caused by the presence of atherosclerotic plaques, and to investigate various other cardiovascular pathologies. In the MACAREN@MOX project, the ECD data are coupled with the reconstructions of the considered carotid cross-sections, obtained via segmentation of

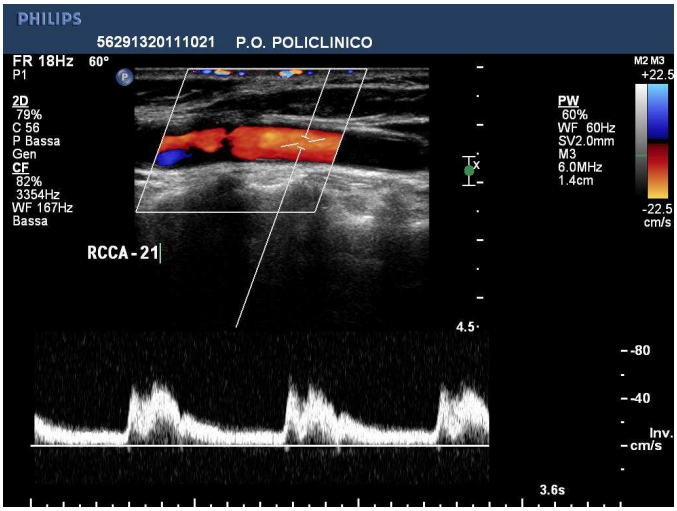

Figure 1: ECD scan in the central acquisition beam at the carotid cross-section located 2 cm prior to the carotid bifurcation.

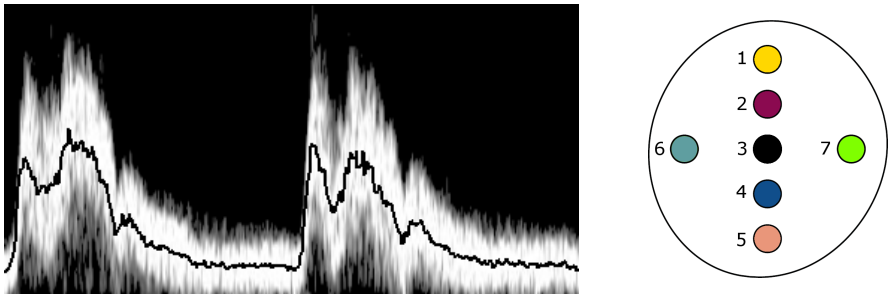

Figure 2: Left: ECD signal in the central beam of the carotid section (beam 3), with superimposed mean velocity. Right: Cross-section of the carotid artery, 2 cm prior to the carotid bifurcation, as reconstructed from MRI data, with indication of the location of the seven beams where the ECD signal is acquired; different colors and numbers are used to indicate the different beams; the same colors and numbers are used in Figures 3 and 12.

magnetic resonance imaging (MRI) data.

Figure 1 displays one of the ECD images. The right panel of Figure 2 shows the section of the carotid, reconstructed from MRI data, and displays the seven beams over which the ECD signal is acquired. The scan in Figure 1 corresponds to the central beam (beam 3). The left panel of Figure 2 zooms-in the acquired signal over the time lapse of about two heart-beats. The ECD scan provides the time-evolving histogram of the velocities (in the longitudinal direction of the vessel) of the blood particles sampled within the

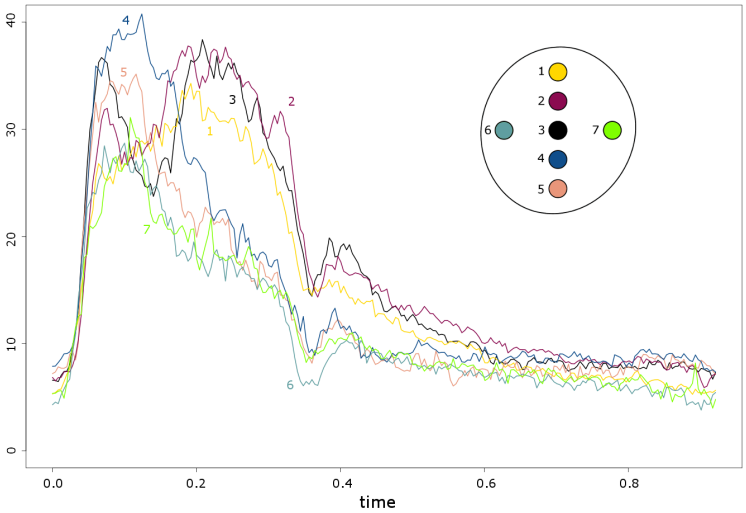

Figure 3: Mean velocity measured in the seven beams on the artery cross-section. Different colors and numbers are used to denote the signals over the seven beams, following the color and number scheme displayed in the upper right corner of the figure.

acquisition beam. Specifically, the y-axis of this plot represents the velocity of the sampled blood particles, with the gray-scaled intensity of pixels corresponding to the number of sampled blood particles having velocity within a certain velocity class. The solid black line superimposed to this signal, in the left panel of Figure 2, corresponds to the mean velocity. Figure 3 reports the mean velocities over the seven acquisition beams, for one heart-beat; different colors and numbers are used to indicate the recorded mean velocities over the different beams, following the color and number scheme shown in the upper right corner of the same figure.

The central goal of the project consists in accurately estimating the time-dependent blood-flow velocity field over the carotid cross-section, starting from these seven spatially dependent functional data, for each of the patients involved in the study. The described data setting though presents some peculiarities, hindering the applicability of both classical and recently proposed techniques for spatio-temporal data, as well as of the available methods for spatially dependent functional data. First of all, the shape of the domain, the carotid cross section, influences the spatio-temporal blood flow velocity field, and hence must be explicitly considered during the estimation process. Unfortunately, almost all the available techniques naturally work over rectangular or tensorized domain. Moreover, in this applied problem there are specific conditions that the estimates must satisfy at the boundary of the spatial domain: the estimated blood flow velocity must in fact be zero at the arterial wall, the boundary of the spatial domain, due to the friction between the wall and the blood particles (the so-called no-slip conditions). Finally, due to the cross shaped pattern of the observations, highlighted in the right panel of Figure 2, isotropic and spatially stationary methods return non-physiological estimates, as already illustrated in [5]; the latter work focused on estimating the blood-flow field at a specific time instant, the systolic peak.

In contrast, we can here profit of a detailed problem-specific information, that can be formalized in terms of a time-dependent PDE, modeling the spatio-temporal behavior of the phenomenon under study. Using the proposed approach, this problem-specific information, thus formalized, can be profitably included in the estimation process, to define an anisotropic and spatially non-stationary estimator that yields physiological estimates. Section 8 illustrates how this and the other issues mentioned above can efficiently be tackled by the proposed approach.

1.2 Structure of the work

The paper is organized as follows. Section 2 introduces the modeling of the spatio-temporal variation of the phenomenon under study via a time dependent PDE. Section 3 describes the proposed spatio-temporal regression with time-dependent PDE regularization, under the simplifying assumption that the functional data are available continuously over time; this simpli-

fied sampling design is the one considered by kriging for functional data. The well-posedness of the estimation problem is proven. Section 4 describes the discretization of the estimation problem by means of the finite element method in space and the finite difference method in time. Section 6 extends the model to various realistic sampling design, including geostatistical and areal/interval data; also in these cases, the well-posedness of the estimation problem is shown. Section 7 reports some simulation studies that compare the proposed method to spatio-temporal kriging and to smoothing methods based on two regularizing terms. Section 8 shows the application to blood velocity field estimation. Finally, Section 9 outlines future research directions. All technical details and proofs are deferred to the Appendix.

2 Modeling problem-specific information on a spatio-temporal field via a time dependent PDE

Let $\Omega \subset \mathbb{R}^2$ be a bounded spatial domain with boundary $\partial\Omega \in C^2$, and let $[t_{\text{start}}, t_{\text{end}}] \subset \mathbb{R}$ be a temporal domain. To lighten the notation, without loss of generality, we set $t_{\text{start}} = 0$ and $t_{\text{end}} = T$. We want to estimate a spatio-temporal field $f_0(\mathbf{p}, t) : \Omega \times [0, T] \to \mathbb{R}$ in the presence of problem-specific information on f_0. Specifically, we assume to have a prior knowledge on the phenomenon under study that can be described in terms of a time-dependent (so-called parabolic) PDE, viz.

$$\partial f_0/\partial t + Lf_0 = u \tag{1}$$

where $\partial f_0/\partial t$ is the time derivative of f_0 and L is a linear second order differential operator defined as

$$Lf = -\operatorname{div}(\mathbf{K}\nabla f) + \mathbf{b} \cdot \nabla f + cf \tag{2}$$

with $\mathbf{K} \in \mathbb{R}^{2\times 2}$ the symmetric and positive definite diffusion tensor, $\mathbf{b} \in \mathbb{R}^2$ the transport vector, and $c \geq 0$ the reaction term. The diffusion, transport and reaction terms may vary over space, i.e., $\mathbf{K} = \mathbf{K}(\mathbf{p})$, $\mathbf{b} = \mathbf{b}(\mathbf{p})$ and $c = c(\mathbf{p})$. The modeling of the space-time variation of the field by the PDE specified in (1) and (2) is very flexible. For example, with the diffusion term $\mathbf{K}(\mathbf{p})$ we can model non-stationary and anisotropic diffusion effects; with the transport term $\mathbf{b}(\mathbf{p})$ we can model non-stationary unidirectional effects; with the reaction term $c(\mathbf{p})$ we can model non-stationary shrinking effects.

In addition, the problem specific information is specified in terms of boundary conditions that f_0 satisfies on the boundary $\partial\Omega$ of the spatial domain Ω, and an initial condition $s(\mathbf{p})$ at time $t = 0$. Various boundary conditions may be considered, such as homogeneous or non-homogeneous Dirichlet, Neumann, Robin (or mixed) conditions, thus enabling a very flexible modeling of the behavior of the spatio-temporal field at the boundaries of

the domain of interest. In particular, Dirichlet boundary conditions concern the value of the field f_0 on $\partial\Omega$; Neumann boundary conditions involve the evaluation of the normal derivative of f_0 on $\partial\Omega$, thus controlling the flow across the boundary; Robin boundary conditions specify a linear combination of the values of the field and the values of its normal derivative on $\partial\Omega$; finally, mixed conditions allow to divide the boundary of Ω, and impose different types of boundary conditions in different portions of $\partial\Omega$. The possible boundary conditions can be summarized as $\mathcal{B}_c f_0 = h$ on $\partial\Omega \times (0,T]$, with

$$\mathcal{B}_c f_0 = \begin{cases} f_0 & \text{on } \Gamma_D \times (0,T], \\ \mathbf{K}\nabla f_0 \cdot \boldsymbol{\nu} & \text{on } \Gamma_N \times (0,T], \\ \mathbf{K}\nabla f_0 \cdot \boldsymbol{\nu} + \chi f_0 & \text{on } \Gamma_R \times (0,T], \end{cases} \qquad h = \begin{cases} h_D & \text{on } \Gamma_D \times (0,T], \\ h_N & \text{on } \Gamma_N \times (0,T], \\ h_R & \text{on } \Gamma_R \times (0,T], \end{cases}$$

where $\boldsymbol{\nu}$ is the outward unit normal vector to $\partial\Omega$, $\chi \in \mathbb{R}$ is a positive constant, and $\partial\Omega = \bar{\Gamma}_D \cup \bar{\Gamma}_N \cup \bar{\Gamma}_R$, with $\Gamma_D, \Gamma_N, \Gamma_R$ not overlapping. The boundary conditions are said homogeneous if $h = 0$.

The parabolic PDE is thus specified by

$$\begin{cases} \partial f_0/\partial t + L f_0 = u & \text{in } \Omega \times (0,T), \\ f_0(\mathbf{p}, 0) = s(\mathbf{p}) & \text{in } \Omega, \\ \mathcal{B}_c f_0 = h & \text{on } \partial\Omega \times (0,T]. \end{cases} \tag{3}$$

3 Model for continuous data with spatial dependence

For simplicity of exposition, we start by considering the setting where we have observations $z_i(t)$ at the spatial locations $\mathbf{p}_i = (x_i, y_i) \in \Omega$, for $i \in \{1, \ldots, n\}$, and these data are available continuously over time, for $t \in [0,T]$. This is the setting considered by kriging for functional data; see, e.g., [15, 33]. Realistic sampling designs, where the data are not available continuously over time, will be considered in Section 6. We assume that the data $z_i(t)$ are noisy observations of the deterministic spatio-temporal field $f_0 : \Omega \times [0,T] \to \mathbb{R}$. Specifically, we suppose that for all $i \in \{1, \ldots, n\}$,

$$z_i(t) = f_0(\mathbf{p}_i, t) + \varepsilon_i(t)$$

where $\mathrm{E}\{\varepsilon_i(t)\} = 0$, $\mathrm{var}\{\varepsilon_i(t)\} = \sigma^2$, $\mathrm{cov}\{\varepsilon_i(t), \varepsilon_j(t^*)\} = 0$ for $i \neq j$, and $\mathrm{cov}\{\varepsilon_i(t), \varepsilon_i(t^*)\} = 0$ for $|t - t^*| > \delta$ for some $\delta > 0$.

We want to take advantage of the problem-specific knowledge on the phenomenon under study, specified in terms of a time-dependent PDE, as described in Section 2. To this end, we propose to estimate f_0 by minimizing the following penalized sum-of-square-error functional

$$J_T(f) = \sum_{i=1}^{n} \int_0^T \{f(\mathbf{p}_i, t) - z_i(t)\}^2 + \lambda \int_0^T \int_\Omega (\partial f/\partial t + Lf - u)^2, \tag{4}$$

where $\lambda > 0$ is a smoothing parameter. The functional J_T trades off a data fitting criterion, consisting in the sum of $L^2[0,T]$ errors, and a model fidelity criterion, formalized as a regularizing term and involving the misfit from the PDE that models the phenomenon under study. For very small values of the smoothing parameter, the estimated function is rough, while for large values the estimated function becomes smoother, approaching the solution of the regularizing PDE.

For simplicity, the data fitting criterion considered in the functional J_T in (4) does not account for the short range temporal correlation of the error. One could also consider a data fitting criterion that instead accounts for it. However, as it will be clear from Section 4, if the temporal step used in the discretization of the estimator is larger than δ, this short range correlation has no effect on the estimator and the two formulations would indeed coincide, once the problem is discretized.

3.1 Functional space where the estimation functional is well-defined

Denote by $H^k(\Omega)$ the Sobolev space of functions $\ell : \Omega \to \mathbb{R}$ that are in $L^2(\Omega)$ and whose derivatives up to the order k are in $L^2(\Omega)$, equipped with the norm

$$\|\ell\|_{H^k(\Omega)} = \left(\sum_{|\alpha| \leq k} \|D^\alpha \ell\|_{L^2}^2 \right)^{1/2},$$

where $\alpha = (\alpha_1, \alpha_2)$ and $D^\alpha \ell = \partial^{|\alpha|} \ell / \partial x^{\alpha_1} \partial y^{\alpha_2}$ denotes the derivative of order $|\alpha| = \alpha_1 + \alpha_2$, with $D^0 \ell = \ell$. Denote by $L^2(0,T;H^k(\Omega))$ the space of functions defined over $[0,T]$, taking values in $H^k(\Omega)$, and such that

$$\int_0^T \|f(t)\|_{H^k(\Omega)}^2 \, dt < \infty.$$

Define the space

$$V = \{ f \in L^2(0,T;H^2(\Omega)) : \partial f / \partial t \in L^2(0,T;L^2(\Omega)) \\ + \text{boundary and initial conditions} \}.$$

This space contains the functions such that

$$\|f\|_V^2 = \int_0^T \|f(t)\|_{H^2(\Omega)}^2 \, dt + \int_0^T \|\partial f(t)/\partial t\|_{L^2(\Omega)}^2 \, dt < \infty.$$

Both the error term and the regularization term in $J_T(f)$ are well defined in V, since this space contains functions continuous in space and square integrable in time, such that

$$\int_0^T \left\{ \sup_{\mathbf{p} \in \bar{\Omega}} f(\mathbf{p}, t) \right\}^2 dt < \infty;$$

this follows from the embedding $H^2(\Omega) \subset C(\bar{\Omega})$, if $\Omega \subset \mathbb{R}^d$ with $d \in \{1, 2, 3\}$. Hence, the functional $J_T(f)$ is well defined for $f \in V$.

3.2 Estimation problem

The estimation problem is formulated as follows.

Problem 1. *Find $\hat{f} \in V$ such that*

$$\hat{f} = \operatorname*{argmin}_{f \in V} J_T(f).$$

Denote by $\delta_{\mathbf{p}_i}$ the Dirac mass located in $\mathbf{p}_i$. The existence and uniqueness of the estimator is stated in the following proposition.

Proposition 1. *Under suitable regularity assumptions (see A), the solution of Problem 1, with J_T given in (4), exists and is unique. It is obtained by solving a coupled system of two time-dependent PDEs, viz.*

$$\begin{cases} \partial \hat{f}/\partial t + L\hat{f} = u + \hat{g} & \text{in } \Omega \times (0, T], \\ \hat{f}(\mathbf{p}, 0) = s(\mathbf{p}) & \text{in } \Omega, \\ \mathcal{B}_c \hat{f} = h & \text{on } \partial\Omega \times (0, T], \end{cases} \tag{5}$$

$$\begin{cases} -\partial \hat{g}/\partial t + L^* \hat{g} = -\sum_{i=1}^{n} (\hat{f} - z_i)\delta_{\mathbf{p}_i}/\lambda & \text{in } \Omega \times [0, T), \\ \hat{g}(\mathbf{p}, T) = 0 & \text{in } \Omega, \\ \mathcal{B}_c{}^* \hat{g} = 0 & \text{on } \partial\Omega \times [0, T), \end{cases} \tag{6}$$

where $\hat{g} \in L^2(0, T; L^2(\Omega))$ is the misfit of the penalized PDE, i.e., $\hat{g} = \partial \hat{f}/\partial t + L\hat{f} - u$, L^* is the adjoint operator of L, i.e.,

$$L^* \hat{g} = -\operatorname{div}(\mathbf{K} \nabla \hat{g}) - \mathbf{b} \cdot \nabla \hat{g} + \{c - \operatorname{div}(\mathbf{b})\}\hat{g},$$

and $\mathcal{B}_c{}^*$ is the boundary condition operator of the adjoint problem, i.e.,

$$\mathcal{B}_c{}^* g = \begin{cases} g & \text{on } \Gamma_D, \\ \mathbf{K}\nabla g \cdot \boldsymbol{\nu} + \mathbf{b} \cdot \boldsymbol{\nu} g & \text{on } \Gamma_N, \\ \mathbf{K}\nabla g \cdot \boldsymbol{\nu} + (\mathbf{b} \cdot \boldsymbol{\nu} + \chi) g & \text{on } \Gamma_R. \end{cases}$$

The proof of Proposition 1 is detailed in A, where the regularity assumptions are specified. Eq. (5) is a so-called forward parabolic PDE, while Eq. (6) is a so-called backward parabolic PDE characterized by the opposite sign for the time derivative, and an ending condition instead of the starting one. The second equation is homogeneous both in the boundary conditions and in the ending condition.

To save space, in the following we consider only homogeneous Dirichlet boundary conditions that are those relevant for the applied problem motivating this research. All other boundary conditions may be handled similarly to what described in [4] for SR-PDE estimators over space only.

3.3 Weak formulation of the estimation problem

Let $a(\cdot,\cdot)$ be the bilinear form associated to the operator L, defined as

$$a(\hat{f},\psi) = \int_\Omega \big(\mathbf{K}\nabla\hat{f}\cdot\nabla\psi + \mathbf{b}\cdot\nabla\hat{f}\psi + c\hat{f}\psi\big). \tag{7}$$

The coupled system of PDEs (5)–(6), with homogeneous Dirichlet boundary conditions, has the following equivalent formulation

$$\begin{cases} \displaystyle\int_\Omega \frac{\partial\hat{f}}{\partial t}\,\psi + a(\hat{f},\psi) - \int_\Omega \hat{g}\psi = \int_\Omega u\psi & \text{if } t\in(0,T), \\ \hat{f}(\mathbf{p},0) = s(\mathbf{p}) & \\ \displaystyle -\lambda\int_\Omega \frac{\partial\hat{g}}{\partial t}\,\varphi + \lambda a(\varphi,\hat{g}) + \sum_{i=1}^{n}\hat{f}(\mathbf{p}_i,\cdot)\varphi(\mathbf{p}_i) = \sum_{i=1}^{n} z_i\varphi(\mathbf{p}_i) & \text{if } t\in(0,T), \\ \hat{g}(\mathbf{p},T) = 0, & \end{cases} \tag{8}$$

for all ψ, φ in V, where V is specified with homogeneous Dirichlet boundary conditions. This so-called weak formulation is particularly convenient for the discretization of the estimation problem, as detailed below.

4 Discrete estimator

Problem 1 cannot be solved analytically. For this reason we approximate the PDE system (5)–(6) with the Finite Difference method in time and the finite element method in space. These two methods are classical techniques used in numerical analysis to approximate the solution of parabolic PDEs; see, e.g., [36]. In particular, the approximation we consider is based on the equivalent weak formulation of the estimation problem given in (8).

4.1 Discretization in space via finite elements

For simplicity, we assume here that the spatial domain Ω is polygonal. If the original spatial domain is not polygonal, we simply need to approximate it by a polygonal domain Ω that closely approximate the original spatial domain. Let $\mathcal{T}_h$ be a triangulation of Ω, where h represents the characteristic mesh size, i.e., the maximum length of the triangle edges in the triangulation. The left panel of Figure 4 shows the triangulation used for the application to the blood flow velocity field estimation, introduced in Section 1 and detailed in Section 8. We consider the space V_h^r of globally continuous and piecewise polynomial functions over the triangulation, which are polynomials of order r for $r \geq 1$, once restricted to any triangle in the triangulation, i.e.,

$$V_h^r = \{v \in C^0(\bar{\Omega}) : \forall_{K\in\mathcal{T}_h}\ v|_K \in \mathbb{P}^r(K)\}.$$

Let $N_h = \dim(V_h^r)$. To define a set of N_h basis $\psi_1, \ldots, \psi_{N_h}$ that span such space, it is convenient to consider the so-called nodes of the triangulation, denoted by $\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_{N_h}$. For linear finite elements, the nodes coincide with the vertices of the triangles in $\mathcal{T}_h$. For higher-order finite elements, the nodes are a super-set of the triangle vertices; for instance, for quadratic finite elements the nodes coincide with the triangle vertices and the middle points of the triangle edges. For each $j \in \{1, \ldots, N_h\}$, the basis ψ_j is then associated to one node $\boldsymbol{\xi}_j$, and it is a locally supported piecewise polynomial function of order r that takes value 1 at the associated node and value 0 on all other nodes, i.e., $\psi_j(\boldsymbol{\xi}_k) = \delta_{jk}$, where $\delta_{jk} = 1$ if $j = k$ and $\delta_{jk} = 0$ if $j \neq k$. The right panel of Figure 4 shows an example of linear finite element basis.

When considering homogeneous Dirichlet boundary conditions, the value of the function at the boundary of Ω is fixed to 0. In this case, we can use the finite element space $V_{h,0}^r$ of dimension $N_{h,0}$, defined as

$$V_{h,0}^r = \{v \in C^0(\bar{\Omega}) : v|_{\partial\Omega} = 0 \text{ and } \forall_{K \in \mathcal{T}_h} v|_K \in \mathbb{P}^r(K)\},$$

which only necessitates the internal nodes of the triangulation and the associated basis functions, while all bases associated to boundary nodes can be discarded.

We set $\boldsymbol{\psi} = (\psi_1, \ldots, \psi_{N_{h,0}})^\top$, $\boldsymbol{\psi}_x = (\partial\psi_1/\partial x, \ldots, \partial\psi_{N_{h,0}}/\partial x)^\top$ and $\boldsymbol{\psi}_y = (\partial\psi_1/\partial y, \ldots, \partial\psi_{N_{h,0}}/\partial y)^\top$, and we define the $n \times N_{h,0}$ matrix $\boldsymbol{\Psi}$ of evaluations of the $N_{h,0}$ finite elements basis at the n data locations, i.e.,

$$\boldsymbol{\Psi} = \begin{bmatrix} \boldsymbol{\psi}^\top(\mathbf{p}_1) \\ \vdots \\ \boldsymbol{\psi}^\top(\mathbf{p}_n) \end{bmatrix}.$$

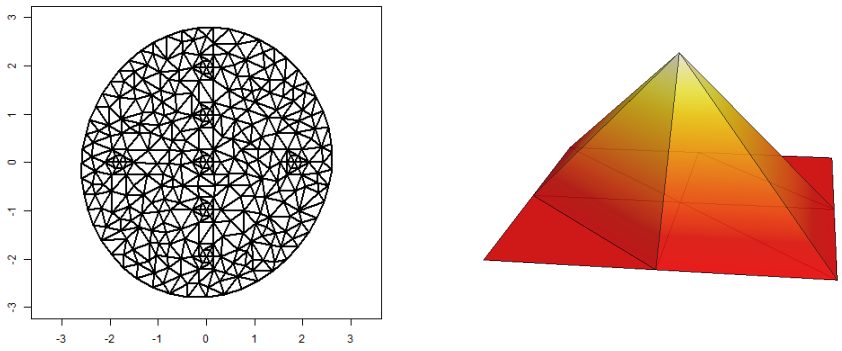

Figure 4: Left: Triangulation of the carotid cross-section displayed in the left panel of Figure 2. Right: A linear finite element basis function on a triangulation.

4.2 Discretization in time via finite differences

In order to discretize the problem in time we use the finite difference method. For the sake of simplicity, we consider N_T uniformly spaced temporal instants in $[0,T]$, named $\tau^0,\dots,\tau^{N_T}$, such that $\tau^0=0$, $\tau^k=k\cdot\Delta t$ and $\tau^{N_T}=T$. We hence discretize the time derivatives as

$$\frac{\hat{f}_h(\,\cdot\,,\tau^k)-\hat{f}_h(\,\cdot\,,\tau^{k-1})}{\Delta t}\approx\frac{\partial\hat{f}_h}{\partial t}(\,\cdot\,,\tau^k).$$

The discretization in time of the system is then obtained by means of an Implicit Euler scheme [37], as detailed in Section 4.3.

4.3 Details of the discretization

Our discretization of the estimation problem is based on the weak formulation (8) of the system of PDEs (5)–(6). The discretization of (8) in space is obtained replacing the infinite-dimensional space V (with the specified homogeneous Dirichlet boundary conditions) by the finite-dimensional space $V^r_{h,0}$:

$$\begin{cases}\displaystyle\int_\Omega\frac{\partial\hat{f}_h}{\partial t}\,\psi_h+a(\hat{f}_h,\psi_h)-\int_\Omega\hat{g}_h\psi_h=\int_\Omega u\psi_h & \text{if } t\in(0,T),\\ \hat{f}_h(\mathbf{p},0)=s_h(\mathbf{p}),\\ \displaystyle-\lambda\int_\Omega\frac{\partial\hat{g}_h}{\partial t}\,\varphi_h+\lambda a(\varphi_h,\hat{g}_h)+\sum_{i=1}^n\hat{f}_h(\mathbf{p}_i,\cdot)\varphi_h(\mathbf{p}_i) & \\ \qquad\qquad=\displaystyle\sum_{i=1}^n z_i\varphi_h(\mathbf{p}_i) & \text{if } t\in(0,T),\\ \hat{g}_h(\mathbf{p},T)=0,\end{cases}\tag{9}$$

for all $\psi_h,\varphi_h\in V^r_{h,0}$, where $\hat{f}_h(\cdot,t),\hat{g}_h(\cdot,t)\in V^r_{h,0}$, for all $t\in[0,T]$, and s_h is the interpolation of the initial condition $s(\mathbf{p})$ in $V^r_{h,0}$. Notice that (9) is only a space discretization. The application of the Implicit Euler scheme then leads to the following discretization in space and time of the variational system (8):

$$\begin{cases}\displaystyle\int_\Omega\frac{\hat{f}_h^k-\hat{f}_h^{k-1}}{dt}\,\psi_h+a(\hat{f}_h^k,\psi_h)-\int_\Omega\hat{g}_h^{k-1}\psi_h=\int_\Omega u^k\psi_h & \text{for } k\in\{1,\dots,N_T\},\\ \hat{f}_h^0=s_h,\\ \displaystyle-\lambda\int_\Omega\frac{\hat{g}_h^k-\hat{g}_h^{k-1}}{dt}\,\varphi_h+\lambda a(\varphi_h,\hat{g}_h^{k-1}) & \\ \qquad\displaystyle+\sum_{i=1}^n\hat{f}_h^k(\mathbf{p}_i)\varphi_h(\mathbf{p}_i)=\sum_{i=1}^n z_i^k\varphi_h(\mathbf{p}_i) & \text{for } k\in\{1,\dots,N_T\},\\ \hat{g}_h^{N_T}=0.\end{cases}\tag{10}$$

To write this system in matrix formulation, define the following matrices in $\mathbb{R}^{N_{h,0}\times N_{h,0}}$:

$$\mathbf{R}(c) = \int_\Omega c\boldsymbol{\psi}\boldsymbol{\psi}^\top, \quad \mathbf{R} = \mathbf{R}(1) = \int_\Omega \boldsymbol{\psi}\boldsymbol{\psi}^\top,$$
$$\mathbf{R}_x(\mathbf{b}) = \int_\Omega \mathbf{b}_1\boldsymbol{\psi}\boldsymbol{\psi}_x^\top, \quad \mathbf{R}_y(\mathbf{b}) = \int_\Omega \mathbf{b}_2\boldsymbol{\psi}\boldsymbol{\psi}_y^\top,$$
$$\mathbf{R}_{xx}(\mathbf{K}) = \int_\Omega \mathbf{K}_{11}\boldsymbol{\psi}_x\boldsymbol{\psi}_x^\top, \quad \mathbf{R}_{yy}(\mathbf{K}) = \int_\Omega \mathbf{K}_{22}\boldsymbol{\psi}_y\boldsymbol{\psi}_y^\top,$$
$$\mathbf{R}_{xy}(\mathbf{K}) = \int_\Omega \mathbf{K}_{12}(\boldsymbol{\psi}_x\boldsymbol{\psi}_y^\top + \boldsymbol{\psi}_y\boldsymbol{\psi}_x^\top),$$

where $\mathbf{K}_{ij}$ and $\mathbf{b}_j$ are the elements of the diffusion tensor matrix $\mathbf{K}$ and of the transport vector $\mathbf{b}$. We can thus write the discretization in the finite element space of the bilinear form $a(\cdot,\cdot)$ defined in (7) as

$$\mathbf{A}(\mathbf{K},\mathbf{b},c) = \mathbf{R}_{xx}(\mathbf{K}) + \mathbf{R}_{xy}(\mathbf{K}) + \mathbf{R}_{yy}(\mathbf{K}) + \mathbf{R}_x(\mathbf{b}) + \mathbf{R}_y(\mathbf{b}) + \mathbf{R}(c).$$

Denote by $\hat{f}_h^k$ and $\hat{g}_h^k$ the finite element approximations of $\hat{f}_h(\cdot,\tau^k)$ and $\hat{g}_h(\cdot,\tau^k)$ respectively, and set $u^k(\cdot) = u(\cdot,\tau^k)$ and $z_i^k = z_i(\tau^k)$ with $k \in \{0,\dots,N_T\}$. Let $\hat{\mathbf{f}}^k$, $\hat{\mathbf{g}}^k$, $\mathbf{s}$ be the vectors such that $\hat{f}_h^k = \boldsymbol{\psi}^\top\hat{\mathbf{f}}^k$, $\hat{g}_h^k = \boldsymbol{\psi}^\top\hat{\mathbf{g}}^k$ and $s_h = \boldsymbol{\psi}^\top\mathbf{s}$. Moreover, for $k \in \{1,\dots,N_T\}$, define the vectors containing the values of z_i at time τ^k, $\mathbf{z}^k = (z_1^k,\dots,z_n^k)^\top$, and set $\mathbf{u}_j^k = \int_\Omega u^k\psi_j$. We can write the system (10) in matrix form as

$$\begin{cases} \left(\mathbf{A} + \dfrac{1}{dt}\mathbf{R}\right)\hat{\mathbf{f}}^k - \dfrac{1}{dt}\mathbf{R}\hat{\mathbf{f}}^{k-1} - \mathbf{R}\hat{\mathbf{g}}^{k-1} = \mathbf{u}^k & \text{for } k \in \{1,\dots,N_T\}, \\ \hat{\mathbf{f}}^0 = \mathbf{s} \\ -\dfrac{\lambda}{dt}\mathbf{R}\hat{\mathbf{g}}^k + \lambda\left(\mathbf{A}^\top + \dfrac{1}{dt}\mathbf{R}\right)\hat{\mathbf{g}}^{k-1} + \boldsymbol{\Psi}^\top\boldsymbol{\Psi}\hat{\mathbf{f}}^k = \boldsymbol{\Psi}^\top\mathbf{z}^k & \text{for } k \in \{1,\dots,N_T\}, \\ \hat{\mathbf{g}}^{N_T} = 0. \end{cases}$$

The discrete surface estimators $\hat{f}_h^k$ are thus obtained solving the system above. Let $\hat{\mathbf{f}}^\top = ((\hat{\mathbf{f}}^1)^\top,\dots,(\hat{\mathbf{f}}^{N_T})^\top)$, $\hat{\mathbf{g}}^\top = ((\hat{\mathbf{g}}^0)^\top,\dots,(\hat{\mathbf{g}}^{N_T-1})^\top)$, $\mathbf{z}^\top = ((\mathbf{z}^1)^\top,\dots,(\mathbf{z}^{N_T})^\top)$ and $\tilde{\mathbf{u}}^\top = ((\mathbf{u}^1 + \mathbf{R}\mathbf{s}/\Delta t)^\top,(\mathbf{u}^2)^\top,\dots,(\mathbf{u}^{N_T})^\top)$. Introduce the matrix $D \in \mathbb{R}^{N_T\times N_T}$ associated with the derivation in time, viz.

$$\mathbf{D} = \frac{1}{\Delta t}\begin{bmatrix} 1 & 0 & & \\ -1 & 1 & & \\ & \ddots & \ddots & \\ & & -1 & 1 \end{bmatrix}$$

and define the tensor product matrices

$$\tilde{\boldsymbol{\Psi}} = I_{N_T}\otimes\boldsymbol{\Psi}, \quad \tilde{\mathbf{A}} = I_{N_T}\otimes\mathbf{A} + \mathbf{D}\otimes\mathbf{R}, \quad \tilde{\mathbf{R}} = I_{N_T}\otimes\mathbf{R},$$

where I_d denotes the identity matrix of dimension d. With this notation, we finally obtain the following proposition, that specifies the form of the discrete estimator as solution of a linear system.

Proposition 2. *The discrete solution* $\hat{\mathbf{f}}$ *of Problem 1 is computed solving the system*

$$\begin{bmatrix} \tilde{\mathbf{\Psi}}^\top \tilde{\mathbf{\Psi}} & \lambda \tilde{\mathbf{A}}^\top \\ \tilde{\mathbf{A}} & -\tilde{\mathbf{R}} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{f}} \\ \hat{\mathbf{g}} \end{bmatrix} = \begin{bmatrix} \tilde{\mathbf{\Psi}}^\top \mathbf{z} \\ \tilde{\mathbf{u}} \end{bmatrix}.$$

5 Distributional properties of the estimator

The surface estimator $\hat{f}_h^k$ at time τ^k is a linear function of the observed data values. The fitted values at the time instant τ^k are computed as $\hat{\mathbf{z}}^k = \mathbf{\Psi}\hat{\mathbf{f}}^k$. The vector $\hat{\mathbf{z}}$, containing the fitted values at all the time instants $\tau^0, \ldots, \tau^{N_T}$, can be thus represented as

$$\hat{\mathbf{z}} = \tilde{\mathbf{S}}\mathbf{z} + \tilde{\mathbf{r}},$$

where the smoothing matrix $\tilde{\mathbf{S}} \in \mathbb{R}^{nN_T \times nN_T}$ and the vector $\tilde{\mathbf{r}} \in \mathbb{R}^{nN_T}$ are obtained as

$$\tilde{\mathbf{S}} = \tilde{\mathbf{\Psi}}(\tilde{\mathbf{\Psi}}^\top \tilde{\mathbf{\Psi}} + \lambda \tilde{\mathbf{P}})^{-1} \tilde{\mathbf{\Psi}}^\top, \quad \tilde{\mathbf{r}} = \tilde{\mathbf{\Psi}}(\tilde{\mathbf{\Psi}}^\top \tilde{\mathbf{\Psi}} + \lambda \tilde{\mathbf{P}})^{-1} \lambda \tilde{\mathbf{P}} \tilde{\mathbf{A}}^{-1} \mathbf{u},$$

with $\tilde{\mathbf{P}}$ denoting the penalty matrix

$$\tilde{\mathbf{P}} = \mathbf{P}(\mathbf{K}, \mathbf{b}, c) = \tilde{\mathbf{A}}^\top (\tilde{\mathbf{R}})^{-1} \tilde{\mathbf{A}}.$$

The smoothing matrix $\tilde{\mathbf{S}}$ has the typical form obtained in a penalized regression problem.

Thanks to the linearity of the estimator $\hat{\mathbf{z}}$ in the observations, we can easily derive its mean and variance-covariance structure, and obtain classical inferential tools as point-wise confidence bands and prediction intervals at a fixed point location and time instant. Let $\mathbf{f}_0 = \big(f_0(\mathbf{p}_1, \tau^1), \ldots, f_0(\mathbf{p}_n, \tau^1), \ldots, f_0(\mathbf{p}_1, \tau^{N_T}), \ldots, f_0(\mathbf{p}_n, \tau^{N_T})\big)^\top$ be the column vector of evaluations of the true function f_0 at the n data locations and N_T time instants used for the temporal discretization. Recall that $\mathrm{E}(\mathbf{z}) = \mathbf{f}_0$. We can thus compute the expected value of the estimator $\hat{\mathbf{z}}$ as

$$\mathrm{E}(\hat{\mathbf{z}}) = \tilde{\mathbf{S}}\mathbf{f}_0 + \tilde{\mathbf{r}}.$$

Suppose in addition that the time discretization is such that $\Delta t > \delta$, so that, for any couple of discretization instants τ^k, τ^ℓ, the noise process satisfies $\mathrm{cov}\{\varepsilon_i(\tau^k)\varepsilon_j(\tau^\ell)\} = 0$ if $i \neq j$ or $k \neq \ell$ and $\mathrm{var}\{\varepsilon_i(\tau^k)\} = \sigma^2$. We can thus compute the variance-covariance structure of $\hat{\mathbf{z}}$ as

$$\mathrm{cov}(\hat{\mathbf{z}}) = \sigma^2 \tilde{\mathbf{S}} \tilde{\mathbf{S}}^\top.$$

In order to compute the estimate at a generic space-time location $(\mathbf{p}, t)$, we can define the function $\boldsymbol{\phi}$ such that, for $t \in [\tau^{k-1}, \tau^k]$, $\boldsymbol{\phi}(t) = (t - \tau^k)\boldsymbol{e}_{k-1}/\Delta t + (\tau^{k-1} - t)\boldsymbol{e}_k \Delta t$, where $\boldsymbol{e}_k$ is the kth vector of the canonical basis of $\mathbb{R}^{N_T}$. Let

$\tilde{\boldsymbol{\psi}} = \boldsymbol{\phi}^\top \otimes \boldsymbol{\psi}^\top$. The estimator of the field f_0 at the generic spatio-temporal location $(\mathbf{p}, t) \in \Omega \times [0, T]$ is then given by

$$\hat{f}_h(\mathbf{p}, t) = \tilde{\boldsymbol{\psi}}(\mathbf{p}, t)(\tilde{\boldsymbol{\Psi}}^\top \tilde{\boldsymbol{\Psi}} + \lambda \tilde{\mathbf{P}})^{-1}(\tilde{\boldsymbol{\Psi}}^\top \mathbf{z} + \lambda \tilde{\mathbf{P}} \tilde{\mathbf{A}}^{-1} \mathbf{u}).$$

Its mean and variance are given by

$$\begin{aligned}
&\mathrm{E}\{\hat{f}_h(\mathbf{p}, t)\} = \tilde{\boldsymbol{\psi}}(\mathbf{p}, t)(\tilde{\boldsymbol{\Psi}}^\top \tilde{\boldsymbol{\Psi}} + \lambda \tilde{\mathbf{P}})^{-1}(\tilde{\boldsymbol{\Psi}}^\top \mathbf{f}_0 + \lambda \tilde{\mathbf{P}} \tilde{\mathbf{A}}^{-1} \mathbf{u}), \\
&\mathrm{var}\{\hat{f}_h(\mathbf{p}, t)\} = \sigma^2 \tilde{\boldsymbol{\psi}}(\mathbf{p}, t)(\tilde{\boldsymbol{\Psi}}^\top \tilde{\boldsymbol{\Psi}} + \lambda \tilde{\mathbf{P}})^{-1} \tilde{\boldsymbol{\Psi}}^\top \tilde{\boldsymbol{\Psi}} (\tilde{\boldsymbol{\Psi}}^\top \tilde{\boldsymbol{\Psi}} + \lambda \tilde{\mathbf{P}})^{-1} \tilde{\boldsymbol{\psi}}(\mathbf{p}, t)^\top.
\end{aligned}$$

The covariance of the field estimator at any two spatio-temporal locations $(\mathbf{p}_1, t_1)$ and $(\mathbf{p}_2, t_2) \in \Omega \times [0, T]$ is given by

$$\begin{aligned}
&\mathrm{cov}\{\hat{f}_h(\mathbf{p}_1, t_1), \hat{f}_h(\mathbf{p}_2, t_2)\} = \\
&\qquad \sigma^2 \tilde{\boldsymbol{\psi}}(\mathbf{p}_1, t_1)(\tilde{\boldsymbol{\Psi}}^\top \tilde{\boldsymbol{\Psi}} + \lambda \tilde{\mathbf{P}})^{-1} \tilde{\boldsymbol{\Psi}}^\top \tilde{\boldsymbol{\Psi}} (\tilde{\boldsymbol{\Psi}}^\top \tilde{\boldsymbol{\Psi}} + \lambda \tilde{\mathbf{P}})^{-1} \tilde{\boldsymbol{\psi}}(\mathbf{p}_2, t_2)^\top.
\end{aligned}$$

Both the mean and the covariance structure of the estimator are thus induced by the regularizing term. Since we are dealing with a linear estimator, we can use $\mathrm{tr}(\tilde{\mathbf{S}})$ as a measure of the equivalent degrees of freedom of the estimator; see, e.g., [8] and [22]. We can hence estimate σ^2 as

$$\hat{\sigma}^2 = \frac{1}{nN_T - \mathrm{tr}(\tilde{\mathbf{S}})} (\hat{\mathbf{z}} - \mathbf{z})^\top (\hat{\mathbf{z}} - \mathbf{z}).$$

The smoothing parameter λ may be selected via generalized cross-validation, minimizing the index

$$GCV = \frac{1}{nN_T \{1 - \mathrm{tr}(\tilde{\mathbf{S}})/nN_T\}^2} (\hat{\mathbf{z}} - \mathbf{z})^\top (\hat{\mathbf{z}} - \mathbf{z}). \tag{11}$$

Studying the convergence of the proposed estimator when the number of observations goes to infinity (infill asymptotics) is not easy, mainly due to the complexity of the PDE regularization and the non-trivial domain. The results obtained for univariate and multivariate splines [11, 12], where the regularization has a simpler form and does not depend on the domain of interest, cannot unfortunately be extended to the estimators here considered. Nevertheless, Arnone [2] studies the convergence of the estimators defined over space only, introduced in [4, 5] and here extended to spatio-temporal domains: in the simpler setting involving only space, both the infinite-dimensional estimator and the finite-dimensional one, obtained after discretization, are shown to be consistent and to achieve the optimal rates of convergence for non-parametric estimators [42].

6 Different sampling designs

We now consider more realistic sampling designs, where we do not assume that the data are observed continuously over time.

Sampling design 1. *Point-wise observations in space and interval observations in time*

Let $\mathbf{p}_1, \ldots, \mathbf{p}_n \in \Omega$ be n spatial locations, and let $T_1, \ldots, T_m \subset [0,T]$ be m disjoint temporal intervals. Assume that for all $i \in \{1, \ldots, n\}$ and $j \in \{1, \ldots, m\}$,

$$z_{ij} = \frac{1}{|T_j|} \int_{T_j} f_0(\mathbf{p}_i, t)dt + \varepsilon_{ij},$$

where the errors ε_{ij} are independent, with zero mean, and variance proportional to $1/|T_j|$. In this case, we estimate f_0 by minimizing with respect to $f \in V$ the following functional:

$$J_T(f) = \sum_{i=1}^{n} \sum_{j=1}^{m} \frac{1}{|T_j|} \left[\int_{T_j} \{f(\mathbf{p}_i, t) - z_{ij}\}dt \right]^2 + \lambda \int_0^T \int_\Omega (\partial f/\partial t + Lf - u)^2.$$

Sampling design 2. *Areal observations in space and point-wise observations in time*

Let $D_1, \ldots, D_n \subset \Omega$ be n disjoint spatial subdomains, and let $t_1, \ldots, t_m$ be m time instants, with $0 = t_1 < \cdots < t_m \leq T$. Assume that for all $i \in \{1, \ldots, n\}$ and $j \in \{1, \ldots, m\}$,

$$z_{ij} = \frac{1}{|D_i|} \int_{D_i} f_0(\mathbf{p}, t_j)d\mathbf{p} + \varepsilon_{ij},$$

where the errors ε_{ij} are independent, with zero mean, and variance proportional to $1/|D_i|$. We estimate f_0 by minimizing the functional

$$J_T(f) = \sum_{i=1}^{n} \sum_{j=1}^{m} \frac{1}{|D_i|} \left[\int_{D_i} \{f(\mathbf{p}, t_j) - z_{ij}\}d\mathbf{p} \right]^2 + \lambda \int_0^T \int_\Omega (\partial f/\partial t + Lf - u)^2. \tag{12}$$

Sampling design 3. *Areal observations in space and interval observations in time*

Let $D_1, \ldots, D_n \subset \Omega$ be n disjoint spatial subdomains, and let let $T_1, \ldots, T_m \subset [0,T]$ be m disjoint temporal intervals. Assume that for all $i \in \{1, \ldots, n\}$ and $j \in \{1, \ldots, m\}$,

$$z_{ij} = \frac{1}{|D_i|\,|T_j|} \int_{T_j} \int_{D_i} f_0(\mathbf{p}, t)d\mathbf{p}dt + \varepsilon_{ij},$$

where the errors ε_{ij} are independent, with zero mean, and variance proportional to $1/(|D_i||T_j|)$. We estimate f by minimizing the functional

$$J_T(f) = \sum_{i=1}^{n} \sum_{j=1}^{m} \frac{1}{|D_i|\,|T_j|} \left[\int_{T_j} \int_{D_i} \{f(\mathbf{p}, t) - z_{ij}\}d\mathbf{p}dt \right]^2 + \lambda \int_0^T \int_\Omega (\partial f/\partial t + Lf - u)^2.$$

6.1 General formulation including different sampling designs

The three sampling designs detailed above can be unified under a general formulation. Indeed, for specific choices of functions $\alpha_i(\mathbf{p})$ and $\beta_j(t)$, the models and the associated estimation problems corresponding to the sampling designs 1–3 can be obtained as special cases of the model

$$z_{ij} = \frac{1}{\int_\Omega \int_0^T \alpha_i(\mathbf{p})\beta_j(t) d\mathbf{p}dt} \int_\Omega \int_0^T f_0(\mathbf{p},t)\alpha_i(\mathbf{p})\beta_j(t) d\mathbf{p}dt + \varepsilon_{ij} \tag{13}$$

for $i \in \{1,\dots,n\}$, and $j \in \{1,\dots,m\}$, where the errors ε_{ij} are independent, with zero mean, and variance proportional to $1/\int_0^T \int_\Omega \alpha_i\beta_j$, with the associated estimation functional

$$J_T(f) = \sum_{i=1}^{n}\sum_{j=1}^{m} \frac{1}{\int_\Omega \int_0^T \alpha_i(\mathbf{p})\beta_j(t) d\mathbf{p}dt} \left[\int_\Omega \int_0^T \{f(\mathbf{p},t) - z_{ij}\}\alpha_i(\mathbf{p})\beta_j(t) d\mathbf{p}dt\right]^2 + \lambda \int_0^T \int_\Omega (\partial f/\partial t + Lf - u)^2. \tag{14}$$

In particular, denote by $\mathbb{I}_{D_i}$ the characteristic function of D_i and by $\delta_{\mathbf{p}_i}(\mathbf{p})$ the Dirac mass located in $\mathbf{p}_i$, and define analogously $\mathbb{I}_{T_j}(t)$ and $\delta_{t_j}(t)$. Using this notation, the model and estimation functional corresponding to sampling design 1 are obtained from (13) and (14) setting $\alpha_i(\mathbf{p}) = \delta_{\mathbf{p}_i}(\mathbf{p})$ and $\beta_j(t) = \mathbb{I}_{T_j}(t)$; those corresponding to sampling design 2 are obtained setting $\alpha_i(\mathbf{p}) = \mathbb{I}_{D_i}(\mathbf{p})$ and $\beta_j(t) = \delta_{t_j}(t)$; finally, those corresponding to sampling design 3 are obtained setting $\alpha_i(\mathbf{p}) = \mathbb{I}_{D_i}(\mathbf{p})$ and $\beta_j(t) = \mathbb{I}_{T_j}(t)$.

6.2 Estimation problem

For the sampling designs described in the previous paragraph, the functional $J_T(f)$ is well defined. Indeed, point-wise evaluations in space are allowed because the functional space V, defined in Section 3, contains functions continuous in space, thanks to the embedding $H^2(\Omega) \subset C(\bar{\Omega})$ if $\Omega \subset \mathbb{R}^d$ with $d \leq 3$. In addition, V contains also functions that are in $C^0(0,T;H^1(\Omega))$. Therefore we can consider point-wise evaluations in time when we have areal data in space (which are continuous linear functionals in $H^1(\Omega)$). Both the penalty term and the least square term are thus well defined for functions in V. Other choices for α and β are of course possible. The proof of existence and uniqueness of the estimator given in this work does not cover the case of data observed point-wise in both space and time simultaneously.

The estimation problem under the general setting considered in Section 6.1 is formulated as Problem 1. The existence and uniqueness of the corresponding estimator is stated in the following proposition.

Proposition 3. *Under suitable regularity assumptions (see A) the solution of Problem 1, with J_T given in* (14) *and α, β as in Section 6.1, exists and is unique and is obtained by solving:*

$$\begin{cases} \partial \hat{f}/\partial t + L\hat{f} = u + \hat{g} & \text{in } \Omega \times (0,T], \\ \hat{f}(\mathbf{p},0) = s(\mathbf{p}) & \text{in } \Omega, \\ \mathcal{B}_c \hat{f} = h & \text{on } \partial\Omega \times (0,T], \end{cases}$$

$$\begin{cases} -\partial \hat{g}/\partial t + L^* \hat{g} = -\frac{1}{\lambda} \sum_{i=1}^{n} \sum_{i=1}^{m} \frac{\alpha_i \beta_j}{\int_\Omega \int_0^T \alpha_i \beta_j} \int_\Omega \int_0^T (\hat{f} - z_{ij}) \alpha_i \beta_j & \text{in } \Omega \times [0,T), \\ \hat{g}(\mathbf{p},T) = 0 & \text{in } \Omega, \\ {\mathcal{B}_c}^* \hat{g} = 0 & \text{on } \partial\Omega \times [0,T). \end{cases}$$

The proof of Proposition 3 is detailed in A. Note that we obtain a coupled system similar to (5)–(6), with only a difference in the right-hand side of the second equation, due to the different sampling designs here considered.

The coupled system in Proposition 3 can be discretized similarly to the discretization of (5)–(6), detailed in Section 4. Moreover, the properties of the corresponding estimators can be derived analogously, along the same lines in Section 5. Finally, also for these different sampling designs, the smoothing parameter λ may be chosen via the minimization of the GCV defined in Eq. (11).

7 Simulation studies

In this section we present some simulation studies that compare the proposed space-time regression with time-dependent PDE regularization (ST-tPDE) to four competing methods. The first competing method that we consider is spatio-temporal kriging with a separable variogram, with parameters estimated from the empirical variogram. This method is implemented using the functions `krigeST` and `fit.StVariogram` of the `R` package `gstat` [35]. The other methods we consider are based on differential regularization with two roughness terms that account separately for the regularity of the field in space and in time. All these methods use a tensor product approach.

The first method, denoted by the acronym TPS, adopts a thin plate spline basis in space and a cubic B-spline basis in time; the spatial penalty is the thin plate spline energy and the temporal penalty is the L^2 norm of the second derivative in time. The second method, denoted by the acronym SOAP and proposed by [3, 29], uses soap film smoothing in space [46] and cubic B-splines in time; the penalization is composed by the L^2 norm of the Laplacian in space and the L^2 norm of the second derivative in time. Both TPS and SOAP are implemented using the function `gam` of the `R` package mgcv [45]. The last model we consider, denoted by the acronym ST-PDE

and proposed in [7], employs finite elements in space and cubic B-Splines in time; this method penalizes the L^2 norm of the Laplacian in space and the L^2 norm of the second derivative in time. Both the proposed ST-tPDE method and the ST-PDE method by [7] are implemented in R and C++, based on the R package `fdaPDE` [27].

7.1 First simulation study: Areal observations in space and point-wise observations in time of a smooth function

For the first simulation study, we aim at comparing the performances of the proposed ST-tPDE to those of the competing ST-PDE presented in [7] in a simulation setting that mimics the applied problem stimulating this research. In particular, we consider areal observations in space and point-wise in time, corresponding to the sampling design 2. We cannot here compare to kriging, TPS or SOAP, because these methods are not currently implemented for areal data. Specifically, we consider a quasi-circular domain Ω in Figure 4, corresponding to the section of the carotid artery of one of the patients in MACAREN@MOX project; we moreover set the temporal domain to $[0, 1]$. We consider a function f_0 solution of the following heat equation, viz.

$$\begin{cases} \partial f_0/\partial t - \Delta f_0 = 0 & \text{in } \Omega \times (0, T), \\ f_0(\mathbf{p}, 0) = s(\mathbf{p}) & \text{in } \Omega, \\ \mathcal{B}_c f_0 = 0 & \text{on } \partial\Omega \times (0, T], \end{cases}$$

where the initial condition $s(\mathbf{p})$ is the solution of

$$\begin{cases} \Delta s = 1 & \text{in } \Omega, \\ \mathcal{B}_c s = 0 & \text{on } \partial\Omega. \end{cases}$$

We sample from this f_0, over the seven beams displayed in Figure 3 and at 11 equally-spaced time instants in $[0, 1]$, adding a Gaussian noise with mean zero and standard deviation $\sigma = 0.4$, corresponding to approximately 5% of the signal range.

The ST-PDE method in [7] is implemented using the penalty

$$\lambda_T \int_0^T \int_\Omega (\partial^2 f/\partial t^2)^2 + \lambda_S \int_0^T \int_\Omega (\Delta f)^2.$$

The proposed ST-tPDE is implemented with the penalty

$$\lambda \int_0^T \int_\Omega (\gamma \partial f/\partial t - \Delta f)^2.$$

Moreover, the initial condition on the field is estimated from the data, using the SR-PDE described in [5], with the penalization of the Laplacian

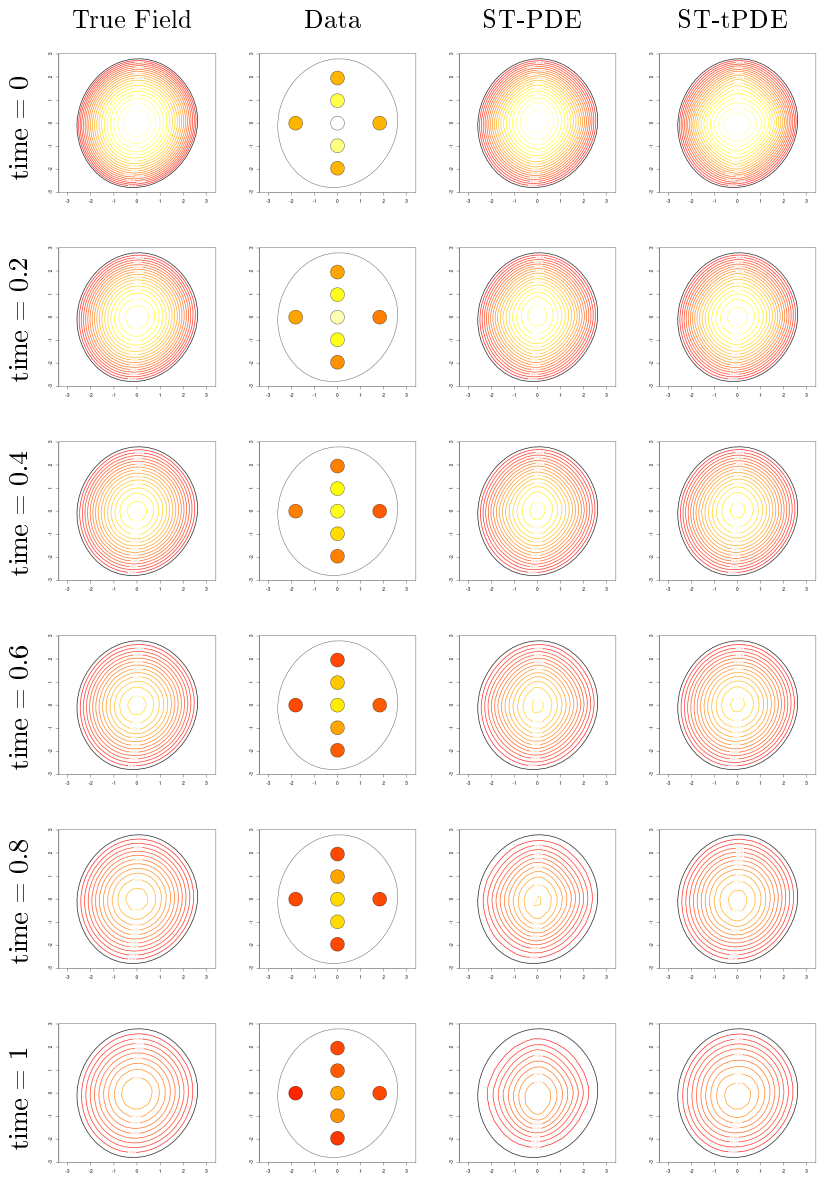

Figure 5: First simulation study. First column: True function, evaluated at different time instants. Second column: Sampled data at the same time instants. Third column: Corresponding estimates provided by ST-PDE. Fourth column: corresponding estimates provided by ST-tPDE.

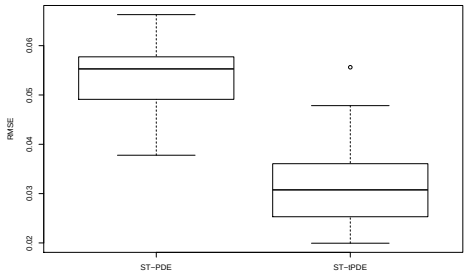

Figure 6: First simulation study. Boxplots of the RMSE over 50 simulation replicates of the estimates of the spatio-temporal field obtained by ST-PDE and ST-tPDE.

Δf. Both ST-PDE and ST-tPDE are thus considering the correct isotropic smoothing. Dirichlet homogeneous boundary conditions are enforced for both methods. For the spatial discretization, both methods employ linear finite elements over the triangulation shown in the left panel of Figure 4. For the time discretization, ST-PDE uses cubic B-splines with nodes at 11 time instants equally spaced in $[0, 1]$, while ST-tPDE uses finite differences at the same time instants. The smoothing parameters λ_S, λ_T and for ST-PDE and λ, γ for ST-tPDE are chosen via GCV.

Figure 5 shows in the first column the true field at different time instants, in the second column the data sampled at the same time instants, and in the third and fourth columns the corresponding estimates provided by ST-PDE and by the proposed ST-tPDE. It is difficult to visually appreciate the differences between the estimates provided by the two methods, which are only apparent at the last time instant. For this reason, to quantify these differences, we consider 50 repetitions of this simulation, corresponding to 50 different noise generations. Figure 6 shows the boxplots of the RMSE of the space-time field estimates given by the two methods over the 50 replicates. The RMSE is computed over a regular grid with 341 points in space and 41 points in time. The boxplots highlights that the best estimates are provided by the proposed ST-tPDE.

7.2 Second simulation study: Point-wise observations in space and time of a smooth function

In this second simulation we want to compare the proposed ST-tPDE to ST-PDE, TPS, SOAP and kriging. Since TPS, SOAP and kriging are not currently implemented to handle areal data, we here consider point-wise evaluations of f_0 in both space and time. Proposition 3 does not cover this sampling design, and the existence and uniqueness of the solution of the

infinite-dimension estimation Problem 1 has not been proven for this case; nevertheless, the corresponding discrete ST-tPDE estimator is well defined.

The true field f_0, the spatial domain Ω, and the temporal domain are the same considered in the first simulation study. We sample from f_0 at 100 space locations and at 11 equi-spaced time instants over $[0, 1]$, adding a Gaussian noise with mean zero and standard deviation of $\sigma = 0.4$ corresponding to approximately 5% of the signal range.

We implement kriging with marginally spherical variograms, both in space and in time. Other standard choices of variograms models have also been considered, with comparable results; hence to save space, these are not reported here. We add 50 spatial locations on the boundary of the domain, for which we set the data equal to zero, in order to improve the performance of kriging. TPS and SOAP are implemented under the standard settings in the `mgcv` package. ST-PDE and ST-tPDE are implemented as detailed in Section 7.1, with the only difference that the triangulation used for the discretization in space has 100 interior nodes, corresponding to the data locations. For SOAP, ST-PDE and ST-tPDE, homogeneous Dirichlet boundary conditions are used. All the smoothing parameters of TPS, SOAP, ST-PDE and ST-tPDE are chosen via GCV.

Figure 7 shows in the first column the true field evaluated at different time instants, in the second column the data sampled at the same instants, and from the third to the last columns the corresponding estimates provided respectively by kriging, TPS, SOAP, ST-PDE and the proposed ST-tPDE. From these plots we can see that kriging is not able to reconstruct the right isolines, while the other methods are performing better. Figure 8 shows the boxplots of the RMSE of the space-time field estimates over 50 replicates of the noise generation, with the RMSE computed as detailed in Section 7.1. From the boxplots we can see that kriging is giving the worst estimate in term of RMSE, while the best estimate is provided by ST-tPDE.

7.3 Third simulation study: Point-wise observations in space and time of a random field

As in the previous simulation study, we consider point-wise observations in space and time. The domain Ω is circular with radius $r = 1$; the temporal domain is $[0, 1]$. We generate from a spatio-temporal random field, with separable Matérn covariance structure, with smoothness parameter $\nu = 2$, constraining the field to be zero at the boundary of spatial domain or, more precisely, at 100 equally spaced points on $\partial\Omega$. The field is sampled at the 120 space locations and 11 equidistant time instants in the interval $[0, 1]$, adding a small Gaussian noise with mean zero and standard deviation equal to 0.035.

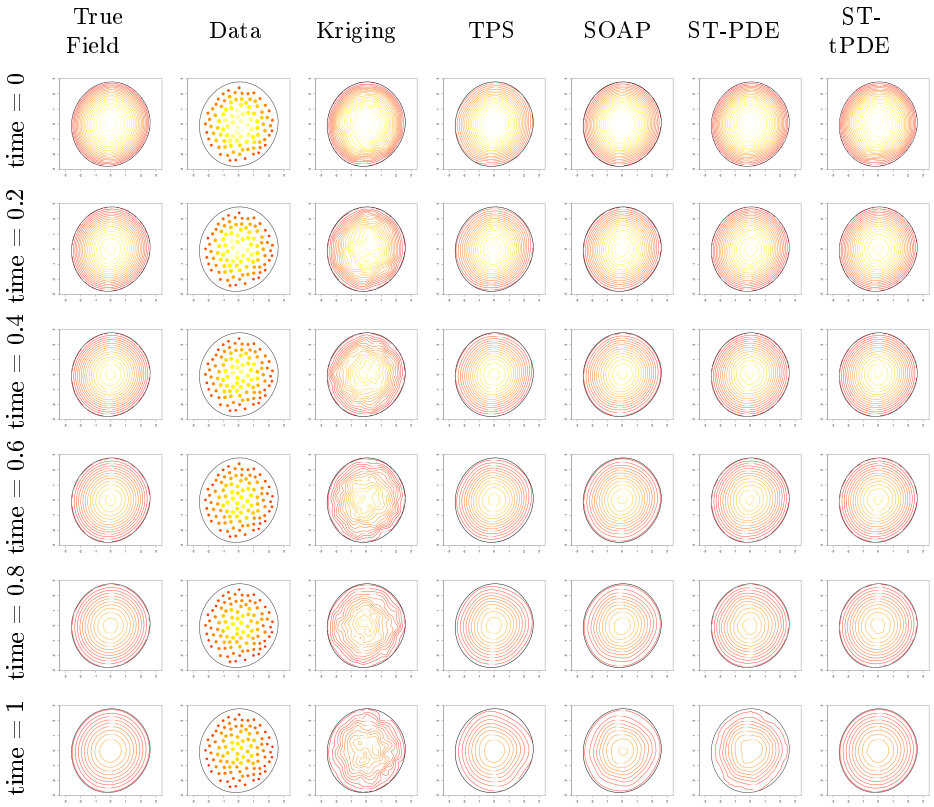

Figure 7: Second simulation study. First column: true function evaluated at different time instants. Second column: sampled data. Third column to last column: estimates provided by kriging, TPS, SOAP, ST-PDE and ST-tPDE.

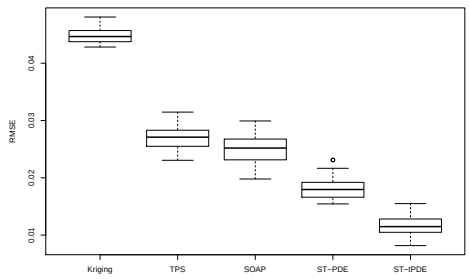

Figure 8: Second simulation study. Boxplots of the RMSE over 50 simulation replicates of the estimates of the spatio-temporal field obtained by kriging, TPS, SOAP, ST-PDE and ST-tPDE.

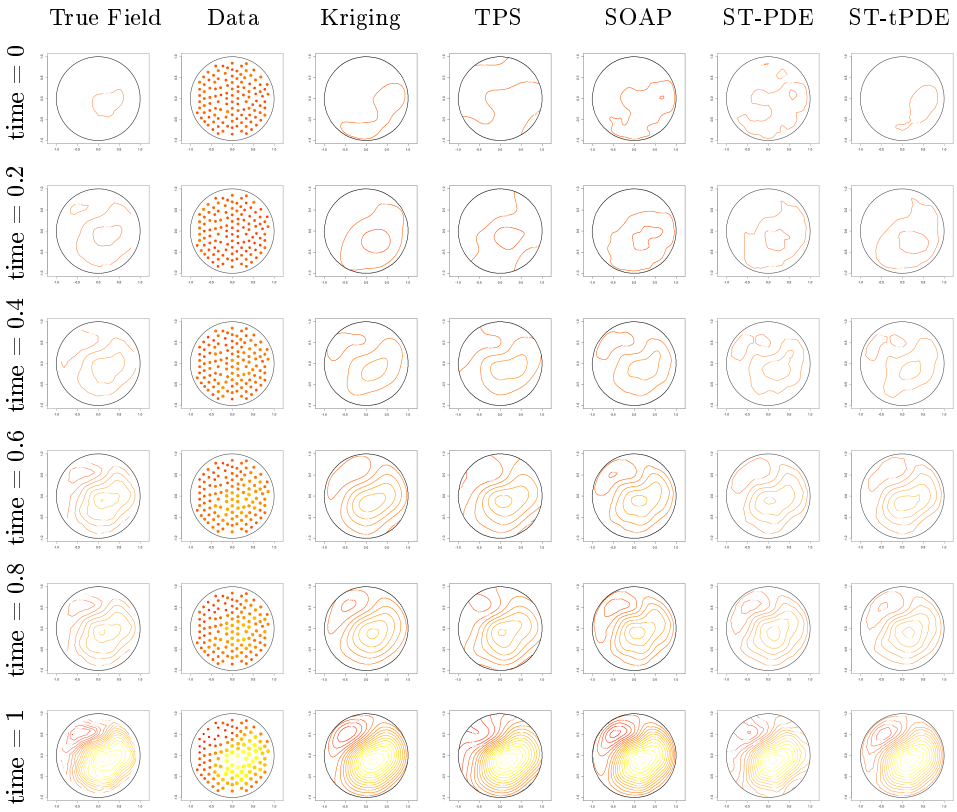

Figure 9: Third simulation study. First column: True function evaluated at different time instants. Second column: Sampled data. Third column to last column: Estimates provided by kriging, TPS, SOAP, ST-PDE and ST-tPDE.

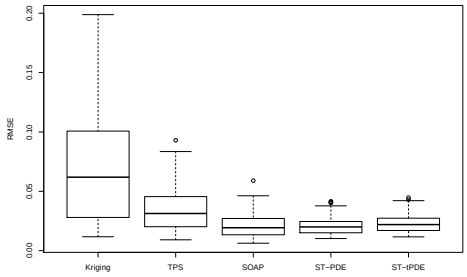

Figure 10: Third simulation study. Boxplots of the RMSE over 100 simulation replicates of the estimates of the spatio-temporal field obtained by kriging, TPS, SOAP, ST-PDE and ST-tPDE.

We implement kriging with the right covariance structure, separable Matérn in space and time, and using the right smoothness parameter $\nu = 2$. We implement the other methods as detailed for the second simulation study.

Figure 9 shows the true random field at some temporal instants, the corresponding data and the estimates provided by kriging, TPS, SOAP, ST-PDE and ST-tPDE. In order to compare the performances of the methods, we compute RMSE over 100 repetitions of this simulation, corresponding to different generation of both the random field and the noise. The RMSE is computed over a regular grid with 248 points in space and 41 points in time. Figure 10 shows the boxplots of the RMSE of the space-time field estimates given by the five methods. A visual inspection of the boxplots highlights that kriging provides the worst estimates, also characterized by the largest variance. This was not expected, since this method employs the same covariance structure used to generate the signal, and even the correct smoothness parameter. SOAP, ST-PDE and ST-tPDE have comparable performances with ST-PDE reaching in this case slightly smaller RMSE. This small advantage of ST-PDE over the proposed ST-tPDE is expected as the true field f_0, in this simulation study, has a separable covariance structure in space and time.

8 Application to blood flow velocity field estimation

We now want to apply the proposed ST-tPDE to the estimation of the dynamic blood velocity field on a cross-section of the common carotid artery. More precisely, our aim is to estimate the mean velocity field during an average heart beat.

As shown in Figure 2, the ECD scan runs over more than one heart beat. We thus compute a mean signal from the record over multiple heart beats, cutting the signal at starting time of the systolic phase, which is an easily detectable landmark. The signals over the seven beams are obtained from separate ECD measurements, as the measurement device can only scan one beam at a time. The measurements in different beams are thus aligned, using again as landmark the starting time of the systolic phase. Figure 3 displays the data for one subject in the study: the seven signals represent the mean velocity of the blood cells measured in the seven beams on the cross-section of the carotid. Different colors and numbers are used to denote the signals over the seven beams, following the color and number scheme displayed in the upper right corner of the figure. The ECD signals over the seven beams have different shapes. Specifically, the ECD signals corresponding to the central beam, beam 3, and to the beam in the upper part of the section, beam 2, have two peaks in the systolic phase, with the highest velocities being reached during the second peak. The ECD signals over the beams in the lower and lateral part of the artery section, beams 4 to 7, have only one main peak, that is earlier in time with respect to the main peak over the beams 1 to 3. We can also observe that higher velocities are reached over the beams in the central and lower part of the artery section, beams 2 to 5, with respect to the upper and lateral beams, beams 1, 6 and 7.

As detailed in [5], which considered these data for a fixed time instant, the systolic peak, we have a detailed problem specific information about blood fluid dynamics; see, e.g., [19]. This information can be conveniently translated into a partial differential equation, along with the physiological boundary conditions. In particular, we can here consider the parabolic PDE $\gamma \partial f/\partial t + Lf = 0$, where the spatial operator L is the same used in [5] for the estimation of the blood flow velocity at the systolic peak time. Specifically, L is given by the following diffusion, transport and reaction terms:

$$\mathbf{K}(x,y) = \begin{bmatrix} y^2 + \kappa_1 x^2 & (\kappa_1 - 1)xy \\ (\kappa_1 - 1)xy & x^2 + \kappa_1 y^2 \end{bmatrix} + \kappa_2 (R^2 - x^2 - y^2)\mathbf{I}_2,$$

$$\mathbf{b}(x,y) = (\beta x, \beta y)^\top, \quad c = 0,$$

where R is the artery radius, $R = 2.8$, while the hyperparameters in the diffusion and transport terms are set to $\kappa_1 = 0.1$, $\kappa_2 = 0.2$, and $\beta = 0.5$, on the basis of problem-specific information. Finally, the relative strength of the space and time derivatives is controlled via the parameter γ, which is set to 0.1. See [5] for the details of the derivation.

The starting velocity profile h_0, corresponding to the velocity field at the end of the diastolic phase, is estimated via the SR-PDE in [5], with the same specification of the spatial operator L. Likewise in [5], we moreover impose homogeneous Dirichlet boundary conditions on the wall of the carotid cross-section, i.e., $f|_{\partial\Omega} = 0$, corresponding to the physiological no-slip conditions.

The sampling design of these data corresponds to the sampling design 2, detailed in Section 6, i.e., areal data in space and point-wise data in time. The space-time velocity field is thus estimated minimizing the functional $J_T(f)$ in (12), with the PDE described above. The estimation problem is discretized in space using linear finite elements defined on the triangulation shown in the left panel of Figure 4. The discretization in time is obtained by means of the finite difference method, with a time discretization grid of $N_T = 41$ uniformly spaced points during the time of the heartbeat (0.92 seconds for the considered patient).

The estimated dynamic surface is represented in Figure 11 at fixed instants in time. We can observe that during the heartbeat, the shape of the velocity field is subject to strong variations. During the first instants of the systolic phase, the velocity field has a strong asymmetry with higher values in the lower part of the artery cross-section; the plot in the bottom right

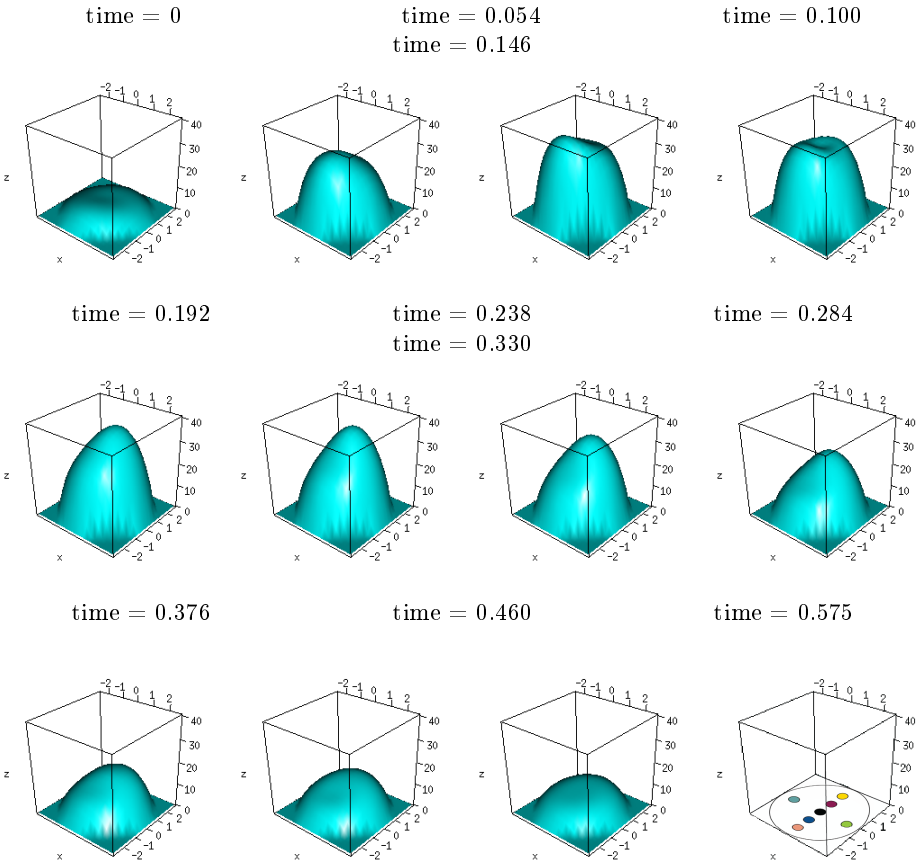

Figure 11: Blood flow velocity field estimated by ST-tPDE, at different time instants. The figure in the bottom right corner indicates the orientation of the carotid cross-section.

corner of Figure 11 gives the orientation of the carotid cross-section for this figure. In the subsequent instants, the shape of the velocity field changes, assuming higher values in the upper right part of the cross-section. These asymmetries and eccentricities of the blood flow are of strong interest to the medical doctors, to investigate how the heamodynamics may influence the pathogenesis of atherosclerosis. During the diastolic phase, however, the estimated velocity field is symmetric and flat.

We can compare the obtained estimates with the original data in Figure 3. Figure 12 displays the estimated velocity in the central point of the beams. Notice that the estimated dynamic surface captures very well the main features of the ECD signals. Moreover the estimate of the mean velocity on each beam borrows strength from the proximity of other beams, taking into account the spatial structure of the phenomenon. Penalizing a parabolic PDE that summarizes the problem-specific knowledge on the phenomenon thus allows to obtain a physiological estimate of the velocity field.

9 Directions of future research

Various extensions of the proposed model can be considered and will be the object of future research.

First, spatio-temporal varying covariates could be included in the model, using a semiparametric approach analogous to the one considered in [41].

Second, using a generalized linear framework similar to [44], we could model variables of interest having any distribution in the exponential family, such as Poisson counts, binomial or Gamma distributed outcomes, scattered over spatio-temporal domains. This model extension would significantly broaden the scope of the proposed model.

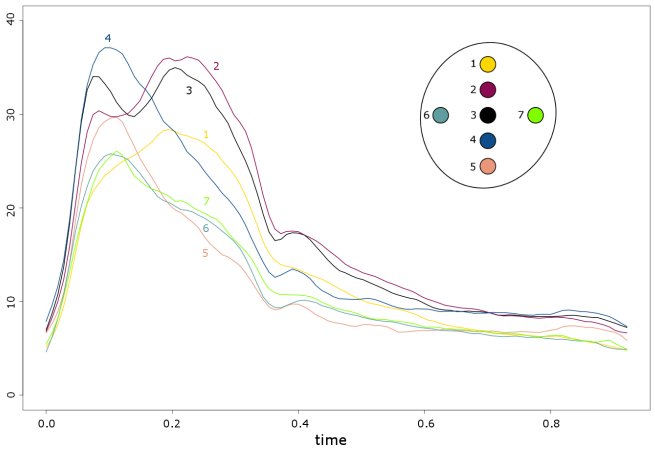

Figure 12: Estimated blood-flow velocity at the beams' centers.

A very interesting possible development would consist in combining the proposed method with a technique for data-driven estimation of the hyperparameters in the regularized PDE. In the described application to blood-flow velocity field estimation, these hyperparameters have been fixed on the basis of prior knowledge on the phenomenon under study, following the derivation in [5]. However, the problem specific knowledge may be not as detailed as to suggest specific values for such parameters. A promising approach in this respect is offered by the parameter cascading technique proposed in [38, 10, 47]. This technique is explored in the framework of spatial regression with differential regularization in [6], in the simpler case of only spatial data and with a simpler PDE with only a stationary diffusion operator $Lf = \mathrm{div}(\mathbf{K}\nabla f)$.

Another fascinating generalization would consist in handling spatio-temporal data distributed over general non-planar domains. This could be done using non-planar finite elements, as in [26], or isogeometric analysis based on non-rational B-spline basis, as in [43]. The generalization to non-planar domains would enable advanced applications in the life sciences, as well as in the geosciences and engineering. For instance, in the neurosciences, this would permit the analysis of spatio-temporal neuroimaging data associated to neuronal activity over the cerebral cortex, the highly convoluted thin sheet of neuronal tissue that hosts most of the neuronal activity. In cardiovascular research, this would for instance allow the study of spatio-temporal varying haemodynamical stresses exerted by blood-flow over the arterial wall. This study would be fundamental for advancing our knowledge on aneurysms' pathogenesis. A first study in this direction, considering a fixed time instant, the systolic peak, has been carried out in [17]. The use of non-rational B-spline basis, instead of finite elements, would instead be particularly suited to engineering applications, being these bases extensively used in computer aided design.

Acknowledgments

We are grateful to Sandro Salsa, for very helpful discussions on an earlier version of this work. We are also grateful to Christian Vergara, P.I. of the MACAREN@MOX project, and to Maurizio Domanin, the principal medical doctor of the project.

A Proof of the solution of the estimation problem

To prove the existence and the uniqueness of the solution of Problem 1, we consider the quantity $g \in \mathcal{G}_T = L^2(0,T;L^2(\Omega))$, defined as $g = \partial f/\partial t + Lf - u$, where L is the second order elliptic operator (2) and u is the known forcing term. We moreover define the space $V_{T,0} = \{v \in V : \ \mathcal{B}_c v = 0 \text{ and } v(\mathbf{p},0) = 0\}$, which represents the space of functions in V with homogeneous boundary

conditions and homogeneous initial value, and we introduce the operator $B_T : L^2(0,T;L^2(\Omega)) \to V_{T,0}$, such that $B_T\tilde{u}$ is the unique solution of the PDE (3) with forcing term $\tilde{u}$ and homogeneous boundary conditions, i.e., $\partial_t(B_T\tilde{u}) + L(B_T\tilde{u}) = \tilde{u}$ in $\Omega \times (0,T]$, $B_T\tilde{u}(\mathbf{p},0) = 0$ in Ω and $\mathcal{B}_c(B_T\tilde{u}) = 0$ on $\partial\Omega \times (0,T]$. We make the following assumptions.

Assumption 1. *$\Gamma_D \neq \emptyset$, so that a Poincaré inequality holds, i.e.,*

$$\|v\|_{L^2(\Omega)} \leq C_P \|\nabla v\|_{L^2(\Omega)}.$$

Assumption 2. *The parameters of the PDE are such that for all $\tilde{u} \in L^2(0,T;L^2(\Omega))$, there exists a unique solution f_0 of the PDE (3), which moreover satisfies*

$$f_0 \in L^2(0,T;H^2(\Omega)) \cap C^0(0,T;H^1_{\Gamma_D}(\Omega)),$$

$$\partial f_0/\partial t \in L^2(0,T;L^2(\Omega))$$

and $\|f_0\|_V \leq \beta \|\tilde{u}\|_{L^2(0,T;L^2(\Omega))}$.

Under Assumptions 1 and 2, thanks to the well-posedness and the regularity of the PDE (3), the operator B_T is an isomorphism between $L^2(0,T;L^2(\Omega))$ and $V_{T,0}$. Moreover, the following inequality holds:

$$\int_0^T \|B_T\tilde{u}(t)\|^2_{H^2(\Omega)} \leq C \int_0^T \|\tilde{u}(t)\|^2_{L^2(\Omega)}. \tag{15}$$

The solution of the PDE (3) can thus be written as $f = f_b + B_T\tilde{u}$, where f_b is the solution of the PDE with homogeneous forcing term, non-homogeneous initial value and non-homogeneous boundary conditions.

Existence and uniqueness of the estimator $\hat{f}$ can hence be obtained thanks to a classical result of calculus of variations. We recall here the result stated, e.g., in [28].

Theorem 1. *Let $\mathcal{G}$ be an Hilbert space, $\mathcal{A} : \mathcal{G} \times \mathcal{G} \to \mathbb{R}$ a continuous, coercive and symmetric bilinear form in $\mathcal{G}$, $\mathcal{L} : \mathcal{G} \to \mathbb{R}$ a linear operator over $\mathcal{G}$, and c a constant. If the functional $J(g)$ has the form*

$$J(g) = \mathcal{A}(g,g) + \mathcal{L}g + c, \tag{16}$$

then there exists a unique $\hat{g} \in \mathcal{G}$ such that $J(\hat{g}) = \inf_{\mathcal{G}} J(g)$. Moreover $\hat{g}$ satisfies the following Euler–Lagrange equation

$$\forall_{\varphi \in \mathcal{G}} \; (J'(\hat{g}), \varphi) = 2\mathcal{A}(\hat{g}, \varphi) + \mathcal{L}\varphi = 0. \tag{17}$$

Exploiting this theorem, we can prove Proposition 1 and 3.

Proof of Proposition 1. Recalling the definitions of g, $\mathcal{G}_T$, B_T given at the beginning of this appendix, and the definition of f_b given below equation (15), we can write any $f \in V$ as an affine transformation of g, by $f = f_b + B_T(u+g)$, and we can re-express the functional (4) as the following functional $J_{T,g}$ over $\mathcal{G}_T$:

$$J_{T,g}(g) = J_T\{f_b + B_T(u+g)\} = \\ = \sum_{i=1}^{n} \int_0^T \{B_T(u+g)(\mathbf{p}_i,t) + f_b(\mathbf{p}_i,t) - z_i(t)\}^2 + \lambda \int_0^T \|g(t)\|^2_{L^2(\Omega)}\,.$$

Moreover, we can rewrite $J_{T,g}$ in the quadratic form (16) of Theorem 1, setting

$$\mathcal{A}(g,\varphi) = \sum_{i=1}^{n} \int_0^T B_T g(\mathbf{p}_i,t) B_T\varphi(\mathbf{p}_i,t) + \lambda \int_0^T \int_\Omega g\varphi,$$
$$\mathcal{L}\varphi = 2\sum_{i=1}^{n} \int_0^T B_T\varphi(\mathbf{p}_i,t)\left\{B_T u(\mathbf{p}_i,t) + f_b(\mathbf{p}_i,t) - z_i(t)\right\},$$
$$c = \sum_{i=1}^{n} \int_0^T \{B_T u(\mathbf{p}_i,t) + f_b(\mathbf{p}_i,t) - z_i(t)\}^2\,.$$

Given that B_T, the point-wise evaluation of a function and the integration on an interval are linear operators, we have that $\mathcal{A}(g,\varphi)$ is a bilinear form on $\mathcal{G}_T$. Moreover, it is continuous in $\mathcal{G}_T$. Indeed, thanks to the embedding $H^2(\Omega) \subset C(\bar{\Omega})$ if $\Omega \subset \mathbb{R}^d$ with $d \leq 3$, and thanks to (15), we have that

$$\int_0^T |B_T g(\mathbf{p}_i,t)|^2 \leq \int_0^T \|B_T g(t)\|^2_{C(\bar{\Omega})} \leq \\ \leq C\int_0^T \|B_T g(t)\|^2_{H^2(\Omega)} \leq \bar{C}\int_0^T \|g(t)\|^2_{L^2(\Omega)}\,.$$

We thus obtain that

$$\mathcal{A}(g,\varphi) \leq (\bar{C}n + \lambda)\left\{\int_0^T \|g(t)\|^2_{L^2(\Omega)}\right\}^{1/2}\left\{\int_0^T \|\varphi(t)\|^2_{L^2(\Omega)}\right\}^{1/2}.$$

Finally, the operator $\mathcal{A}(g,\varphi)$ is coercive in $\mathcal{G}_T$ because

$$\mathcal{A}(g,g) = \sum_{i=1}^{n} \int_0^T |B_T g(\mathbf{p}_i,t)|^2 + \lambda \int_0^T \|g(t)\|^2_{L^2(\Omega)} \geq \lambda \int_0^T \|g(t)\|^2_{L^2(\Omega)}\,.$$

As a result of the fact that the bilinear form $\mathcal{A}(\cdot,\cdot)$ is continuous and coercive in the Hilbert space $\mathcal{G}_T$, that the operator $\mathcal{L}: \mathcal{G}_T \to \mathbb{R}$ is linear, and

that c is a constant, Theorem 1 states the existence and the uniqueness of $\hat{g} = \text{argmin}_{g \in \mathcal{G}_T} J_{T,g}(g)$. From the bijectivity of $B_T : L^2(0,T;L^2(\Omega)) \to V_{T,0}$, we then deduce the existence and uniqueness of $\hat{f} = f_b + B_T(\hat{g} + u) = \text{argmin}_{f \in V} J_T(f)$.

Thanks to Theorem 1 and the definition of the operator B_T, we can obtain the surface estimator $\hat{f}$ as the solution of the PDE

$$\begin{cases} \partial \hat{f}/\partial t + L\hat{f} = u + \hat{g} & \text{in } \Omega \times (0,T], \\ \hat{f}(\mathbf{p},0) = s(\mathbf{p}) & \text{in } \Omega, \\ \mathcal{B}_c \hat{f} = h & \text{on } \partial\Omega \times (0,T], \end{cases} \tag{18}$$

where $\hat{g}$ is obtained as the solution of Eq. (17), that can be written as

$$\frac{1}{2}\left(J'(\hat{g}), \varphi\right) = \sum_{i=1}^{n} \int_0^T \left(B_T g + B_T u + f_b - z_i\right) B_T \varphi \; \delta_{\mathbf{p}_i} + \lambda \int_0^T \int_\Omega g\varphi = 0.$$

This equation corresponds to the PDE

$$\begin{cases} -\partial \hat{g}/\partial t + L^* \hat{g} = -\sum_{i=1}^{n} (\hat{f} - z_i)\delta_{\mathbf{p}_i}/\lambda & \text{in } \Omega \times [0,T), \\ \hat{g}(\mathbf{p},T) = 0 & \text{in } \Omega, \\ \mathcal{B}_c{}^* \hat{g} = 0 & \text{on } \partial\Omega \times [0,T), \end{cases} \tag{19}$$

where $\delta_{\mathbf{p}_i}$ is the Dirac mass located in $\mathbf{p}_i$, L^* is the adjoint operator of L, defined in (2) and $\mathcal{B}_c{}^*$ is the operator that defines the boundary conditions of the adjoint problem. The surface estimator can thus be written as the solution of a coupled system of the second order PDEs (18) and (19). □

Proof of Proposition 3. The same strategy used to prove Proposition 1 can be followed to prove Proposition 3. The only difference is the presence of α_i and β_j. Clearly, the point-wise evaluation of a function and the integration on an interval are linear operators; thus the expressions of α_i and β_j given in Section 6.1 preserve the bilinearity of $\mathcal{A}$.

The other important point is the continuity of $\mathcal{A}$. For the sampling design 1, i.e., point-wise observations in space and interval observations in time, the continuity of $\mathcal{A}$ is ensured by the fact that B_T is an isomorphism between $L^2(0,T;L^2(\Omega))$ and $V_{T,0}$ and by the embedding $H^2(\Omega) \subset C(\bar{\Omega})$. For the sampling design 2, i.e., areal observations in space and point-wise observations in time, the continuity of $\mathcal{A}$ holds thanks to the fact that $B_T g \in C^0(0,T;H^1(\Omega))$. Finally, for the sampling design 3, i.e., areal observations in space and interval observations in time, the continuity of $\mathcal{A}$ trivially holds. □

References

[1] M.C. Aguilera-Morillo, M. Durbán, A.M. Aguilera, Prediction of functional data with spatial dependence: A penalized approach, Stochastic Environ Res Risk Assess 31 (2017) 7–22.

[2] E. Arnone, Regression with PDE penalization for modelling functional data with spatial and spatio-temporal dependence, PhD thesis, Politecnico di Milano, 2018.

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

[4] L. Azzimonti, F. Nobile, L.M. Sangalli, P. Secchi, Mixed finite elements for spatial regression with PDE penalization, SIAM/ASA J. Uncert. Quant. 2 (2014) 305–335.

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

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

[7] M.S. Bernardi, L.M. Sangalli, G. Mazza, J.O. Ramsay, A penalized regression model for spatial functional data with application to the analysis of the production of waste in Venice province, Stoch. Environ. Res. Risk Assess. 31 (2017) 23–38.

[8] A. Buja, T.J. Hastie, R.J. Tibshirani, Linear smoothers and additive models, Ann. Statist. (1989) 153–555.

[9] W. Caballero, R. Giraldo, J. Mateu, A universal kriging approach for spatial functional data, Stoch. Environ. Res. Risk Assess. 27 (2013) 1553–1563.

[10] J. Cao, G.F. Fussmann, J.O. Ramsay, Estimating a predator-prey dynamical model with the parameter cascades method, Biometrics 64 (2008) 959–967.

[11] D.D. Cox, Asymptotics for m-type smoothing splines. Ann. Statist. 11 (1983) 530–551.

[12] D.D. Cox, Multivariate smoothing spline functions, SIAM J. Numer. Anal. 21 (1984) 798–813.

[13] N. Cressie, Statistics for Spatial Data, Wiley, New York, 2015.

[14] N. Cressie, C.K. Wikle, Statistics for Spatio-Temporal Data, Wiley, Hoboken, NJ, 2011.

[15] P. Delicado, R. Giraldo, C. Comas, J. Mateu, Statistics for spatial functional data: Some recent contributions, Environmetrics 21 (2010) 224–239.

[16] P.J. Diggle, P.J. Ribeiro Jr., Model-Based Geostatistics, Springer, New York, 2007.

[17] B. Ettinger, S. Perotto, L.M. Sangalli, Spatial regression models over two-dimensional manifolds, Technical Report 54/2012, MOX - Dipartimento di Matematica, Politecnico di Milano, 2012. Available at `http://mox.polimi.it/it/progetti/pubblicazioni/`.

[18] F. Ferraty. P. Vieu, Nonparametric Functional Data Analysis: Theory and Practice, Springer, New York, 2006.

[19] L. Formaggia, A. Quarteroni, A. Veneziani, Eds., Cardiovascular Mathematics: Modeling and Simulation of the Circulatory System, Springer-Verlag Mailand, 2009.

[20] R. Giraldo, P. Delicado, J. Mateu, Ordinary kriging for functionvalued spatial data, Environ. Ecol. Stat. 18 (2011) 411–426, 2011.

[21] O. Grujic, A. Menafoglio, J. Caers, Cokriging for multivariate Hilbert space valued random fields, application to multifidelity computer code emulation. Stoch. Environ. Res. Risk Assess. 2017. to appear.

[22] T.J. Hastie, R.J. Tibshirani, Generalized Additive Models, Chapman & Hall, London, 1990.

[23] L. Horváth, P. Kokoszka, Inference for Functional Data with Applications, Springer, New York, 2012.

[24] R. Ignaccolo, J. Mateu, R. Giraldo, Kriging with external drift for functional data for air quality monitoring, Stoch. Environ. Res. Risk Assess. 28 (2014) 1171–1186.

[25] P. Kokoszka, M. Reimherr, Introduction to Functional Data Analysis, CRC Press, Boca Raton, FL, 2017.

[26] E. Lila, J.A.D. Aston, L.M. Sangalli, Smooth principal component analysis over two-dimensional manifolds with an application to neuroimaging, Ann. Appl. Stat. 10 (2016) 1854–1879.

[27] E. Lila, L.M. Sangalli, J.O. Ramsay, L. Formaggia, `fdaPDE`: Regression with Partial Differential Regularizations, Using the Finite Element Method, 2016. R Package version 0.1-1.

[28] J.L. Lions, Optimal Control of Systems Governed by Partial Differential Equations, Springer, New York, Berlin, 1971.

[29] G. Marra, D.L. Miller, L. Zanin, Modelling the spatiotemporal distribution of the incidence of resident foreign population, Stat. Neerl. 66 (2012) 133–160.

[30] J. Mateu, E. Romano, Advances in spatial functional statistics, Stoch. Environ. Res. Risk Assess. 31 (2017) 1–6.

[31] A. Menafoglio, P. Secchi, M. Dalla Rosa, A universal kriging predictor for spatially dependent functional data of a Hilbert space, Electron. J. Statist. 7 (2013) 2209–2240.

[32] A. Menafoglio, A. Guadagnini, P. Secchi, A kriging approach based on Aitchison geometry for the characterization of particle-size curves in heterogeneous aquifers, Stoch. Environ. Res. Risk Assess. 28 (2014) 1835–1851.

[33] A. Menafoglio, P. Secchi, Statistical analysis of complex and spatially dependent data: A review of object oriented spatial statistics, Europ. J. Oper. Res. 258 (2017) 401–410. 2017.

[34] D. Nerini, P. Monestiez, C. Manté, Cokriging for spatial functional data, J. Multivariate Anal. 101 (2010) 409–418.

[35] E.J. Pebesma, Multivariable geostatistics in S: The `gstat` package, Comput. Geosciences 30 (2004) 683–691.

[36] A. Quarteroni, R. Sacco, F. Saleri, Numerical Mathematics, Springer-Verlag, Berlin, 2007.

[37] A. Quarteroni, A. Valli, Numerical Approximation of Partial Differential Equations, Springer, Berlin, 2008.

[38] J.O. Ramsay, G. Hooker, D. Campbell, J. Cao, Parameter estimation for differential equations: A generalized smoothing approach, J. R. Stat. Soc. Ser. B (Stat. Methodol.) 69 (2007) 741–796.

[39] J.O. Ramsay, B.W. Silverman. Functional Data Analysis, Springer, New York, 2005.

[40] T. Ramsay, Spline smoothing over difficult regions, J. R. Stat. Soc. Ser. B (Stat. Methodol.) 64 (2002) 307–319.

[41] L.M. Sangalli, J.O. Ramsay, T. Ramsay, Spatial spline regression models, J. R. Stat. Soc. Ser. B (Stat. Methodol.) 75 (2013) 681–703.

[42] C.J. Stone, Optimal global rates of convergence for nonparametric regression. Ann. Statist. 10 (1982) 1040–1053.

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

[44] M. Wilhelm, L.M. Sangalli, Generalized spatial regression with differential regularization, J. Stat. Comput. Simul. 86 (2016) 2497–2518.

[45] S.N. Wood, Generalized additive models: An introduction with R, CRC Press, Boca Raton, FL, 2006.

[46] S.N. Wood, M.V. Bravington, S.L. Hedley, Soap film smoothing, J. R. Stat. Soc. Ser. B (Stat. Methodol.) 70 (2008) 931–955.

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

MOX Technical Reports, last issues

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

46/2018 Riccobelli, D.; Ciarletta, P.
Morpho-elastic model of the tortuous tumour vessels

47/2018 Stefanucci, M.; Sangalli, L.M.; Brutti, P.
PCA-based discrimination of partially observed functional data, with an application to Aneurisk65 dataset

44/2018 Bernardi, M.S.; Sangalli, L.M.
Modelling spatially dependent functional data by spatial regression with differential regularization

45/2018 Bernardi, M.S.; Carey, M.; Ramsay, J.O.; Sangalli, L.M.
Modeling spatial anisotropy via regression with partial di fferential regularization

43/2018 Fontana, L.; Masci, C.; Ieva, F.; Paganoni, A.M.
Performing Learning Analytics via Generalized Mixed-Effects Trees

42/2018 Antonietti, P.F.; Melas, L.
Algebraic multigrid schemes for high-order discontinuous Galerkin methods

41/2018 Mazzieri, I.; Melas, L.; Smerzini, C.; Stupazzini, M.
The role of near-field ground motion on seismic risk assessment in large urban areas

38/2018 Domanin, M.; Gallo, D.; Vergara, C.; Biondetti, P.; Forzenigo, L.V.; Morbiducci, U.
Prediction of long term restenosis risk after surgery in the carotid bifurcation by hemodynamic and geometric analysis

39/2018 Ferro, N.; Micheletti, S.; Perotto, S.
Density-based inverse homogenization with anisotropically adapted elements

40/2018 Chiappa, A.S.; Micheletti, S.; Peli, R.; Perotto, S.
Mesh adaptation-aided image segmentation