

Mixed Finite Elements for Spatial Regression with PDE Penalization*

Laura Azzimonti[†], Fabio Nobile[‡], Laura M. Sangalli[†], and Piercesare Secchi[†]

Abstract. We study a class of models at the interface between statistics and numerical analysis. Specifically, we consider nonparametric regression models for the estimation of spatial fields from pointwise and noisy observations, which account for problem-specific prior information, described in terms of a partial differential equation governing the phenomenon under study. The prior information is incorporated in the model via a roughness term using a penalized regression framework. We prove the well-posedness of the estimation problem, and we resort to a mixed equal order finite element method for its discretization. Moreover, we prove the well-posedness and the optimal convergence rate of the proposed discretization method. Finally the smoothing technique is extended to the case of areal data, particularly interesting in many applications.

Key words. mixed finite element method, fourth order problems, nonparametric regression, smoothing

AMS subject classifications. 65N30, 65N15, 65N21, 62G08, 62G20

DOI. 10.1137/130925426

1. Introduction. In this work we study the properties of a nonparametric regression technique for the estimation of bidimensional or three dimensional fields on bounded domains starting from pointwise noisy evaluations. The technique is particularly well suited for applications in physics, engineering, biomedicine, etc., where prior knowledge of the field might be available from physical principles and should be taken into account in the field estimation or smoothing process. We consider, in particular, phenomena where the field can be described by a partial differential equation (PDE) and has to satisfy some known boundary conditions.

Spatial regression with PDE penalization (SR-PDE) has been developed in [1] for the estimation of the blood velocity field on an artery section, from echo-Doppler data. This technique has very broad applicability since PDEs are commonly used to describe phenomena behavior in many fields of physics, mechanics, biology, and engineering. Many applications of particular interest can be named: the estimation of the concentration of pollutant released in water or in the air and transported by the stream or by the wind starting from noisy observations, the estimation of temperature or pressure fields from electronic control units or sensors in environmental sciences, and many other phenomena in physical and biological sciences or engineering. In this work we focus on phenomena that are well described by linear

*Received by the editors June 17, 2013; accepted for publication (in revised form) April 3, 2014; published electronically July 22, 2014. This work was supported by MIUR Ministero dell'Istruzione dell'Università e della Ricerca, *FIRB Futuro in Ricerca* research project "Advanced statistical and numerical methods for the analysis of high dimensional functional data in life sciences and engineering" (<http://mox.polimi.it/users/sangalli/firbSNAPLE.html>), and by the program Dote Ricercatore Politecnico di Milano - Regione Lombardia, research project "Functional data analysis for life sciences."

<http://www.siam.org/journals/juq/2/92542.html>

[†]MOX - Dipartimento di Matematica, Politecnico di Milano, 20133 Milano, Italy (laura.azzimonti@polimi.it, laura.sangalli@polimi.it, piercesare.secchi@polimi.it).

[‡]MATHICSE-CSQI, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland (fabio.nobile@epfl.ch).

second order elliptic PDEs, typically transport-reaction-diffusion problems. Methods similar to SR-PDE, applied to problems in hemodynamics, are described in [9, 10, 24].

SR-PDE uses a functional data analysis approach (see, e.g., [22]) and generalizes classical spatial smoothing techniques, such as thin-plate splines. SR-PDE, in fact, estimates the surface or the field minimizing a penalized least square functional, with the roughness penalty involving a partial differential operator. Many methods for surface estimation define the estimate as the minimizer of a penalized sum-of-square-error functional, with the penalty term involving a simple partial differential operator. Thin-plate spline smoothing, for example, penalizes an energy functional in \mathbb{R}^2 that involves second order derivatives. The minimizer of this functional belongs to the linear space generated by the Green's functions associated to the bi-Laplacian (see [30] for details). Thin-plate spline smoothing was first extended to the case of bounded domains in [26], where the thin-plate energy is computed only over the bounded domain of interest. Since the minimizer cannot be directly characterized, it is approximated by a surface in the space of tensor product B-splines. Recently, more complex smoothing methods have been developed that deal with general bounded domains in \mathbb{R}^2 and general boundary conditions. Some examples are soap-film smoothing, described in [31], and the spatial spline regression models described in [25], which generalize the finite element L-splines introduced in [23]. These methods estimate bidimensional surfaces on complex bounded domains penalizing the Laplace operator of the surface as a measure of the local curvature. Soap-film smoothing approximates the minimizer of the penalized least square functional with a linear combination of Green's functions of the bi-Laplacian on the domain of interest, centered on the vertices of a lattice. On the other hand, finite element L-splines and spatial spline regression models solve directly the PDE associated to the penalized least square functional by means of a mixed finite element method.

Following the approach presented in [23, 25], we propose estimating the field minimizing a least square functional regularized with the L^2 -norm over the domain of interest of the misfit of a second order PDE, $Lf = u$, modeling the phenomenon under study. The important novelty with respect to the methods cited above is that the problem-specific prior information, formalized in the PDE, is here used to model the phenomenon space variation. Furthermore, SR-PDE allows for important modeling flexibility in this direction, accounting, for instance, for space anisotropy and nonstationarity in a straightforward way.

We assume here that all the parameters appearing in the operator L and the boundary conditions are known, while the forcing term in the PDE is not completely determined. This approach is similar to that used in control theory when a distributed control on the forcing term is considered; see, for example, [18, 19, 21]. However, the viewpoint that we have in this work is quite different from the classical control theory framework. We are in fact interested in a surface estimation problem starting from partial, pointwise, and noisy evaluations, instead of considering an integral target. For this reason, different questions have to be addressed in this partial and pointwise data framework, with respect to the issues classically considered in control theory. It is, for instance, necessary to require higher regularity to the field to ensure that the penalized least square functional is well defined. On the other hand, classical issues handled in control theory, such as the presence of constraints in the space of controls, are not relevant in this framework. Similarly, SR-PDE can be interpreted as an inverse problem [11], where the goal is specifically the determination of a field starting from direct partial and

pointwise observations. A similar setting is considered in [27].

The penalized least square functional has a unique minimum in the Sobolev space H^2 , and the minimum is the solution of a fourth order problem. In order to prove the existence and the uniqueness of the estimator we resort to a mixed approach for fourth order problems since the penalized error functional is not necessarily convex in H^2 . Accordingly, a mixed equal order finite element method, similar to classical mixed methods described, for example, in [5, 20], is used for discretizing the estimation problem. Other classical conforming and nonconforming methods (see [5] and references therein) or more recent discontinuous Galerkin methods (see, e.g., [3, 28, 15]) can be used for the discretization of the fourth order problem. However, in the specific case considered here the mixed finite element method is a convenient choice since the problem under exam can be written as a system of second order PDEs. Moreover, the mixed approach also provides a good approximation of second order derivatives of the field that can be useful in order to compute physical quantities of interest.

The proposed mixed equal order finite element discretization is known to have a suboptimal convergence rate when applied to fourth order problems with arbitrary boundary conditions, and, in particular, the first order approximation might not converge to the exact solution (see, e.g., [4, 5]). However, we are able to prove the optimal convergence of the proposed discretization method for the specific set of boundary conditions that are naturally associated to the smoothing problem, whenever the true underlying field satisfies exactly those conditions. The theoretical results are confirmed by numerical experiments.

The inspected convergence concerns the study of the bias of the estimator, while the study of the variance of the estimator and the convergence when the number of observations goes to infinity will be the subject of a future work. These topics are studied in the classical setting of smoothing splines (see, e.g., [6]), thin-plate splines, or multidimensional smoothing splines (see, e.g., [7, 8, 16] and references therein), but the results proved in these classical settings cannot be directly generalized to SR-PDE models.

The proposed smoothing technique is also extended to the case of areal data, i.e., data that represent linear quantities computed on some subdomains; this data framework is frequent in many applications. For instance, in the case of the driving problem considered in [1], which concerns the blood velocity field estimation from echo-Doppler acquisitions, the data represent the mean velocity of blood on some subdomains on the considered artery section. The properties of the estimator in the areal setting are obtained along the same lines followed for pointwise observations.

The paper is organized as follows. Section 2 introduces the SR-PDE model used for pointwise observations. Section 3 proves the well-posedness of the estimation problem, and section 4 obtains a bound for the bias of the estimator. Section 5 describes the mixed finite element method used for the discretization of the estimation problem and proves the well-posedness of the discrete problem. Section 6 proves the convergence of the proposed mixed finite element method and provides a bound for the bias of the finite element estimator. Section 7 extends the models to the case of areal data and presents the asymptotic results in this setting. Section 8 presents the numerical experiments supporting the theoretical results. Section 9 outlines future research directions.

2. Surface estimator for pointwise data. Consider a bounded, regular, open domain $\Omega \subset \mathbb{R}^d$ with $d \leq 3$ and a regular function $f_0 : \Omega \rightarrow \mathbb{R}$ to be estimated from noisy observations. Let z_i , for $i = 1, \dots, n$, be n observations that represent noisy evaluations of the field f_0 at points $\mathbf{p}_i \in \Omega$. We assume in this work that the location points \mathbf{p}_i are fixed and known without error. As is usually done in nonparametric and smoothing problems, we assume the classical additive error model for the observations:

$$(2.1) \quad z_i = f_0(\mathbf{p}_i) + \epsilon_i,$$

where ϵ_i , $i = 1, \dots, n$, are independent errors with mean $\mathbb{E}[\epsilon_i] = 0$, and constant variance $\text{Var}(\epsilon_i) = \sigma^2$.

In addition to the observations z_i , it is supposed that we have a physical knowledge of the phenomenon under study and that this prior knowledge can be described by means of a differential operator. Specifically, we can formalize this as a PDE that f_0 satisfies:

$$(2.2) \quad \begin{cases} Lf_0 = \tilde{u} & \text{in } \Omega, \\ \mathcal{B}_c f_0 = \gamma & \text{on } \partial\Omega, \end{cases}$$

where the operator L and the boundary conditions are completely determined and fixed, while the forcing term $\tilde{u} = u + g_0 \in L^2(\Omega)$ is composed of a known and fixed part $u \in L^2(\Omega)$, motivated by the application, and an unknown term $g_0 \in L^2(\Omega)$, which will be estimated from data. The parameters of the PDE and the boundary conditions could as well be considered partly unknown and estimated from data, but in this work we assume them to be known and fixed. We focus on second order elliptic operators; in particular, L is a diffusion-transport-reaction operator

$$(2.3) \quad Lf_0 = -\text{div}(\mathbf{K}\nabla f_0) + \mathbf{b} \cdot \nabla f_0 + cf_0$$

with smooth and bounded parameters. The matrix $\mathbf{K} \in \mathbb{R}^{d \times d}$ is a symmetric and positive definite diffusion tensor, $\mathbf{b} \in \mathbb{R}^d$ is the transport vector, and $c \geq 0$ is the reaction term. These parameters can be spatially varying in Ω ; i.e., $\mathbf{K} = \mathbf{K}(\mathbf{x})$, $\mathbf{b} = \mathbf{b}(\mathbf{x})$, and $c = c(\mathbf{x})$, with $\mathbf{x} \in \Omega$. The boundary conditions of the PDE are homogeneous or nonhomogeneous Dirichlet, Neumann, or Robin (or mixed) conditions. All the admissible boundary conditions are summarized in

$$(2.4) \quad \mathcal{B}_c f_0 = \begin{cases} f_0 & \text{on } \Gamma_D, \\ \mathbf{K}\nabla f_0 \cdot \boldsymbol{\nu} & \text{on } \Gamma_N, \\ \mathbf{K}\nabla f_0 \cdot \boldsymbol{\nu} + \chi f_0 & \text{on } \Gamma_R, \end{cases} \quad \gamma = \begin{cases} \gamma_D & \text{on } \Gamma_D, \\ \gamma_N & \text{on } \Gamma_N, \\ \gamma_R & \text{on } \Gamma_R, \end{cases}$$

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. In what follows, we make the following assumption.

Assumption 1. $|\Gamma_D| > 0$ so that a Poincaré inequality holds, i.e.,

$$(2.5) \quad \|v\|_{L^2(\Omega)} \leq C_P \|\nabla v\|_{L^2(\Omega)}$$

$\forall v$ vanishing on Γ_D .

In order to estimate the field f_0 , starting from the observations z_1, \dots, z_n and using the a priori knowledge of the phenomenon, we propose minimizing the penalized sum-of-square-error functional

$$(2.6) \quad J(f) = \frac{1}{n} \sum_{i=1}^n (f(\mathbf{p}_i) - z_i)^2 + \lambda \int_{\Omega} (Lf - u)^2$$

over the set of functions $V_{\gamma} = \{v \in H^2(\Omega) : \mathcal{B}_c v = \gamma\}$, where $H^2(\Omega)$ is the Sobolev space of functions in $L^2(\Omega)$ with first and second derivatives in $L^2(\Omega)$; notice that the boundary conditions (2.4) are imposed directly in V_{γ} . Even if in this case we are considering fixed and deterministic boundary conditions, it would be possible to account for uncertainty on the boundary conditions, for instance, by including in the least square functional a dedicated regularizing term, provided that the observations are also available on the boundary.

The functional $J(f)$ is composed of a data fitting criterion, consisting of classical least square errors, and a model fitting criterion, formalized as a roughness term that penalizes the misfit of a PDE governing the phenomenon. Because of the inclusion of a PDE in the definition of the statistical model, the proposed method can be seen as a regularized least square analogous to the Bayesian inverse problems presented, e.g., in [27]. In particular, the least square term in $J(f)$ corresponds to a log-likelihood for Gaussian errors, while the regularizing term effectively translates the prior knowledge of the surface. With respect to [27], besides the different model framework and different estimation approaches, we also deal with a larger class of operators, also including nonstationary anisotropic diffusion, transport, and reaction terms.

The penalization of the general second order operator (2.3) strongly influences the spatial structure of the estimated field. It is thus important to choose correct diffusion, transport, reaction, and boundary terms, which translate the prior knowledge of the phenomenon. An application of this method to a real problem can be found in [1], where the penalized differential operator describes the physiological knowledge of a blood velocity profile. Notice also that by penalizing the misfit of the PDE $Lf_0 - u$, where $u \in L^2(\Omega)$ is the known part of the forcing term, we are actually minimizing the contribution of the unknown forcing term g_0 . This unknown forcing term can be interpreted as the classical control term introduced in PDE optimal control theory; see, for example, [18, 19, 21]. Even though the formalism is similar, in this work we use a different viewpoint with respect to the classical control theory. We consider, in fact, partial, pointwise, and noisy observations, instead of the classical integral targets. For this reason new issues, concerning, for instance, the regularity of the solution, have to be addressed. It is in fact necessary to require $f \in H^2(\Omega)$ for the functional $J(f)$ to be well defined, thanks to the embedding $H^2(\Omega) \subset C(\bar{\Omega})$ if $\Omega \subset \mathbb{R}^d$ with $d \leq 3$. For data in \mathbb{R}^d with $d > 3$ one has to require more regularity in order to obtain $f \in C(\Omega)$; in particular one needs $f \in H^s(\Omega)$ with $s > d/2$; see, e.g., [14].

In the functional $J(f)$ the contribution of the data fitting criterion and of the model fitting criterion is tuned by means of the smoothing parameter λ . A large literature is devoted to the optimal choice of this parameter in classical statistical contexts; see, e.g., [17, 22] and references therein. Standard methods are, for example, the Akaike information criterion (AIC), the Bayesian information criterion (BIC), and the generalized cross-validation (GCV)

criterion. Other methods for the choice of the parameter λ , arising from the inverse problems literature, are the unbiased predictive risk estimator (UPRE), the discrepancy principle, and the L-curve method; see, e.g., [29].

The estimation problem is formulated as follows.

Problem 1. Find $\hat{f} \in V_\gamma$ such that

$$\hat{f} = \operatorname{argmin}_{f \in V_\gamma} J(f).$$

As will be shown in the next section, this problem is well-posed if we assume some regularity on the parameters of the PDE and on the domain Ω . In particular, in the case $d \leq 3$, we make the following assumption.

Assumption 2. The parameters of the PDE are such that $\forall \tilde{u} \in L^2(\Omega)$ there exists a unique solution f_0 of the PDE (2.2), which, moreover, satisfies $f_0 \in H^2(\Omega)$.

The Lax–Milgram theorem guarantees the existence and the uniqueness of the solution of the PDE (2.2) in $H^1(\Omega)$ when the parameters of the PDE \mathbf{K} , \mathbf{b} , and c satisfy some classical requirements, for example, \mathbf{K}_{ij} , \mathbf{b}_j , $c \in L^\infty(\Omega)$, \mathbf{K} is symmetric and uniformly elliptic, i.e., $\boldsymbol{\xi}^T \mathbf{K}(\mathbf{x}) \boldsymbol{\xi} \geq \alpha_K \forall \mathbf{x} \in \Omega$ and $\forall \boldsymbol{\xi} \in \mathbb{R}^d$, $\mathbf{b} \cdot \boldsymbol{\nu} \geq 0$ on $\Gamma_N \cup \Gamma_R$, $-1/2 \operatorname{div}(\mathbf{b}(\mathbf{x})) + c(\mathbf{x}) \geq -\alpha_K/C_P$, where α_K is the ellipticity constant and C_P is the Poincaré constant, $\chi \in L^\infty(\partial\Omega)$, $\chi \geq 0$, and $\gamma_D \in H^{1/2}(\partial\Omega)$, $\gamma_N \in H^{-1/2}(\partial\Omega)$, $\gamma_R \in H^{-1/2}(\partial\Omega)$.

To guarantee that the solution of the PDE is in $H^2(\Omega)$ we need to require extra regularity on the domain Ω , on the parameters of the PDE, and on the boundary conditions: Ω is a polygonal convex domain or its boundary $\partial\Omega$ is a curve of class C^2 , \mathbf{K}_{ij} is Lipschitz continuous, and $\gamma_D \in H^{3/2}(\partial\Omega)$, $\gamma_N \in H^{1/2}(\partial\Omega)$, $\gamma_R \in H^{1/2}(\partial\Omega)$. If the boundary conditions imposed are mixed, they have to satisfy some joint conditions in order not to reduce the regularity of the solution; see [14] for more details.

3. Well-posedness analysis. To analyze the well-posedness of Problem 1 we introduce a new quantity $g \in \mathcal{G} = L^2(\Omega)$ that represents the misfit of the PDE in the penalizing term. This new quantity, $g \in \mathcal{G}$, is defined as $g = Lf - u$, where L is the second order elliptic operator (2.3), and is the classical distributed control term in PDE optimal control theory.

It is useful to also introduce the space $V_0 = \{v \in H^2(\Omega) : \mathcal{B}_c v = 0\}$, which represents the space of functions in $H^2(\Omega)$ with homogeneous boundary conditions, and the operator $B : L^2(\Omega) \rightarrow V_0$ such that $B\tilde{u}$ is the unique solution of the PDE (2.2) with forcing term \tilde{u} and homogeneous boundary conditions, i.e., $L(B\tilde{u}) = \tilde{u}$ in Ω and $\mathcal{B}_c(B\tilde{u}) = 0$ on $\partial\Omega$. Under Assumptions 1 and 2, thanks to the well-posedness and the H^2 -regularity of the PDE (2.2), the operator B is an isomorphism between the spaces L^2 and V_0 and the H^2 -norm of Bu is equivalent to the L^2 -norm of u ; i.e., there exist two positive constants C_1 and C_2 such that

$$(3.1) \quad C_1 \|u\|_{L^2(\Omega)} \leq \|Bu\|_{H^2(\Omega)} \leq C_2 \|u\|_{L^2(\Omega)}.$$

The solution of the PDE (2.2) can thus be written as $f = f_b + B\tilde{u}$, where f_b is the solution of the PDE with the homogeneous forcing term and nonhomogeneous boundary conditions, i.e., $L(f_b) = 0$ in Ω and $\mathcal{B}_c(f_b) = \gamma$ on $\partial\Omega$.

Existence and uniqueness of the estimator \hat{f} are obtained thanks to classical results of calculus of variations. We recall here the result stated, e.g., in [19].

Theorem 1. *If the functional $J(g)$ has the form*

$$(3.2) \quad J(g) = \mathcal{A}(g, g) + \mathcal{L}g + c,$$

where $\mathcal{A} : \mathcal{G} \times \mathcal{G} \rightarrow \mathbb{R}$ is a continuous, coercive, and symmetric bilinear form in \mathcal{G} , $\mathcal{L} : \mathcal{G} \rightarrow \mathbb{R}$ is a linear operator, c is a constant, and \mathcal{G} is a Hilbert space, 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:

$$(3.3) \quad (J'(\hat{g}), \varphi) = 2\mathcal{A}(\hat{g}, \varphi) + \mathcal{L}\varphi = 0 \quad \forall \varphi \in \mathcal{G}.$$

The existence and uniqueness of the estimator are stated in the following theorem.

Theorem 2. *Under Assumptions 1–2, the solution of Problem 1 exists and is unique.*

Proof. Thanks to the definition of g we can write f as an affine transformation of g , i.e., $f = f_b + B(u + g)$, and the functional (2.6) as

$$(3.4) \quad J_g(g) = J(f_b + B(u + g)) = \frac{1}{n} \sum_{i=1}^n (f_b(\mathbf{p}_i) + B(u + g)(\mathbf{p}_i) - z_i)^2 + \lambda \|g\|_{L^2(\Omega)}^2.$$

This reformulation of the functional J is very useful since we can now write J_g in the quadratic form (3.2), where

$$\begin{aligned} \mathcal{A}(g, \varphi) &= \frac{1}{n} \sum_{i=1}^n Bg(\mathbf{p}_i)B\varphi(\mathbf{p}_i) + \lambda \int_{\Omega} g\varphi, \\ \mathcal{L}\varphi &= \frac{2}{n} \sum_{i=1}^n B\varphi(\mathbf{p}_i)(f_b(\mathbf{p}_i) + Bu(\mathbf{p}_i) - z_i), \\ c &= \frac{1}{n} \sum_{i=1}^n (f_b(\mathbf{p}_i) + Bu(\mathbf{p}_i) - z_i)^2. \end{aligned}$$

Clearly, $\mathcal{A}(g, \varphi)$ is a bilinear form since both B and the pointwise evaluation of a function are linear operators. Moreover, it is continuous in \mathcal{G} ; 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 (3.1) we have that

$$|Bg(\mathbf{p}_i)| \leq \|Bg\|_{C(\bar{\Omega})} \leq C \|Bg\|_{H^2(\Omega)} \leq C \|g\|_{L^2(\Omega)}.$$

We thus obtain that $\mathcal{A}(g, \varphi) \leq (C^2 + \lambda) \|g\|_{L^2(\Omega)} \|\varphi\|_{L^2(\Omega)}$.

Finally, the operator $\mathcal{A}(g, \varphi)$ is coercive in $L^2(\Omega)$ since

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

Due to the fact that the bilinear form $\mathcal{A}(\cdot, \cdot)$ is continuous and coercive in $\mathcal{G} = L^2(\Omega)$, that the linear operator \mathcal{L} is continuous, and that c is a constant, Theorem 1 states the existence and the uniqueness of $\hat{g} = \operatorname{argmin}_{g \in \mathcal{G}} J_g(g)$. From the bijectivity of $B : L^2(\Omega) \rightarrow V_0$ we deduce the existence and uniqueness of $\hat{f} = f_b + B(\hat{g} + u) = \operatorname{argmin}_{f \in V_\gamma} J(f)$. ■

The estimator \hat{f} is obtained by solving

$$(3.5) \quad \begin{cases} L\hat{f} = u + \hat{g} & \text{in } \Omega, \\ \mathcal{B}_c \hat{f} = \gamma & \text{on } \partial\Omega, \end{cases}$$

where \hat{g} is the solution of

$$(3.6) \quad \frac{1}{2} (J'_g(\hat{g}), \varphi) = \int_{\Omega} \frac{1}{n} \sum_{i=1}^n (\hat{f} - z_i) B\varphi \delta_{\mathbf{p}_i} + \lambda \int_{\Omega} \hat{g}\varphi = 0 \quad \forall \varphi \in \mathcal{G}.$$

Recalling that B is an isomorphism between L^2 and V_0 , we can associate to any $\varphi \in \mathcal{G}$ the function $v \in V_0$ such that $v = B\varphi$ (or, equivalently, $Lv = \varphi$ with homogeneous boundary conditions) and we can write the latter equation as

$$(3.7) \quad \frac{1}{2} (J'_g(\hat{g}), v) = \int_{\Omega} \frac{1}{n} \sum_{i=1}^n (\hat{f} - z_i) v \delta_{\mathbf{p}_i} + \lambda \int_{\Omega} \hat{g}Lv = 0 \quad \forall v \in V_0.$$

The estimator \hat{g} , the solution of (3.7), can be interpreted as the solution in the sense of distributions of the PDE

$$(3.8) \quad L^* \hat{g} = -\frac{1}{n\lambda} \sum_{i=1}^n (\hat{f} - z_i) \delta_{\mathbf{p}_i},$$

where $\delta_{\mathbf{p}_i}$ is the Dirac mass located in \mathbf{p}_i and L^* is the adjoint operator of L :

$$(3.9) \quad L^*g = -\operatorname{div}(\mathbf{K}\nabla g) - \mathbf{b} \cdot \nabla g + (c - \operatorname{div}(\mathbf{b}))g.$$

Remark 1. Should \hat{g} be sufficiently smooth, we could integrate twice (3.7) by parts and obtain \hat{g} as the solution of

$$(3.10) \quad \begin{cases} L^* \hat{g} = -\frac{1}{n\lambda} \sum_{i=1}^n (\hat{f} - z_i) \delta_{\mathbf{p}_i} & \text{in } \Omega, \\ \mathcal{B}_c^* \hat{g} = 0 & \text{on } \partial\Omega, \end{cases}$$

where the boundary conditions associated to the adjoint problem are

$$(3.11) \quad \mathcal{B}_c^* g = \begin{cases} g & \text{on } \Gamma_D, \\ \mathbf{K}\nabla g \cdot \boldsymbol{\nu} + \mathbf{b} \cdot \boldsymbol{\nu} g & \text{on } \Gamma_N, \\ \mathbf{K}\nabla g \cdot \boldsymbol{\nu} + (\mathbf{b} \cdot \boldsymbol{\nu} + \chi)g & \text{on } \Gamma_R. \end{cases}$$

The estimator (\hat{f}, \hat{g}) would then satisfy the coupled problem

$$(3.12) \quad \begin{cases} L\hat{f} = u + \hat{g} & \text{in } \Omega, \\ \mathcal{B}_c \hat{f} = \gamma & \text{on } \partial\Omega, \end{cases} \quad \begin{cases} L^* \hat{g} = -\frac{1}{n\lambda} \sum_{i=1}^n (\hat{f} - z_i) \delta_{\mathbf{p}_i} & \text{in } \Omega, \\ \mathcal{B}_c^* \hat{g} = 0 & \text{on } \partial\Omega, \end{cases}$$

which is analogous to the mixed formulation of fourth order problems, described, for example, in [4, 5, 20]. Usually, \hat{g} is not regular enough for the strong formulation (3.10) to be well defined. However, we will derive the finite element discretization, described in detail in section 5, starting from the weak formulation of the system (3.12).

4. Bias of the estimator. The penalty term in the functional $J(f)$ induces a bias in the estimator \hat{f} unless the unknown part of the forcing term g_0 is equal to the null function and the true underlying field f_0 satisfies exactly the penalized PDE, namely $Lf_0 = u$. Now we want to quantify this bias.

The estimators \hat{f} and \hat{g} are defined, respectively, as the solutions of the linear problems (3.5) and (3.7). Thanks to the linearity of (3.7), we can write $\hat{g} = \hat{g}^* + \hat{g}_w$, where \hat{g}^* and \hat{g}_w are, respectively, the solution of the equations

$$(4.1) \quad \int_{\Omega} \frac{1}{n} \sum_{i=1}^n (B\hat{g}^* + Bu + f_b - f_0) v \delta_{\mathbf{p}_i} + \lambda \int_{\Omega} \hat{g}^* Lv = 0 \quad \forall v \in V_0,$$

$$(4.2) \quad \int_{\Omega} \frac{1}{n} \sum_{i=1}^n (B\hat{g}_w - \epsilon_i) v \delta_{\mathbf{p}_i} + \lambda \int_{\Omega} \hat{g}_w Lv = 0 \quad \forall v \in V_0.$$

Indeed it is trivial to show that $\hat{g}^* + \hat{g}_w$ satisfies (3.7) since $z_i = f_0(\mathbf{p}_i) + \epsilon_i$ and $\hat{f} = f_b + B(u + \hat{g})$.

Thanks to the linearity of (3.5), we can now write the surface estimator as $\hat{f} = \hat{f}^* + \hat{w}$, where $\hat{f}^* = f_b + B(u + \hat{g}^*)$ and $\hat{w} = B\hat{g}_w$. It is easy to see that \hat{f}^* and \hat{w} both satisfy a minimization problem of the form (2.6). It follows that the estimation problem can be split in two independent parts:

$$(4.3) \quad \hat{f} = \underset{f \in V_{\gamma}}{\operatorname{argmin}} \left[\frac{1}{n} \sum_{i=1}^n (f(\mathbf{p}_i) - f_0(\mathbf{p}_i))^2 + \lambda \int_{\Omega} (Lf - u)^2 \right] \\ + \underset{w \in V_0}{\operatorname{argmin}} \left[\frac{1}{n} \sum_{i=1}^n (w(\mathbf{p}_i) - \epsilon_i)^2 + \lambda \int_{\Omega} (Lw)^2 \right].$$

The first functional on the right-hand side of (4.3) involves the observations without noise and the nonhomogeneous PDE in the penalty term (both the forcing term and the boundary conditions are nonhomogeneous). The minimizer \hat{f}^* of this functional is deterministic. The second functional on the right-hand side of (4.3) involves instead pure noise data and the homogeneous PDE (both the forcing term and the boundary conditions are homogeneous). The minimizer \hat{w} of this functional is thus a random term. Notice that the random term has zero mean, i.e., $\mathbb{E}[\hat{w}] = 0$. Indeed, from (4.2) it follows that

$$\int_{\Omega} \frac{1}{n} \sum_{i=1}^n B\mathbb{E}[\hat{g}_w] v \delta_{\mathbf{p}_i} + \lambda \int_{\Omega} \mathbb{E}[\hat{g}_w] Lv = \frac{1}{n} \sum_{i=1}^n \mathbb{E}[\epsilon_i] v \delta_{\mathbf{p}_i} = 0 \quad \forall v \in V_0.$$

Exploiting the arbitrariness of $v \in V_0$, we obtain $\mathbb{E}[\hat{g}_w] = 0$ and $\mathbb{E}[\hat{w}] = B\mathbb{E}[\hat{g}_w] = 0$. For this reason the deterministic term \hat{f}^* corresponds to the average of the surface estimator, i.e., $\mathbb{E}[\hat{f}] = \hat{f}^*$.

We are now interested in quantifying the bias of the surface estimator $\mathbb{E}[\hat{f}] - f_0$ induced by the penalizing term. In particular it is natural to study the bias in the norm on V_0 induced by the functional $J(f)$, i.e.,

$$(4.4) \quad \|f\|_J^2 = \frac{1}{n} \sum_{i=1}^n f(\mathbf{p}_i)^2 + \lambda \int_{\Omega} (Lf)^2 \quad \forall f \in V_0.$$

Lemma 1. *The norm (4.4) of the bias of \hat{f} is bounded by*

$$(4.5) \quad \left\| \mathbb{E}[\hat{f}] - f_0 \right\|_J^2 \leq 4\lambda \|Lf_0 - u\|_{L^2(\Omega)}^2.$$

Proof. In order to obtain the inequality (4.5) we can use the optimality of $\mathbb{E}[\hat{f}]$ in the minimization of the first functional on the right-hand side of (4.3), with respect to any other function in V_γ . We have in fact that

$$\begin{aligned} \left\| \mathbb{E}[\hat{f}] - f_0 \right\|_J^2 &= \frac{1}{n} \sum_{i=1}^n (\mathbb{E}[\hat{f}](\mathbf{p}_i) - f_0(\mathbf{p}_i))^2 + \lambda \left\| L(\mathbb{E}[\hat{f}] - f_0) \right\|_{L^2(\Omega)}^2 \\ &\leq \frac{1}{n} \sum_{i=1}^n (\mathbb{E}[\hat{f}](\mathbf{p}_i) - f_0(\mathbf{p}_i))^2 + 2\lambda \left\| L\mathbb{E}[\hat{f}] - u \right\|_{L^2(\Omega)}^2 + 2\lambda \|Lf_0 - u\|_{L^2(\Omega)}^2 \\ &\leq 2 \left[\frac{1}{n} \sum_{i=1}^n (\mathbb{E}[\hat{f}](\mathbf{p}_i) - f_0(\mathbf{p}_i))^2 + \lambda \left\| L\mathbb{E}[\hat{f}] - u \right\|_{L^2(\Omega)}^2 \right] + 2\lambda \|Lf_0 - u\|_{L^2(\Omega)}^2 \\ &\leq 2\lambda \|Lf_0 - u\|_{L^2(\Omega)}^2 + 2\lambda \|Lf_0 - u\|_{L^2(\Omega)}^2. \quad \blacksquare \end{aligned}$$

This result means that the estimator is asymptotically unbiased in the norm $\|\cdot\|_J$ if either $\|Lf_0 - u\|_{L^2} = 0$ or $\lambda = \lambda(n) \rightarrow 0$ for $n \rightarrow +\infty$. The condition $\|Lf_0 - u\|_{L^2(\Omega)} = 0$ means that the real field f_0 is in the kernel of the penalty term, while the condition $\lambda = \lambda(n) \rightarrow 0$ for $n \rightarrow +\infty$ means that we consider the roughness parameter to be a decreasing function of the number of observations and in particular that the more observations we have, the less we penalize the PDE misfit.

5. Finite element estimator. The estimation problem presented in section 2 is infinite dimensional and cannot be solved analytically. To reduce this infinite dimensional problem to a finite dimensional one, we discretize the PDE system (3.12) with the finite element method; this method has already been used in this framework, for example, in [23, 25, 13]. The finite element approximation of the system (3.12) can be regarded as a naive mixed finite element method for the discretization of Problem 1. More complex methods for the discretization of fourth order problems could be used: in [5], for example, some conforming and nonconforming methods for the discretization of fourth order problems are introduced, while in [3, 28, 15] more recent discontinuous Galerkin methods are described.

Let \mathcal{T}_h be a regular and quasi-uniform triangulation of the domain, which for convenience we assume here to be polygonal and convex, and $h = \max_{K \in \mathcal{T}_h} \text{diam}(K)$ be the characteristic mesh size (see, e.g., [2]). Notice that the mesh \mathcal{T}_h can be defined independently of the location of the observations $\mathbf{p}_1, \dots, \mathbf{p}_n$. We consider the space V_h^r of piecewise continuous polynomial functions of degree $r \geq 1$ on the triangulation:

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

and $V_{h,\Gamma_D}^r = V_h^r \cap H_{\Gamma_D}^1(\Omega)$, where $H_{\Gamma_D}^1(\Omega) = \{v \in H^1(\Omega) : v|_{\Gamma_D} = 0\}$.

In order to discretize the PDE system (3.12) we define the bilinear forms

$$(5.1) \quad \begin{aligned} r(g, v) &= \int_{\Omega} gv, & l(f, \psi) &= \frac{1}{n} \sum_{i=1}^n f(\mathbf{p}_i) \psi(\mathbf{p}_i), \\ a(f, \psi) &= \int_{\Omega} (\mathbf{K} \nabla f \cdot \nabla \psi + \mathbf{b} \cdot \nabla f \psi + cf\psi) + \int_{\Gamma_R} \chi f \psi, \end{aligned}$$

the last one being the bilinear form associated to the operator L ; we also introduce the linear operator $F(\psi) = \int_{\Omega} u\psi + \int_{\Gamma_N} \gamma_N \psi + \int_{\Gamma_R} \gamma_R \psi$.

Now let $f_{D,h} \in V_h^r$ be a lifting of the nonhomogeneous Dirichlet conditions, i.e., $f_{D,h}|_{\Gamma_D} = \gamma_{D,h}$, where $\gamma_{D,h}$ is the interpolant of γ_D in the space of piecewise continuous polynomial functions of degree r on the Dirichlet boundary Γ_D . The finite element approximation of the system (3.12) becomes

$$(5.2) \quad \begin{cases} \frac{1}{\lambda} l(\hat{f}_h, \psi_h) + a(\psi_h, \hat{g}_h) = \frac{1}{n\lambda} \sum_{i=1}^n z_i \psi_h(\mathbf{p}_i) & \forall \psi_h \in V_{h,\Gamma_D}^r, \\ a(\hat{f}_h, v_h) - r(\hat{g}_h, v_h) = F(v_h) & \forall v_h \in V_{h,\Gamma_D}^r, \end{cases}$$

with $(\hat{f}_h - f_{D,h}, \hat{g}_h) \in V_{h,\Gamma_D}^r \times V_{h,\Gamma_D}^r$.

In this and the following section, we need a slightly stronger regularity assumption on the PDE (2.2); in particular, we need its solution to be in a Sobolev space $W^{2,p}$, where $W^{s,p}(\Omega)$ is the space of functions in $L^p(\Omega)$ with derivatives up to order s in $L^p(\Omega)$.

Assumption 3. The parameters of the PDE are such that for every $\tilde{u} \in L^p(\Omega)$ there exists a unique solution $f_0 \in W^{2,p}(\Omega)$ for some $p > d$.

Lemma 2. Under Assumption 3, there exists $h_0 > 0$ such that for every $h \leq h_0$, problem (5.2) has a unique solution.

Proof. The proof mimics the strategy used to prove the existence and the uniqueness of the estimator at the continuous level in Theorem 2.

Let $\varphi_h \in V_h^r$, and let $B : L^2(\Omega) \rightarrow V_0$ be the operator defined in section 3 such that $\psi = B\varphi_h$ is the solution of

$$a(\psi, v) = \int_{\Omega} \varphi_h v \quad \forall v \in H_{\Gamma_D}^1.$$

We define the operator B_h as the discretization of the operator B ; i.e., $\psi_h = B_h \varphi_h \in V_{h,\Gamma_D}^r$ is the solution of

$$(5.3) \quad a(\psi_h, v_h) = \int_{\Omega} \varphi_h v_h \quad \forall v_h \in V_{h,\Gamma_D}^r.$$

It is easy to show that the operator B_h is stable in the L^∞ -norm, i.e.,

$$\|\psi_h\|_{L^\infty(\Omega)} \leq C \|\varphi_h\|_{L^2(\Omega)}.$$

We have in fact that

$$\|\psi_h\|_{L^\infty(\Omega)} \leq \|\psi - \psi_h\|_{L^\infty(\Omega)} + \|\psi\|_{L^\infty(\Omega)}.$$

Thanks to the H^2 -elliptic regularity of the PDE (2.2) (see Assumption 2) we have that

$$\|\psi\|_{L^\infty(\Omega)} \leq C \|\psi\|_{H^2(\Omega)} \leq C \|\varphi_h\|_{L^2(\Omega)},$$

while thanks to Assumption 3 and the Sobolev inequality (see, e.g., [2])

$$(5.4) \quad \|w\|_{L^\infty(\Omega)} \leq C \|w\|_{W^{1,p}(\Omega)} \quad \forall w \in W^{1,p}(\Omega), \quad \forall p > d,$$

where $W^{1,p}(\Omega)$ is the space of functions in $L^p(\Omega)$ with first derivatives in $L^p(\Omega)$, we obtain the bound for the error term in the L^∞ -norm

$$\begin{aligned} \|\psi - \psi_h\|_{L^\infty(\Omega)} &\leq C \|\psi - \psi_h\|_{W^{1,p}(\Omega)} \leq C \inf_{v_h \in V_{h,\Gamma_D}^r} \|\psi - v_h\|_{W^{1,p}(\Omega)} \\ &\leq Ch \|\psi\|_{W^{2,p}(\Omega)} \leq Ch \|\varphi_h\|_{L^p(\Omega)} \leq Ch^{1+\min\{0, \frac{d}{p} - \frac{d}{2}\}} \|\varphi_h\|_{L^2(\Omega)}. \end{aligned}$$

In the last step we have used an inverse inequality (see, e.g., [12]), and taking $p = 2d/(d-2)$, which is larger than d for $d \leq 3$, we conclude that

$$\|\psi - \psi_h\|_{L^\infty(\Omega)} \leq C \|\varphi_h\|_{L^2(\Omega)}.$$

We now define the operators \mathcal{A}_h and \mathcal{L}_h as the discretization of the operators \mathcal{A} and \mathcal{L} defined in section 3:

$$\begin{aligned} \mathcal{A}_h(g_h, \varphi_h) &= \frac{1}{n} \sum_{i=1}^n B_h g_h(\mathbf{p}_i) B_h \varphi_h(\mathbf{p}_i) + \lambda \int_{\Omega} g_h \varphi_h, \\ \mathcal{L}_h \varphi_h &= \frac{2}{n} \sum_{i=1}^n B_h \varphi_h(\mathbf{p}_i) (f_{b,h}(\mathbf{p}_i) + B_h u(\mathbf{p}_i) - z_i), \end{aligned}$$

where $f_{b,h}$ is the discretization of f_b . The operator \mathcal{A}_h is coercive in L^2 ; in fact,

$$\mathcal{A}_h(g_h, g_h) = \frac{1}{n} \sum_{i=1}^n (B_h g_h(\mathbf{p}_i))^2 + \lambda \int_{\Omega} g_h^2 \geq \lambda \|g_h\|_{L^2(\Omega)}^2.$$

Thanks to the stability in the L^∞ -norm of the operator B_h , the operators \mathcal{A}_h and \mathcal{L}_h are both continuous:

$$\begin{aligned} \mathcal{A}_h(g_h, \varphi_h) &\leq C \|B_h g_h\|_{L^\infty(\Omega)} \|B_h \varphi_h\|_{L^\infty(\Omega)} + \lambda \|g_h\|_{L^2(\Omega)} \|\varphi_h\|_{L^2(\Omega)} \\ &\leq C \|g_h\|_{L^2(\Omega)} \|\varphi_h\|_{L^2(\Omega)}, \\ \mathcal{L}_h(\varphi_h) &\leq C \|B_h \varphi_h\|_{L^\infty(\Omega)} \leq C \|\varphi_h\|_{L^2(\Omega)}. \end{aligned}$$

Thanks to the fact that \mathcal{A}_h is continuous and coercive in $L^2(\Omega)$ and \mathcal{L}_h is continuous in $L^2(\Omega)$, the equation

$$2\mathcal{A}_h(g_h, \varphi_h) + \mathcal{L}_h(\varphi_h) = 0 \quad \forall \varphi_h : B_h \varphi_h = \psi_h \in V_{h,\Gamma_D}^r$$

has a unique solution $g_h \in V_{h,\Gamma_D}^r$. This equation corresponds to the first equation of the system (5.2).

Once \hat{g}_h is known, \hat{f}_h is recovered uniquely from the second equation in (5.2). ■

Now let $\{\psi_k\}_{k=1}^{N_h}$ be the Lagrangian basis of the space V_{h,Γ_D}^r , where $N_h = \dim(V_{h,\Gamma_D}^r)$, and let $\boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_{N_h}$ be the nodes associated to the N_h basis functions. Thanks to the Lagrangian property of the basis functions we can write a function $f \in \text{span}\{\psi_1, \dots, \psi_{N_h}\}$ as

$$f(\mathbf{x}) = \sum_{k=1}^{N_h} f(\boldsymbol{\xi}_k) \psi_k(\mathbf{x}) = \mathbf{f}^T \boldsymbol{\psi},$$

where $\mathbf{f} = (f_1, \dots, f_{N_h})^T = (f(\boldsymbol{\xi}_1), \dots, f(\boldsymbol{\xi}_{N_h}))^T$ and $\boldsymbol{\psi} = (\psi_1, \dots, \psi_{N_h})^T$.

Analogously, we define the Lagrangian basis of the space $V_h^r \setminus V_{h,\Gamma_D}^r$ as $\{\psi_k^D\}_{k=1}^{N_h^D}$, where $N_h^D = \dim(V_h^r \setminus V_{h,\Gamma_D}^r)$, and the nodes on the boundary Γ_D as $\boldsymbol{\xi}_1^D, \dots, \boldsymbol{\xi}_{N_h^D}^D$. A lifting $f_{D,h}$ can be constructed in $\text{span}\{\psi_1^D, \dots, \psi_{N_h^D}^D\}$ as $f_{D,h} = \mathbf{f}_D^T \boldsymbol{\psi}^D$, where $\mathbf{f}_D = (f_D(\boldsymbol{\xi}_1^D), \dots, f_D(\boldsymbol{\xi}_{N_h^D}^D))^T$ and $\boldsymbol{\psi}^D = (\psi_1^D, \dots, \psi_{N_h^D}^D)^T$.

We now define the finite element matrices and vectors: $\mathbf{R}_{jk} = \int_{\Omega} \psi_j \psi_k$ is the mass matrix, $\boldsymbol{\Psi}_{ij} = \psi_j(\mathbf{p}_i)$ and $\boldsymbol{\Psi}_{ij}^D = \psi_j^D(\mathbf{p}_i)$ are the matrices of pointwise evaluation of the basis functions, $\mathbf{A}_{jk} = a(\psi_k, \psi_j)$ and $\mathbf{A}_{jk}^D = a(\psi_k^D, \psi_j)$ are the matrices associated to the bilinear form $a(\cdot, \cdot)$, $\mathbf{z} = (z_1, \dots, z_n)$ is the vector containing the observed data, and $\mathbf{u}_j = \int_{\Omega} u \psi_j$, $(\mathbf{h}_N)_j = \int_{\Gamma_N} \gamma_N \psi_j$, and $(\mathbf{h}_R)_j = \int_{\Gamma_R} \gamma_R \psi_j$ are the vectors related to the forcing term and the nonhomogeneous boundary conditions.

The finite element solution \hat{f}_h of the discrete counterpart of the estimation problem can thus be written as

$$\hat{f}_h = \hat{\mathbf{f}}^T \boldsymbol{\psi} + \mathbf{f}_D^T \boldsymbol{\psi}^D,$$

where $\hat{\mathbf{f}}$ is the solution of the linear system

$$(5.5) \quad \begin{bmatrix} \boldsymbol{\Psi}^T \boldsymbol{\Psi} / (n\lambda) & \mathbf{A}^T \\ \mathbf{A} & -\mathbf{R} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{f}} \\ \hat{\mathbf{g}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Psi}^T \mathbf{z} / (n\lambda) - \boldsymbol{\Psi}^T \boldsymbol{\Psi}^D \mathbf{f}_D / (n\lambda) \\ \mathbf{u} + \mathbf{h}_N + \mathbf{h}_R - \mathbf{A}^D \mathbf{f}_D \end{bmatrix}.$$

In [1], exploiting the linearity of the estimator \hat{f}_h in the observations, some classical inferential tools such as approximate pointwise confidence bands and prediction intervals are derived, providing measures for uncertainty quantification within such models. Moreover, the smoothing parameter λ may be selected via standard methods as the AIC, the BIC, the UPRE, and the GCV criterion; see, e.g., [17, 22, 29] and references therein. In particular the GCV, which is available in a closed form for the proposed models, is detailed in [1].

6. Bias of the finite element estimator. To study the convergence of the finite element estimator, it is convenient to reformulate the estimation problem (Problem 1) as a constrained problem on the space

$$(6.1) \quad \mathcal{W} = \{(f, g) \in V_{\gamma} \times \mathcal{G} : Lf - u = g\}.$$

Thus, exploiting the introduction of the adjoint variable \hat{g} , the estimation problem becomes the following.

Problem 2. Find $\hat{f} \in V_{\gamma}$, $\hat{g} \in \mathcal{G}$ such that

$$(\hat{f}, \hat{g}) = \underset{(f,g) \in \mathcal{W}}{\operatorname{argmin}} \frac{1}{n} \sum_{i=1}^n (f(\mathbf{p}_i) - z_i)^2 + \lambda \int g^2.$$

The constrained space \mathcal{W} can be discretized as

$$\mathcal{W}_h = \{(f_h, g_h) \in V_h^r \times V_{h,\Gamma_D}^r : f_h|_{\Gamma_D} = \gamma_{D,h} \text{ and} \\ a(f_h, v_h) - r(g_h, v_h) = F(v_h) \ \forall v_h \in V_{h,\Gamma_D}^r\},$$

where $a(\cdot, \cdot)$, $r(\cdot, \cdot)$, $F(\cdot)$, and $\gamma_{D,h}$ are as defined in section 5, and the finite element estimator $(\hat{f}_h, \hat{g}_h) \in \mathcal{W}_h$ is the solution of the equation

$$\frac{1}{\lambda} l(\hat{f}_h, \psi_h) + a(\psi_h, \hat{g}_h) = \frac{1}{n\lambda} \sum_{i=1}^n z_i \psi_h(\mathbf{p}_i) \quad \forall \psi_h \in V_{h,\Gamma_D}^r.$$

The finite element estimator \hat{f}_h can be split, like its continuous counterpart \hat{f} , in two different terms $\mathbb{E}[\hat{f}_h]$ and \hat{w}_h that are, respectively, the finite element approximation of $\mathbb{E}[\hat{f}]$ and \hat{w} . Reasoning as for the continuous problem, we can easily show that $\mathbb{E}[\hat{w}_h] = 0$ and that the expected value of the finite element estimator $(\mathbb{E}[\hat{f}_h], \mathbb{E}[\hat{g}_h])$ is the solution of the equation

$$(6.2) \quad l(\mathbb{E}[\hat{f}_h], \psi_h) + \lambda a(\psi_h, \mathbb{E}[\hat{g}_h]) = l(f_0, \psi_h) \quad \forall \psi_h \in V_{h,\Gamma_D}^r$$

in the constrained space \mathcal{W}_h .

Neglecting the zero mean term, we now aim at studying the bias of the finite element estimator, $\mathbb{E}[\hat{f}_h] - f_0$, in the norm defined on the constrained space \mathcal{W} ,

$$(6.3) \quad \|\hat{f}_h\|^2 = \|\hat{f}_h\|_n^2 + \lambda \left[\|\hat{f}_h\|_{H^1(\Omega)}^2 + \|\hat{g}_h\|_{L^2(\Omega)}^2 \right],$$

where the seminorm $\|\cdot\|_n$ is the seminorm induced by the bilinear operator $l(\cdot, \cdot)$, i.e.,

$$\|\hat{f}_h\|_n = \frac{1}{n} \sum_{i=1}^n (\hat{f}_h(\mathbf{p}_i))^2.$$

Notice that the norm $\|\cdot\|$ contains both the norm $\|\cdot\|_J$ and the H^1 -norm of \hat{f}_h . In fact, we also need an explicit control on the H^1 -norm of \hat{f}_h to study the convergence properties of the mixed finite element solution of the system (5.2).

The convergence of the bias term is studied when $h \rightarrow 0$, while the number n of observations and the penalty parameter λ are considered fixed. Since we are considering n and λ fixed, we expect to obtain an error bound that contains a term going to zero as $h \rightarrow 0$ and a term that represents the bias induced by the roughness penalty, similarly to the continuous setting in Lemma 1.

Remark 2. One might be tempted to compare $\mathbb{E}[\hat{f}_h]$ to its continuous counterpart $\mathbb{E}[\hat{f}]$. However, due to the presence of $\delta_{\mathbf{p}_i}$ in the forcing term of the dual equation in (3.12), $\mathbb{E}[\hat{f}]$ is not smooth in general. For this reason, in the error analysis proposed in this section, we directly compare $\mathbb{E}[\hat{f}_h]$ with the true underlying field f_0 , which is assumed to be sufficiently smooth.

The bound for the bias of the finite element estimator $\mathbb{E}[\hat{f}_h] - f_0$ is obtained thanks to the following lemma and theorem.

Lemma 3. *Let $g_0 = Lf_0 - u$. The bias of the finite element estimator $(\mathbb{E}[\hat{f}_h], \mathbb{E}[\hat{g}_h]) \in \mathcal{W}_h$ satisfies the inequality*

$$\begin{aligned}
 & \left\| f_0 - \mathbb{E}[\hat{f}_h] \right\|_n^2 + \lambda \left[\left\| f_0 - \mathbb{E}[\hat{f}_h] \right\|_{H^1(\Omega)}^2 + \|g_0 - \mathbb{E}[\hat{g}_h]\|_{L^2(\Omega)}^2 \right] \\
 & \leq C \left\{ \inf_{(\varphi_h, p_h) \in \mathcal{W}_h} \left[\|f_0 - \varphi_h\|_n^2 + \lambda \|f_0 - \varphi_h\|_{H^1(\Omega)}^2 + \lambda \|g_0 - p_h\|_{L^2(\Omega)}^2 \right] \right. \\
 (6.4) \quad & \left. + \lambda \|g_0\|_{L^2(\Omega)}^2 \right\}
 \end{aligned}$$

for some constant $C > 0$ independent of h .

Proof. We set $f_h^* = \mathbb{E}[\hat{f}_h]$ and $g_h^* = \mathbb{E}[\hat{g}_h]$, and we recall that $\|\cdot\|_n$ is the seminorm induced by the bilinear form $l(\cdot, \cdot)$, i.e., $\|f\|_n^2 = l(f, f)$.

In order to prove Lemma 3 we can use the theory of saddle point systems. From (6.2) and the definition of \mathcal{W}_h we have immediately

$$\begin{aligned}
 \frac{1}{\lambda} l(\hat{f}_h^* - f_0, \psi_h) + a(\psi_h, \hat{g}_h^*) &= 0 \quad \forall \psi_h \in V_{h, \Gamma_D}^r, \\
 a(\hat{f}_h^* - \varphi_h, v_h) &= r(\hat{g}_h^* - p_h, v_h) \quad \forall v_h \in V_{h, \Gamma_D}^r, (\varphi_h, p_h) \in \mathcal{W}_h.
 \end{aligned}$$

Choosing $(\varphi_h, p_h) \in \mathcal{W}_h$ we thus obtain

$$\begin{aligned}
 & \left\| \hat{f}_h^* - \varphi_h \right\|_n^2 + \lambda \|\hat{g}_h^* - p_h\|_{L^2(\Omega)}^2 = l(\hat{f}_h^* - \varphi_h, \hat{f}_h^* - \varphi_h) + \lambda r(\hat{g}_h^* - p_h, \hat{g}_h^* - p_h) \\
 & = l(\hat{f}_h^* - f_0, \hat{f}_h^* - \varphi_h) + l(f_0 - \varphi_h, \hat{f}_h^* - \varphi_h) + \lambda r(\hat{g}_h^* - p_h, \hat{g}_h^* - p_h) \\
 & = -\lambda a(\hat{f}_h^* - \varphi_h, \hat{g}_h^*) + l(f_0 - \varphi_h, \hat{f}_h^* - \varphi_h) + \lambda r(\hat{g}_h^* - p_h, \hat{g}_h^* - p_h) \\
 (6.5) \quad & = l(f_0 - \varphi_h, \hat{f}_h^* - \varphi_h) - \lambda r(\hat{g}_h^* - p_h, p_h)
 \end{aligned}$$

since $\hat{f}_h^* - \varphi_h \in V_{h, \Gamma_D}^r$. We now bound the term $\hat{f}_h^* - \varphi_h$ in the H^1 -norm using the coercivity of the bilinear form $a(\cdot, \cdot)$ in $H^1(\Omega)$:

$$\begin{aligned}
 \left\| \hat{f}_h^* - \varphi_h \right\|_{H^1(\Omega)}^2 &\leq \frac{1}{\alpha} a(\hat{f}_h^* - \varphi_h, \hat{f}_h^* - \varphi_h) = \frac{1}{\alpha} r(\hat{g}_h^* - p_h, \hat{f}_h^* - \varphi_h) \\
 &\leq \frac{\sqrt{1 + C_P^2}}{\alpha} \|\hat{g}_h^* - p_h\|_{L^2(\Omega)} \left\| \hat{f}_h^* - \varphi_h \right\|_{H^1(\Omega)},
 \end{aligned}$$

where α is the coercivity constant and C_P is the constant in the Poincaré inequality (2.5),

which holds thanks to Assumption 1. Summing this inequality to (6.5) we obtain

$$\begin{aligned} & \left\| \hat{f}_h^* - \varphi_h \right\|_n^2 + \lambda \left[\frac{\alpha^2}{4(1+C_P^2)} \left\| \hat{f}_h^* - \varphi_h \right\|_{H^1(\Omega)}^2 + \|\hat{g}_h^* - p_h\|_{L^2(\Omega)} \right] \\ & \leq l(f_0 - \varphi_h, \hat{f}_h^* - \varphi_h) - \lambda r(\hat{g}_h^* - p_h, p_h) + \frac{\lambda}{4} \|\hat{g}_h^* - p_h\|_{L^2(\Omega)}^2 \\ & \leq \frac{1}{2} \|f_0 - \varphi_h\|_n^2 + \frac{1}{2} \left\| \hat{f}_h^* - \varphi_h \right\|_n^2 + \frac{\lambda}{2} \|\hat{g}_h^* - p_h\|_{L^2(\Omega)}^2 + 2\lambda \|g_0 - p_h\|_{L^2(\Omega)}^2 \\ & \quad + 2\lambda \|g_0\|_{L^2(\Omega)}^2. \end{aligned}$$

This inequality provides the bound

$$\begin{aligned} & \left\| \hat{f}_h^* - \varphi_h \right\|_n^2 + \lambda \left[\left\| \hat{f}_h^* - \varphi_h \right\|_{H^1(\Omega)}^2 + \|\hat{g}_h^* - p_h\|_{L^2(\Omega)}^2 \right] \\ & \leq C \left\{ \|f_0 - \varphi_h\|_n^2 + \lambda \|g_0 - p_h\|_{L^2(\Omega)}^2 + \lambda \|g_0\|_{L^2(\Omega)}^2 \right\}. \end{aligned}$$

The final error bound (6.4) can now be obtained by triangular inequality and exploiting the arbitrariness of $(\varphi_h, p_h) \in \mathcal{W}_h$. ■

To obtain the desired result, it is necessary to split the error term on the constrained space \mathcal{W}_h in two different errors for $\mathbb{E}[\hat{f}_h]$ and $\mathbb{E}[\hat{g}_h]$ on the space V_h^r . Assuming, moreover, that f_0 and g_0 are in proper Sobolev spaces $W^{s,p}(\Omega)$ we obtain the following result.

Theorem 3. *Under Assumption 3, using finite elements of degree r , if $f_0 \in W^{r+1,p}(\Omega)$ with $f_0|_{\Gamma_D} = \gamma_D$, for $p > d$, and $g_0 \in H^r(\Omega)$ with $g_0|_{\Gamma_D} = 0$, there exists $h_0 > 0$ such that $\forall h \leq h_0$*

$$(6.6) \quad \begin{aligned} & \left\| f_0 - \mathbb{E}[\hat{f}_h] \right\|_n^2 + \lambda \left[\left\| f_0 - \mathbb{E}[\hat{f}_h] \right\|_{H^1(\Omega)}^2 + \|g_0 - \mathbb{E}[\hat{g}_h]\|_{L^2(\Omega)}^2 \right] \\ & \leq C \left\{ h^{2r} \left[(1 + \lambda) |f_0|_{W^{r+1,p}(\Omega)}^2 + |g_0|_{H^r(\Omega)}^2 \right] + \lambda \|g_0\|_{L^2(\Omega)}^2 \right\}. \end{aligned}$$

Proof. In order to prove the result we need to split in two parts the constrained error term

$$\inf_{(\varphi_h, p_h) \in \mathcal{W}_h} \left[\|f_0 - \varphi_h\|_n^2 + \lambda \|f_0 - \varphi_h\|_{H^1(\Omega)}^2 + \lambda \|g_0 - p_h\|_{L^2(\Omega)}^2 \right]$$

in inequality (6.4).

In the following we fix $p_h \in V_{h,\Gamma_D}^r$ and we choose $\varphi_h \in V_h^r$, which satisfies $a(\varphi_h, v_h) = r(p_h, v_h) + F(v_h)$ and $\varphi_h|_{\Gamma_D} = \gamma_{D,h}$, so that $(\varphi_h, p_h) \in \mathcal{W}_h$. Thanks to this choice we obtain the following bound:

$$(6.7) \quad \|f_0 - \varphi_h\|_{H^1(\Omega)}^2 \leq C \left[\|f_0 - z_h\|_{H^1(\Omega)}^2 + \|g_0 - p_h\|_{L^2(\Omega)}^2 \right],$$

where z_h is an arbitrary function in V_h^r such that $z_h|_{\Gamma_D} = \gamma_{D,h}$. This inequality is obtained thanks to the fact that

$$\|f_0 - \varphi_h\|_{H^1(\Omega)}^2 \leq 2 \|f_0 - z_h\|_{H^1(\Omega)}^2 + 2 \|z_h - \varphi_h\|_{H^1(\Omega)}^2$$

and that

$$\begin{aligned} \alpha \|z_h - \varphi_h\|_{H^1(\Omega)}^2 &\leq a(z_h - \varphi_h, z_h - \varphi_h) = a(f_0 - \varphi_h, z_h - \varphi_h) + a(z_h - f_0, z_h - \varphi_h) \\ &= r(g_0 - p_h, z_h - \varphi_h) + a(z_h - f_0, z_h - \varphi_h) \\ &\leq \|g_0 - p_h\|_{L^2(\Omega)} \|z_h - \varphi_h\|_{L^2(\Omega)} + \|z_h - f_0\|_{H^1(\Omega)} \|z_h - \varphi_h\|_{H^1(\Omega)} \\ &\leq C \left[\|g_0 - p_h\|_{L^2(\Omega)} + \|z_h - f_0\|_{H^1(\Omega)} \right] \|z_h - \varphi_h\|_{H^1(\Omega)}. \end{aligned}$$

The term $\|f_0 - \varphi_h\|_n^2$ can be bounded with the L^∞ -norm of the same quantity. We have in fact that

$$\frac{1}{n} \sum_{i=1}^n (f_0(\mathbf{p}_i) - \varphi_h(\mathbf{p}_i))^2 \leq \max_{\mathbf{p}_i} (f_0(\mathbf{p}_i) - \varphi_h(\mathbf{p}_i))^2 \leq \|f_0 - \varphi_h\|_{L^\infty}^2.$$

We now define $f_{0h} \in V_h^r$ such that $a(f_{0h}, \psi_h) - r(g_0, \psi_h) = F(\psi_h) \forall \psi_h \in V_{h,\Gamma_D}^r$ and $f_{0h}|_{\Gamma_D} = \gamma_{D,h}$; the error term can be split in two parts:

$$(6.8) \quad \|f_0 - \varphi_h\|_{L^\infty(\Omega)} \leq \|f_0 - f_{0h}\|_{L^\infty(\Omega)} + \|f_{0h} - \varphi_h\|_{L^\infty(\Omega)}.$$

The first term on the right-hand side of the inequality represents the L^∞ -norm of the finite element error of the elliptic equation. The quantity f_{0h} can in fact be seen as the finite element approximation of the exact solution f_0 . Thanks to the Sobolev inequality (5.4), the L^∞ -norm of the finite element error can be bounded with the $W^{1,p}$ -norm ($p > d$) of the same quantity, i.e.,

$$(6.9) \quad \|f_0 - f_{0h}\|_{L^\infty(\Omega)} \leq C \|f_0 - f_{0h}\|_{W^{1,p}(\Omega)}.$$

Hence (see, e.g., [2])

$$\|f_0 - f_{0h}\|_{W^{1,p}(\Omega)} \leq C \inf_{\substack{z_h \in V_h^r \\ z_h|_{\Gamma_D} = \gamma_{D,h}}} \|f_0 - z_h\|_{W^{1,p}(\Omega)}.$$

The second term on the right-hand side of the inequality can be bounded by

$$\|f_{0h} - \varphi_h\|_{L^\infty(\Omega)} \leq C \|g_0 - p_h\|_{L^2(\Omega)}.$$

This bound is obtained thanks to the L^∞ -stability of the operator $\mathcal{B}_h : L^2(\Omega) \rightarrow V_{h,\Gamma_D}^r$, defined in (5.3) as the operator such that $f_{0h} - \varphi_h = \mathcal{B}_h(g_0 - p_h)$ is the solution of the problem

$$a(f_{0h} - \varphi_h, v_h) = r(g_0 - p_h, v_h) \quad \forall v_h \in V_{h,\Gamma_D}^r.$$

Therefore

$$(6.10) \quad \|f_0 - \varphi_h\|_n \leq C \left(\inf_{\substack{z_h \in V_h^r \\ z_h|_{\Gamma_D} = \gamma_{D,h}}} \|f_0 - z_h\|_{W^{1,p}(\Omega)} + \|g_0 - p_h\|_{L^2(\Omega)} \right).$$

Collecting the bounds (6.7) and (6.10), since $\|g_0 - p_h\|_{L^2(\Omega)} \leq \|g_0\|_{L^2(\Omega)}$, $W^{1,p}(\Omega) \subseteq H^1(\Omega)$ for $p \geq 2$, and Ω is bounded, we obtain for $p > d$, $d = 2, 3$, the unconstrained upper bound

$$(6.11) \quad \left\| f_0 - \hat{f}_h^* \right\|_n^2 + \lambda \left[\left\| f_0 - \hat{f}_h^* \right\|_{H^1(\Omega)}^2 + \|g_0 - \hat{g}_h^*\|_{L^2(\Omega)}^2 \right] \\ \leq C \left\{ (1 + \lambda) \inf_{\substack{z_h \in V_h^r \\ z_h|_{\Gamma_D} = \gamma_D, h}} \|f_0 - z_h\|_{W^{1,p}(\Omega)}^2 + \inf_{p_h \in V_{h,\Gamma_D}^r} \|g_0 - p_h\|_{L^2(\Omega)}^2 + \lambda \|g_0\|_{L^2(\Omega)}^2 \right\}.$$

The classical error bound for the interpolant $\Pi_h^r v \in V_h^r$ of $v \in W^{r+1,p}(\Omega)$ with $p > 1$,

$$(6.12) \quad \|v - \Pi_h^r v\|_{W^{k,p}(\Omega)} \leq Ch^{r+1-k} |v|_{W^{r+1,p}(\Omega)},$$

provides

$$\inf_{\substack{z_h \in V_h^r \\ z_h|_{\Gamma_D} = \gamma_D, h}} \|f_0 - z_h\|_{W^{1,p}(\Omega)} \leq Ch^r |f_0|_{W^{r+1,p}(\Omega)}, \\ \inf_{p_h \in V_{h,\Gamma_D}^r} \|g_0 - p_h\|_{L^2(\Omega)} \leq Ch^{r+1} |g_0|_{H^{r+1}(\Omega)}. \quad \blacksquare$$

Notice that the inequality (6.6) can be split in two terms: the first term on the right-hand side goes to zero for $h \rightarrow 0$ with the optimal convergence rate, while the second term $\|g_0\|_{L^2(\Omega)}^2$ is the same bias term obtained in the error splitting (4.5) and goes to zero when $\lambda \rightarrow 0$.

By closer inspection of the proof of Theorem 3, we can show that it is possible to obtain a better order of convergence requiring a slightly higher regularity to f_0 and g_0 and considering only the seminorm $\|\cdot\|_n$ and the L^2 -norm of the adjoint variable.

Theorem 4. *Under Assumption 3, using finite elements of degree r , if $f_0 \in W^{r+1,\infty}(\Omega)$ with $f_0|_{\Gamma_D} = \gamma_D$ and $g_0 \in H^{r+1}(\Omega)$ with $g_0|_{\Gamma_D} = 0$, there exists $h_0 > 0$ such that $\forall h \leq h_0$*

$$\left\| f_0 - \mathbb{E}[\hat{f}_h] \right\|_n^2 + \lambda \|g_0 - \mathbb{E}[\hat{g}_h]\|_{L^2(\Omega)}^2 \leq C \left\{ h^{2(r+1)} \left[\phi^2(h) \|f_0\|_{W^{r+1,\infty}(\Omega)}^2 + |g_0|_{H^{r+1}(\Omega)}^2 \right] \right. \\ \left. + \lambda \|g_0\|_{L^2(\Omega)}^2 \right\}, \\ \left\| f_0 - \mathbb{E}[\hat{f}_h] \right\|_{H^1(\Omega)}^2 \leq Ch^{2r} \left\{ |f_0|_{H^{r+1}(\Omega)}^2 + |g_0|_{H^r(\Omega)}^2 \right\},$$

where

$$(6.13) \quad \phi(h) = \begin{cases} \log(h), & r = 1, \\ 1, & r > 1. \end{cases}$$

Proof. From Lemma 3 we obtain that

$$\left\| f_0 - \mathbb{E}[\hat{f}_h] \right\|_n^2 + \lambda \|g_0 - \mathbb{E}[\hat{g}_h]\|_{L^2(\Omega)}^2 \leq C \left\{ \inf_{(\varphi_h, p_h) \in \mathcal{W}_h} \left[\|f_0 - \varphi_h\|_n^2 + \lambda \|g_0 - p_h\|_{L^2(\Omega)}^2 \right] \right. \\ \left. + \lambda \|g_0\|_{L^2(\Omega)}^2 \right\}, \\ \left\| f_0 - \mathbb{E}[\hat{f}_h] \right\|_{H^1(\Omega)}^2 \leq C \inf_{(\varphi_h, p_h) \in \mathcal{W}_h} \|f_0 - \varphi_h\|_{H^1(\Omega)}^2.$$

The unconstrained upper bounds can be obtained following the same strategy used in the proof of Theorem 3. The increase in the convergence rate is obtained using a different bound for the first term on the right-hand side of the inequality (6.8), i.e.,

$$\|f_0 - f_{0h}\|_{L^\infty(\Omega)} \leq Ch\phi(h) \|f_0 - f_{0h}\|_{W^{1,\infty}(\Omega)} \leq Ch\phi(h) \inf_{\substack{z_h \in V_h^r \\ z_h|_{\Gamma_D} = \gamma_{D,h}}} \|f_0 - z_h\|_{W^{1,\infty}(\Omega)}.$$

Now exploiting the error bound in the $W^{1,\infty}$ -norm (see, e.g., [2])

$$\inf_{\substack{z_h \in V_h^r \\ z_h|_{\Gamma_D} = \gamma_{D,h}}} \|f_0 - z_h\|_{W^{1,\infty}(\Omega)} \leq Ch^r \|f_0\|_{W^{r+1,\infty}(\Omega)},$$

we obtain

$$\|f_0 - f_{0h}\|_{L^\infty(\Omega)} \leq Ch^{r+1}\phi(h) \|f_0\|_{W^{r+1,\infty}(\Omega)}$$

and the desired result. ■

These rates of convergence are confirmed by the numerical results shown in section 8.

Remark 3. In this work we propose an equal order finite element approximation for \hat{f} and \hat{g} . Equal order finite elements are known to lead to suboptimal convergence rates for the fourth order biharmonic problem (see, e.g., [4, 5]). However, here we are able to recover the optimal convergence rate thanks to the fact that the boundary conditions that are naturally associated to the smoothing problem are the same for \hat{f} and \hat{g} .

Remark 4. It should be noticed that the optimal convergence rate is recovered only if g_0 satisfies exactly the homogeneous Dirichlet boundary conditions on Γ_D , which might be a restrictive hypothesis. If g_0 does not satisfy the Dirichlet boundary conditions, we should expect a boundary error decaying as $h^{1/2}$ in both two and three dimensions.

7. Surface estimator for areal data. The smoothing method presented in section 2 can be extended to the case of areal data that represent linear quantities computed on some subregions. This is useful in many applications of interest, and it is, for instance, the case of the estimation of the blood velocity field from echo-Doppler data presented in [1]; the echo-Doppler data represent, in fact, the mean velocity of the blood cells on a subdomain within an artery and cannot be approximated with pointwise observations.

Let $D_i \subset \Omega$, for $i = 1, \dots, N$, be some subdomains and \bar{z}_i , for $i = 1, \dots, N$, be the mean value of a quantity of interest on the subdomains. Starting from the model for pointwise observations, we can derive the following model for the observations \bar{z}_i :

$$(7.1) \quad \bar{z}_i = \frac{1}{|D_i|} \int_{D_i} f_0 + \eta_i;$$

see [1] for details. The error terms η_i have zero mean and variances $\bar{\sigma}_i^2$. Under the assumption that the number of observations in a subdomain is proportional to its dimension, the variances $\bar{\sigma}_i^2$ depend inversely on the dimension of the beams D_i .

In order to estimate the field we propose minimizing the penalized sum-of-square-error functional

$$(7.2) \quad \bar{J}(f) = \frac{1}{N} \sum_{i=1}^N \frac{1}{|D_i|} \left(\int_{D_i} (f - \bar{z}_i) d\mathbf{p} \right)^2 + \lambda \int_{\Omega} (Lf - u)^2$$

over the space V_γ , defined in section 2. The first term is a weighted least square error functional for areal data over subdomains D_i , weighted with the inverse of the variances $\bar{\sigma}_i^2$, under the assumption that $\bar{\sigma}_i^2 \propto 1/|D_i|$.

Existence and uniqueness of the estimator $\hat{f} = \operatorname{argmin}_{f \in V_\gamma} \bar{J}(f)$ are provided by the following theorem.

Theorem 5. *The estimator \hat{f} exists, is unique, and is obtained solving the system of PDEs*

$$(7.3) \quad \begin{cases} L\hat{f} = u + \hat{g} & \text{in } \Omega, \\ \mathcal{B}_c \hat{f} = \gamma & \text{on } \partial\Omega, \end{cases} \quad \begin{cases} L^* \hat{g} = -\frac{1}{N\lambda} \sum_{i=1}^N \frac{1}{|D_i|} \mathbb{I}_{D_i} \int_{D_i} (\hat{f} - \bar{z}_i) & \text{in } \Omega, \\ \mathcal{B}_c^* \hat{g} = 0 & \text{on } \partial\Omega, \end{cases}$$

where $\hat{g} \in \mathcal{G}$ represents the misfit of the penalized PDE, L^* is the adjoint operator of L , described by (3.9), and \mathcal{B}_c^* is the operator that defines the boundary conditions of the adjoint problem, summarized in (3.11).

The proof is analogous to the proof of Theorem 2. The existence and uniqueness of the estimator are in fact obtained, thanks to Theorem 1, writing the functional $\bar{J}(f)$ as the quadratic form (3.2). The proof of the well-posedness of the problem in the areal case is easier than that presented in section 3, and it is similar to classical results in control theory. Data are in fact distributed, and it is not necessary to require more regularity, as in the case of pointwise observations, for the system of PDEs (7.3) to be well defined.

The estimator is then discretized by means of the mixed finite element method described in section 5. In addition to the matrices used in section 5, we also define the matrices of the spatial average of the basis functions on the subdomains D_i , $\bar{\Psi}_{ik} = 1/|D_i| \int_{D_i} \psi_k$ and $\bar{\Psi}_{ik}^D = 1/|D_i| \int_{D_i} \psi_k^D$, the weight matrix $\mathbf{W} = \operatorname{diag}(|D_1|, \dots, |D_N|)$, and the vector of mean values on subdomains, $\bar{\mathbf{z}} = (\bar{z}_1, \dots, \bar{z}_N)^T$. The finite element estimator \hat{f}_h can thus be written as $\hat{f}_h = \hat{\mathbf{f}}^T \boldsymbol{\psi} + \hat{\mathbf{f}}_D^T \boldsymbol{\psi}^D$, where $\hat{\mathbf{f}}$ is the solution of the linear system

$$(7.4) \quad \begin{bmatrix} \bar{\Psi}^T \mathbf{W} \bar{\Psi} / (N\lambda) & \mathbf{A}^T \\ \mathbf{A} & -\mathbf{R} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{f}} \\ \hat{\mathbf{g}} \end{bmatrix} = \begin{bmatrix} \bar{\Psi}^T \mathbf{W} \bar{\mathbf{z}} / (N\lambda) - \bar{\Psi}^T \bar{\Psi}^D \mathbf{f}_D / (N\lambda) \\ \mathbf{u} + \mathbf{h}_N + \mathbf{h}_R - \mathbf{A}^D \mathbf{f}_D \end{bmatrix}.$$

For more details on the properties of the estimator see [1].

As in the case of pointwise observations we can obtain a bound for the bias of the estimator \hat{f} that corresponds exactly to the bound (4.5). We can also exploit the results obtained in the pointwise case to study the convergence of the bias of the finite element estimator. Moreover, in the areal data setting it is possible to relax the hypothesis on f_0 and g_0 in Theorem 3.

Theorem 6. *Under Assumption 1, using finite elements of degree r , if $f_0 \in H^{r+1}(\Omega)$ with $f_0|_{\Gamma_D} = \gamma_D$ and $g_0 \in H^r(\Omega)$ with $g_0|_{\Gamma_D} = 0$, there exists $h_0 > 0$ such that $\forall h \leq h_0$*

$$(7.5) \quad \begin{aligned} & \frac{1}{N} \sum_{i=1}^N \frac{1}{|D_i|} \left(\int_{D_i} (f_0 - \mathbb{E}[\hat{f}^*]) \right)^2 + \lambda \left[\|f_0 - \mathbb{E}[\hat{f}_h]\|_{H^1(\Omega)}^2 + \|g_0 - \mathbb{E}[\hat{g}_h]\|_{L^2(\Omega)}^2 \right] \\ & \leq C \left\{ h^{2r} (1 + \lambda) \left[\|f_0\|_{H^{r+1}(\Omega)}^2 + \|g_0\|_{H^r(\Omega)}^2 \right] + \lambda \|g_0\|_{L^2(\Omega)}^2 \right\}. \end{aligned}$$

Proof. If we define $\hat{f}_h^* = \mathbb{E}[\hat{f}_h]$, $\hat{g}_h^* = \mathbb{E}[\hat{g}_h]$, and the bilinear form $l(\cdot, \cdot)$ as

$$l(f, \psi) = \frac{1}{N} \sum_{i=1}^N \frac{1}{|D_i|} \int_{D_i} f \int_{D_i} \psi,$$

we can easily obtain the bound (7.5) for the bias in the norm $\|\cdot\|$ defined as

$$\|\hat{f}_h\| = \frac{1}{N} \sum_{i=1}^N \frac{1}{|D_i|} \left(\int_{D_i} \hat{f}_h \right)^2 + \lambda \left[\|\hat{f}_h\|_{H^1(\Omega)}^2 + \|\hat{g}_h\|_{L^2(\Omega)}^2 \right].$$

Since the norm associated to the bilinear form $l(\cdot, \cdot)$ is bounded by the H^1 -norm, we have that

$$\begin{aligned} & \frac{1}{N} \sum_{i=1}^N \frac{1}{|D_i|} \left(\int_{D_i} f_0 - \hat{f}_h^* \right)^2 + \lambda \left[\|f_0 - \hat{f}_h^*\|_{H^1(\Omega)}^2 + \|g_0 - \hat{g}_h^*\|_{L^2(\Omega)}^2 \right] \\ & \leq C \left\{ (1 + \lambda) \inf_{(\varphi_h, p_h) \in \mathcal{W}_h} \left[\|f_0 - \varphi_h\|_{H^1(\Omega)}^2 + \|g_0 - p_h\|_{L^2(\Omega)}^2 \right] + \lambda \|g_0\|_{L^2(\Omega)}^2 \right\}. \end{aligned}$$

The inequality (6.7) still holds for $(\varphi_h, p_h) \in \mathcal{W}_h$ and $z_h \in V_h^r$, and we obtain

$$\begin{aligned} & \frac{1}{N} \sum_{i=1}^N \frac{1}{|D_i|} \left(\int_{D_i} f_0 - \hat{f}_h^* \right)^2 + \lambda \left[\|f_0 - \hat{f}_h^*\|_{H^1(\Omega)}^2 + \|g_0 - \hat{g}_h^*\|_{L^2(\Omega)}^2 \right] \\ & \leq C \left\{ (1 + \lambda) \left[\inf_{\substack{z_h \in V_h^r \\ z_h|_{\Gamma_D} = \gamma_{D,h}}} \|f_0 - z_h\|_{H^1(\Omega)}^2 + \inf_{p_h \in V_{h,\Gamma_D}^r} \|g_0 - p_h\|_{L^2(\Omega)}^2 \right] + \lambda \|g_0\|_{L^2(\Omega)}^2 \right\}. \end{aligned}$$

Using the classical error bound (6.12) we obtain the desired result. ■

8. Numerical simulations. In this section we verify in simple settings the convergence results shown in section 6 for pointwise data and in section 7 for areal data.

8.1. Test 1. We first verify the convergence results obtained for pointwise data. We consider the bidimensional domain $\Omega = [0, 1] \times [0, 1]$, and we assume that the true underlying surface f_0 satisfies the following PDE:

$$(8.1) \quad \begin{cases} \Delta f_0 = 2[x(x-1) + y(y-1)] & \text{in } \Omega, \\ f_0 = 0 & \text{on } \partial\Omega, \end{cases}$$

whose solution, $f_0 = xy(x-1)(y-1)$, is represented in Figure 1, left. We consider the $n = 200$ observation points $\mathbf{p}_1, \dots, \mathbf{p}_n$, represented in Figure 1, right, the roughness parameter λ fixed, and we want to test the convergence of $\|\mathbb{E}[\hat{f}_h] - f_0\|$ when $h \rightarrow 0$. For this reason we solve the estimation problem on different uniform structured meshes with size $h = 1/2, 1/4, \dots, 1/2^9$.

We will consider different settings.

- A. The observations are without noise, i.e., $z_i = f_0(\mathbf{p}_i)$, and the functional $J(f)$ penalizes the misfit of the governing PDE (8.1), i.e., $L = \Delta$ and $u = 2[x(x-1) + y(y-1)]$.
- B. The observations are without noise, i.e., $z_i = f_0(\mathbf{p}_i)$, but the functional $J(f)$ penalizes the misfit of a PDE different from the governing PDE (8.1). In particular, $L = \Delta$ but $u \neq 2[x(x-1) + y(y-1)]$:
 1. the penalized forcing term u is such that $g_0 = Lf_0 - u$ satisfies homogeneous Dirichlet boundary conditions on $\partial\Omega$: $u = 2(x(x-1) + y(y-1))(1 + (x(x-1)y(y-1)))$;

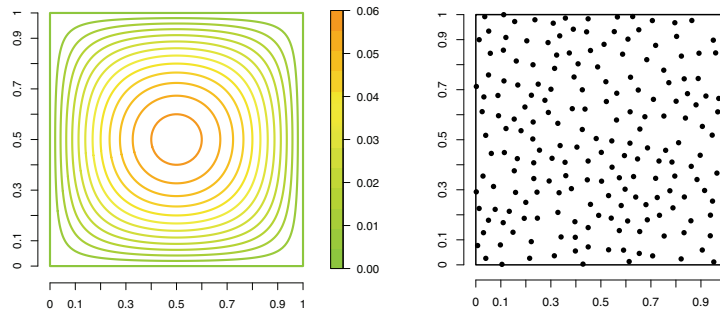


Figure 1. Left: true surface f_0 used for the simulation studies of Test 1; the image displays the isolines $(0, 0.005, 0.01, \dots, 0.06)$. Right: location points sampled uniformly on the domain for Tests 1 and 2.

2. different penalized forcing terms u are considered, such that $g_0 = Lf_0 - u$ does not satisfy homogeneous Dirichlet boundary conditions on $\partial\Omega$:
 - (a) $u = (x(x-1) + y(y-1))$, which corresponds to the knowledge of the real forcing term up to a multiplying constant factor;
 - (b) $u = 2x(x-1)$, which corresponds to the knowledge of only a part of the real forcing term;
 - (c) $u = 2(x(x-1) + y(y-1)) + (x^{10} + y^{10} + (x-1)^{10} + (y-1)^{10})$, which forces g_0 to be equal to -1 on the boundary, with a relatively large boundary layer.
- C. The observations are with noise, i.e., $z_i = f_0(\mathbf{p}_i) + \epsilon_i$, where $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$, and the functional $J(f)$ penalizes the misfit of the governing PDE (8.1), i.e., $L = \Delta$ and $u = 2[x(x-1) + y(y-1)]$.
- D. The observations are with noise, i.e., $z_i = f_0(\mathbf{p}_i) + \epsilon_i$, where $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$, and the functional $J(f)$ penalizes the misfit of a PDE different from the governing PDE (8.1). In particular, $L = \Delta$ but $u = 2(x(x-1) + y(y-1))(1 + (x(x-1)y(y-1)))$, as in case B1.

Case A (no bias, no noise). We solve the estimation problem both with linear and quadratic finite elements fixing the roughness parameter $\lambda = 1$. We recall that we are using the same order of approximation for \hat{f}_h and \hat{g}_h . The results of the linear and the quadratic mixed finite element approximations are shown, respectively, in the left and right panels of Figure 2. In particular we show the convergence of the error $\|\hat{f}_h - f_0\|$ as well as the convergence of each individual term of the norm $\|\cdot\|$, namely $\|\hat{f}_h - f_0\|_n$, $\|\hat{f}_h - f_0\|_{H^1(\Omega)}$, and $\|\hat{g}_h - g_0\|_{L^2(\Omega)}$. Since we are considering the case of observations without noise, $\mathbb{E}[\hat{f}_h] = \hat{f}_h$ and $\mathbb{E}[\hat{g}_h] = \hat{g}_h$. We notice that with both the linear and the quadratic approximations we obtain a rate of convergence equal to or higher than the expected rate for all the error terms. In particular, the H^1 -norm of the error is the dominating term in both the linear and the quadratic approximations, and it decays as h in the case of linear finite elements and as h^2 in the case of quadratic finite elements. All the other terms are negligible. As expected, the norm $\|\cdot\|_n$ of $\hat{f}_h - f_0$ and the L^2 -norm of $\hat{g}_h - g_0$ decay as h^2 in the case of linear finite elements and at least as h^3 in the case of quadratic finite elements.

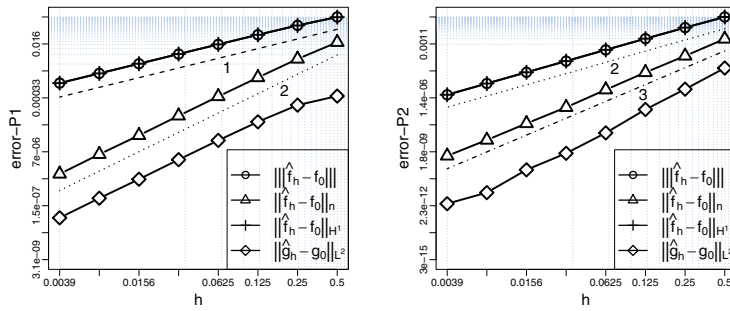


Figure 2. Test 1, case A: convergence rates of the bias of the estimator obtained with $\lambda = 1$ in the norms $\|\hat{f}_h - f_0\|$, $\|\hat{f}_h - f_0\|_n$, $\|\hat{f}_h - f_0\|_{H^1(\Omega)}$, and $\|\hat{g}_h - g_0\|_{L^2(\Omega)}$. Left: linear mixed finite element approximation. Right: quadratic mixed finite element approximation.

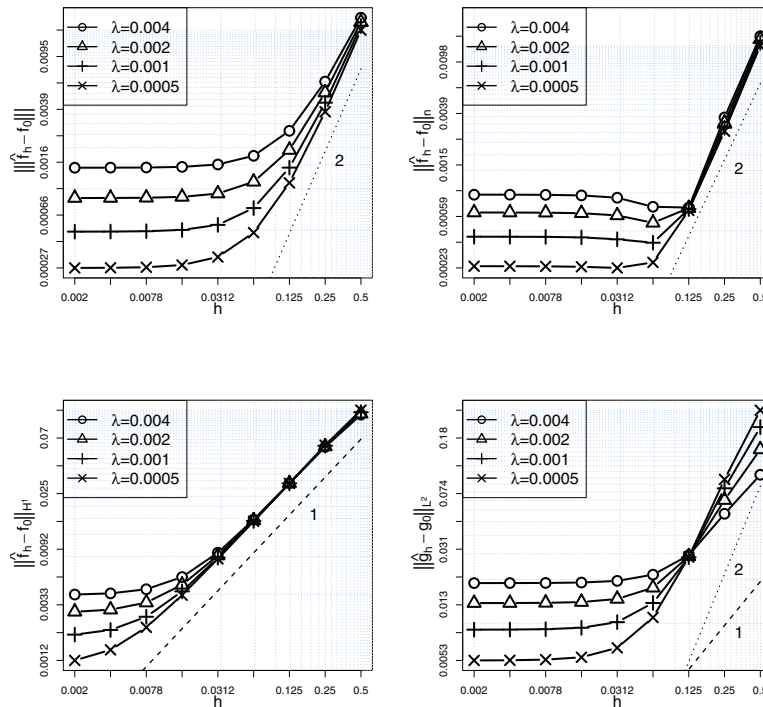


Figure 3. Test 1, case B1: convergence rates of the bias of the estimator obtained with linear finite elements when $\lambda = 0.1, 0.2, 0.4, 0.8$. Top left: $\|\hat{f}_h - f_0\|$, top right: $\|\hat{f}_h - f_0\|_n$, bottom left: $\|\hat{f}_h - f_0\|_{H^1(\Omega)}$, bottom right: $\|\hat{g}_h - g_0\|_{L^2(\Omega)}$.

Case B1 (bias with exact boundary conditions, no noise). We solve the estimation problem with linear finite elements, and we study the convergence for different values of the roughness parameter λ . Recall that, in this case, $g_0 = \Delta f_0 - u \neq 0$ satisfies the homogeneous Dirichlet boundary conditions. Figure 3 shows the rate of convergence of the error in different norms when $\lambda = 0.1, 0.2, 0.4, 0.8$. As in case A, since the observations are without noise, $\mathbb{E}[\hat{f}_h] = \hat{f}_h$

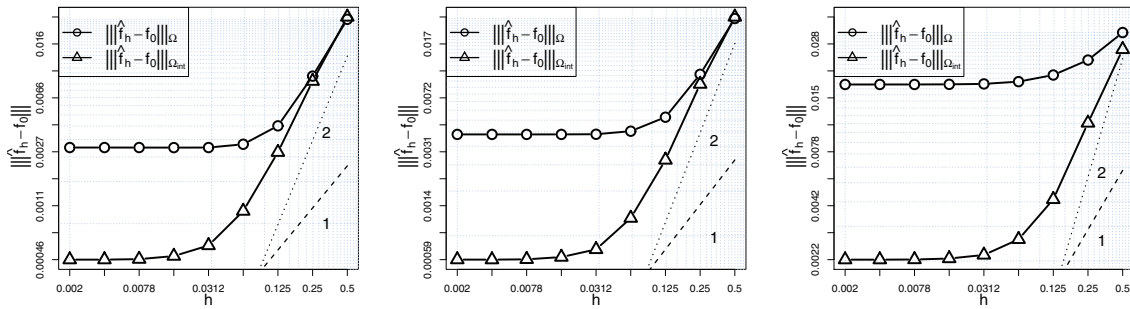


Figure 4. Test 1, case B2: convergence rates of $\|\hat{f}_h - f_0\|_\Omega$ and $\|\hat{f}_h - f_0\|_{\Omega_{\text{int}}}$ using linear finite elements with $\lambda = 0.002$. Left: case B2(a), center: case B2(b), right: B2(c).

and $\mathbb{E}[\hat{g}_h] = \hat{g}_h$. Notice that when the mesh is fine the approximation error in the norm $\|\cdot\|$ asymptotically approaches a value proportional to $\sqrt{\lambda}$, as expected from Theorem 3; this behavior is caused by the presence of the bias term in the error bound (6.6). In this case the dominant term is the L^2 -norm of $\hat{g}_h - g_0$. This term has a different behavior for different values of λ : if λ is sufficiently small, it decays as h^2 before approaching the asymptote; otherwise it decays as h . It is thus necessary to use small values of λ in order to recover the expected convergence rate h^2 . However, even when a large value of λ is used, the rate of convergence of $\|\hat{f}_h - f_0\|$ is at least linear before reaching the saturation level caused by the bias term. The other two terms instead decay with the expected convergence rate for all the values of λ , before approaching the asymptote.

Case B2 (bias with wrong boundary conditions, no noise). We consider three different forcing terms u such that $g_0 = \Delta f_0 - u \neq 0$ does not satisfy the homogeneous Dirichlet boundary conditions. In this case we study the error in the norm $\|\cdot\|$ over the whole domain Ω , which will be denoted by $\|\hat{f}_h - f_0\|_\Omega$, as well as over the subdomain $\Omega_{\text{int}} = [0.1, 0.9] \times [0.1, 0.9]$, denoted by $\|\hat{f}_h - f_0\|_{\Omega_{\text{int}}}$. As highlighted in Remark 4, the former error should be affected by a boundary term decaying as $h^{1/2}$, while the latter does not include the error at the boundary. As in cases A and B1, since the observations are without noise, $\mathbb{E}[\hat{f}_h] = \hat{f}_h$ and $\mathbb{E}[\hat{g}_h] = \hat{g}_h$. The results obtained with the three forcing terms considered in B2(a), B2(b), and B2(c) are represented, respectively, in the left, center, and right panels of Figure 4. In theory we would expect a different rate of convergence for the errors $\|\hat{f}_h - f_0\|_\Omega$ and $\|\hat{f}_h - f_0\|_{\Omega_{\text{int}}}$, which should be more clearly visible when the mesh is fine. On the other hand, the numerical simulations do not display any significant difference between the convergence rates of the two errors in all three cases. This is due to the presence of the bias, which is asymptotically approached by both the error terms, hiding the expected convergence rate. Thus, using a forcing term such that g_0 does not satisfy the homogeneous Dirichlet boundary conditions does not affect the surface estimation too much.

Case C (no bias, with noise). We add some noise to the pointwise evaluations $f_0(\mathbf{p}_i)$ of the surface: for each location point we sample independent errors, $\epsilon_1, \dots, \epsilon_n$, from a zero mean Gaussian distribution $\mathcal{N}(0, \sigma^2)$, with different standard deviations $\sigma = 0.005, 0.01, 0.02$. The first value of σ corresponds to a rather high signal to noise ratio, since the value of the true surface varies from 0 to 0.062, while the last corresponds to a very low signal to noise ratio

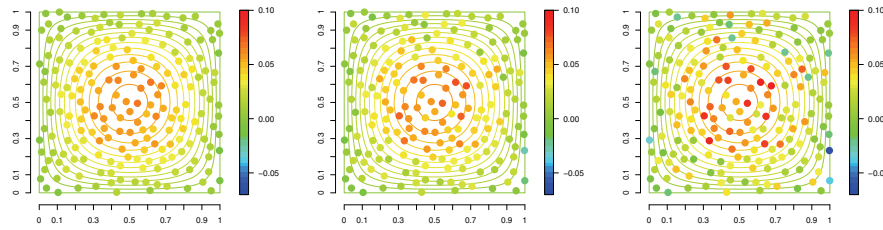


Figure 5. Test 1, case C: value of the observations z_1, \dots, z_n obtained from model (2.1) adding noise with different values of standard deviation σ , superimposed to the true underlying surface f_0 (the image displays the isolines $(0, 0.005, 0.01, \dots, 0.06)$). Left: $\sigma = 0.005$, center: $\sigma = 0.01$, right: $\sigma = 0.02$.

with errors of the same order of magnitude of the variation of f_0 . The experiment is repeated 50 times for each value of σ . Figure 5 shows the values z_i obtained from model (2.1) in the first replicate of the experiment. We can notice that the observations with small additive noise, represented in the left panel of Figure 5, are similar to the evaluation of f_0 in the sampling points, while the observations with large errors, represented in the right panel of the same figure, are far from the true underlying surface. The results obtained solving the estimation problem with linear finite elements, with fixed roughness parameter $\lambda = 1$ and penalizing the exact PDE, are shown in Figure 6. To show the variability of the error convergence over the 50 replicates, we represent the boxplots of the approximation errors obtained for each value of σ and each grid size h . The convergence rate of the experiment representing the median behavior of the convergence rates over 50 repetitions of the experiment is shown in Figure 6 as a thick line superimposed to the boxplots. The median convergence rates and the boxplots are colored according to the corresponding value of σ . Notice that in Figure 6 the represented approximation errors include both the error associated to the finite element discretization and the error associated to the noisy observations. Due to the presence of noise, both $\|\hat{f}_h - f_0\|$ and $\|\hat{f}_h - f_0\|_n$ quite soon reach a saturation limit proportional to the standard deviation of the noise σ . However, further refining the mesh still provides better approximation of the first derivatives, as shown by the convergence of $\|\hat{f}_h - f_0\|_{H^1(\Omega)}$. Notice also that the variability of the error, highlighted in the picture by the height of the boxplots, depends on the grid size and in particular increases as h decreases. This variability does not appear to depend on the standard deviation of the noise, being comparable in the three cases ($\sigma = 0.005, 0.01, 0.02$).

Case D (bias, with noise). Analogously to case C, we add some noise to the pointwise evaluations $f_0(\mathbf{p}_i)$ of the surface: for each location point we sample independent errors, $\epsilon_1, \dots, \epsilon_n$, from a zero mean Gaussian distribution $\mathcal{N}(0, \sigma^2)$, with different standard deviations $\sigma = 0.005, 0.01, 0.02$. We solve the estimation problem with linear finite elements, and we study the convergence for different values of the roughness parameter $\lambda = 0.1, 0.2, 0.4, 0.8$. Recall that, analogously to case B1, $g_0 = \Delta f_0 - u \neq 0$ and it satisfies the homogeneous Dirichlet boundary conditions. For each value of σ and each value of λ , the experiment is repeated 50 times. As in case C, Figure 7 displays the boxplots of the approximation errors, to highlight the variability of the error convergence over the 50 replicates, and the thick lines, to show the convergence rate of the experiment representing the median behavior of the convergence rates. Notice that in Figure 7 the represented approximation errors include the error

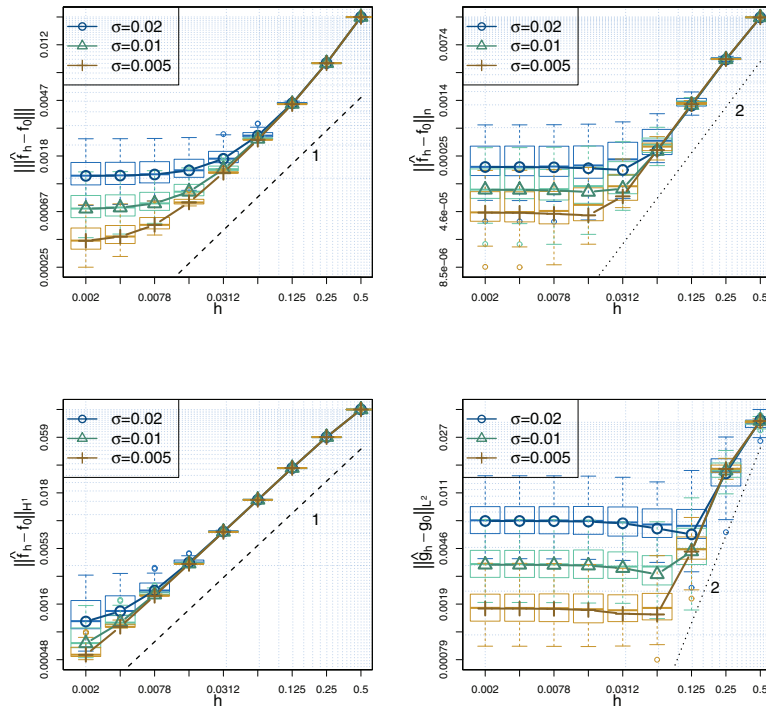


Figure 6. Test 1, case C: convergence rates of the approximation error of the estimator obtained with linear finite elements and $\lambda = 1$ when the error in the observations is generated from a Gaussian distribution with different standard deviations $\sigma = 0.005$ (brown), $\sigma = 0.01$ (green), and $\sigma = 0.02$ (blue). For each value of σ , the thick line corresponds to the convergence rate of the experiment representing the median behavior of the convergence rates. The median convergence rate is superimposed to the boxplots of the approximation errors computed at different values of h . Top left: $\|\hat{f}_h - f_0\|$, top right: $\|\hat{f}_h - f_0\|_n$, bottom left: $\|\hat{f}_h - f_0\|_{H^1(\Omega)}$, bottom right: $\|\hat{g}_h - g_0\|_{L^2(\Omega)}$.

associated to the finite element approximation, the bias induced by the penalty term, and the error associated to the noisy observations. The saturation limit reached by the approximation error depends on the last two terms. In particular, when the roughness parameter λ is large (top panels of Figure 7), the saturation limit obtained for different values of σ is almost the same since the bias is predominant. Instead, as λ decreases (bottom panels of Figure 7), the saturation limit becomes proportional to the standard deviation of the noise, as in case C. Notice also that the variability of the error, highlighted by the height of the boxplots, depends not only on the grid size (likewise in case C) but also on the value of the roughness parameter λ . Indeed, for small values of λ the contribution of the penalty term is negligible and the results are analogous to case C. On the other hand, for large values of λ , the bias induced by the penalty term is predominant and the variability depends on standard deviation of the error; in particular, the larger the standard deviation, the higher the variability of the estimate.

8.2. Test 2. We verify the convergence results in the pointwise data framework also in a different simulation study concerning a diffusion-transport-reaction (DTR) PDE. We consider

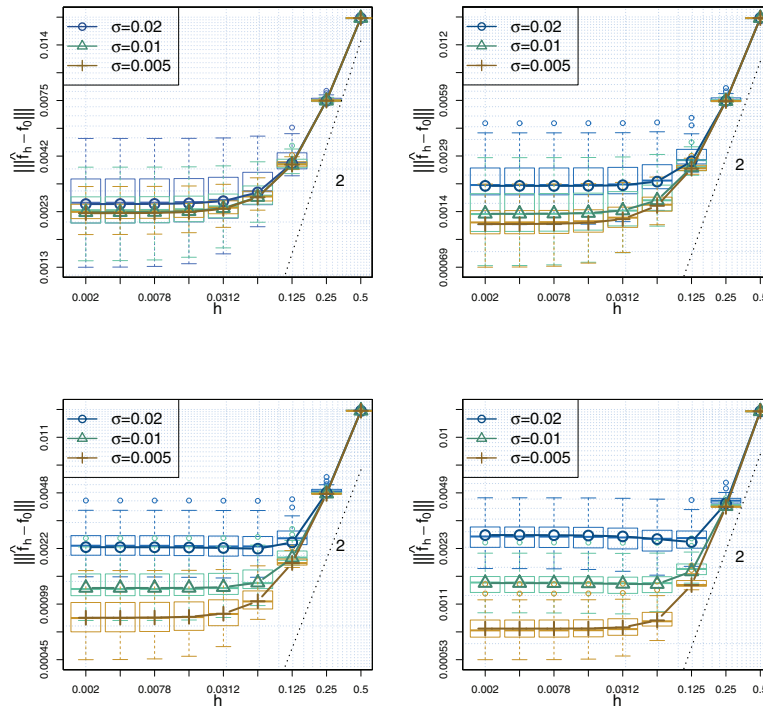


Figure 7. Test 1, case D: convergence rates of $\|\hat{f}_h - f_0\|$ using linear finite elements and different values of λ when the error in the observations is generated from a Gaussian distribution with different standard deviations $\sigma = 0.005$ (brown), $\sigma = 0.01$ (green), and $\sigma = 0.02$ (blue). For each value of σ , the thick line corresponds to the convergence rate of the experiment representing the median behavior of the convergence rates. The median convergence rate is superimposed to the boxplots of the approximation errors computed at different values of h . Top left: $\lambda = 0.8$, top right: $\lambda = 0.4$, bottom left: $\lambda = 0.2$, bottom right: $\lambda = 0.1$.

the domain $\Omega = [0, 1] \times [0, 1]$, and we assume that the true underlying field f_0 satisfies the following PDE:

$$(8.2) \quad \begin{cases} Lf_0 = -1 & \text{in } \Omega, \\ f_0 = 0 & \text{on } \partial\Omega, \end{cases}$$

where the operator L is the DTR operator defined in (2.3) with parameters $\mathbf{K}_{11} = 4$, $\mathbf{K}_{22} = 1$, $\mathbf{K}_{12} = \mathbf{K}_{21} = 0$, $\mathbf{b}_1 = 2$, $\mathbf{b}_2 = 1$, and $c = 1$, \mathbf{K}_{ij} and \mathbf{b}_i being, respectively, the element (i, j) of the diffusion tensor \mathbf{K} and the i th element of the transport vector \mathbf{b} . The solution of the PDE (8.2) is represented in Figure 8. We consider the $n = 200$ observation points $\mathbf{p}_1, \dots, \mathbf{p}_n$ represented in Figure 1, right. For the sake of brevity, in this case we show the convergence of $\|\mathbb{E}[\hat{f}_h] - f_0\| = \|\hat{f}_h - f_0\|$ only when the observations are without noise and the functional $J(f)$ penalizes the misfit of the governing PDE (8.2). The results obtained solving the estimation problem with linear and quadratic finite elements and $\lambda = 1$ on different uniform structured meshes with size $h = 1/2, 1/4, \dots, 1/2^8$ are shown in Figure 9. We can notice that also in the DTR case we obtain a rate of convergence equal to or higher than the expected rate for all the error terms with both the linear and the quadratic approximations. The H^1 -norm is still

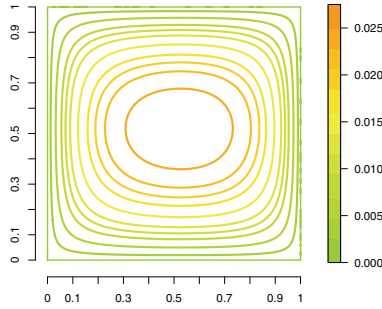


Figure 8. True surface f_0 used for the simulation study of Test 2; the image displays the isolines $(0, 0.0025, 0.005, \dots, 0.03)$.

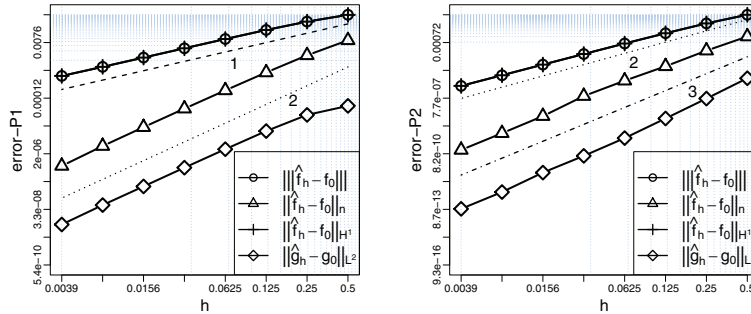


Figure 9. Test 2: convergence rates of the bias of the estimator obtained with $\lambda = 1$ in the norms $\|\hat{f}_h - f_0\|$, $\|\hat{f}_h - f_0\|_n$, $\|\hat{f}_h - f_0\|_{H^1(\Omega)}$, and $\|\hat{g}_h - g_0\|_{L^2(\Omega)}$. Left: linear mixed finite element approximation. Right: quadratic mixed finite element approximation.

the dominating term, while all the other terms are negligible. As in the Laplacian case, the error terms $\|\hat{f}_h - f_0\|_n$ and $\|\hat{g}_h - g_0\|_{L^2(\Omega)}$ decay as h^2 for linear finite elements and at least as h^3 for quadratic finite elements.

8.3. Test 3. Finally, we also verify the convergence results obtained for areal data. We consider the bidimensional domain $\Omega = [0, 1] \times [0, 1]$, and we assume, as in Test 1, that the true underlying surface f_0 satisfies the PDE (8.1), whose solution is represented in Figure 1, left. We consider the $N = 9$ subsets D_1, \dots, D_N represented in Figure 10, left. Each subset is colored according to the corresponding observation value \bar{z}_i . For the sake of brevity, we consider only the case of observations without noise, i.e., $\bar{z}_i = \frac{1}{|D_i|} \int_{D_i} f_0$. We test the convergence of $\|\mathbb{E}[\hat{f}_h] - f_0\| = \|\hat{f}_h - f_0\|$ when $h \rightarrow 0$, solving the estimation problem with linear finite elements and $\lambda = 1$ on different uniform structured meshes with size $h = 1/2^3, 1/2^4, \dots, 1/2^7$. Since we are considering the areal data framework, we substitute the seminorm $\|\cdot\|_n$ in (6.3) with the seminorm $\|f\|_N = \frac{1}{N} \sum_{i=1}^N \frac{1}{|D_i|} (\int_{D_i} f)^2$.

The results obtained minimizing the functional $\bar{J}(f)$, penalized with the misfit of the exact PDE (8.1), are shown in Figure 10, right. We can notice that we obtain the same rate

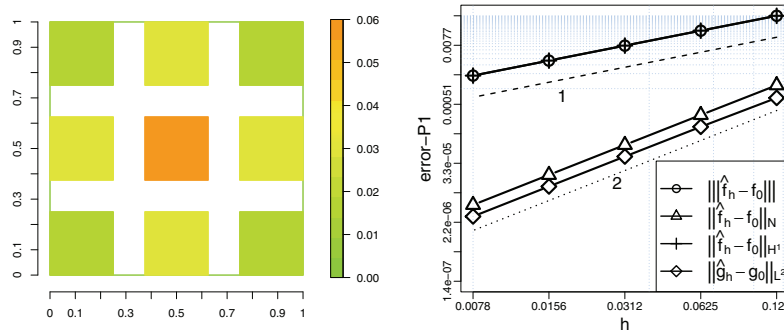


Figure 10. Left: subdomains used in Test 3, colored according to the corresponding observation value \bar{z}_i . Right: convergence rates of the bias of the estimator obtained in Test 3 with $\lambda = 1$ in the norms $\|\hat{f}_h - f_0\|$, $\|\hat{f}_h - f_0\|_N$, $\|\hat{f}_h - f_0\|_{H^1(\Omega)}$, and $\|\hat{g}_h - g_0\|_{L^2(\Omega)}$.

of convergence expected from Theorem 6. In particular, the H^1 -norm is the dominating term decaying as h , while $\|\hat{f}_h - f_0\|_N$ and $\|\hat{g}_h - g_0\|_{L^2(\Omega)}$ decay as h^2 .

9. Conclusion and future work. In this work we have studied the properties of the SR-PDE smoothing technique. This smoothing method has a very broad applicability since PDEs are commonly used to model physical phenomena. The method is actually not applicable to PDEs with discontinuous parameters, with pointwise forcing terms, or defined on irregular domains, due to the extra regularity required for the parameters of the penalized PDE. This request, however, is not restrictive in spatial statistics and in the smoothing framework since the field is normally assumed to be very regular. The proposed mixed finite element method requires, moreover, g_0 to satisfy the Dirichlet boundary conditions on Γ_D . This hypothesis could sometimes be restrictive since it means that the second derivatives of the field at the boundary are clumped to zero; other discretization methods for fourth order problems could be considered in the future. However, we have observed numerically that whenever g_0 does not satisfy the homogeneous Dirichlet boundary conditions on Γ_D , the extra error is of the same order as the bias contribution and therefore does not really compromise the optimal convergence rate of the method.

The convergence studied in this work concerns the bias of the estimator when the characteristic mesh size h goes to zero and neglects instead the error induced by the presence of noise in the observations. Classical results concerning smoothing splines and thin-plate splines (see, e.g., [6, 7, 8, 16]) show the consistency of these estimators when the smoothing parameter $\lambda(n)$ goes to zero, as $n \rightarrow +\infty$, with a proper rate. Unfortunately these results cannot be directly extended to SR-PDE, and a different approach needs to be developed to show the consistency of these models. We are currently studying the (infill) asymptotic properties of the estimator when the number of observations n goes to infinity. In particular we are studying the convergence of the variance term \hat{w} , in both the continuous and the discrete settings, when n goes to infinity, and we are looking for a proper rate of λ that makes both the bias and variance vanish.

It will also be interesting to balance the discretization error induced by the finite element

approximation with the bias of the estimator and the variance term related to the noise of the observations. A possible way to solve the problem is the development of a proper mesh adaptation technique, based on a posteriori estimates of noise, variance, and bias. This technique should locally refine the mesh in order to obtain a local discretization error that is of the same order or smaller than the bias and the noise standard deviation σ .

Acknowledgments. We are grateful to Jim Ramsay for insightful discussions and interesting suggestions. We also thank the editor and referees for their constructive comments that helped us to improve the manuscript.

REFERENCES

- [1] L. AZZIMONTI, L. M. SANGALLI, P. SECCHI, M. DOMANIN, AND F. NOBILE, *Blood flow velocity field estimation via spatial regression with PDE penalization*, J. Amer. Statist. Assoc., to appear; available online from <http://mox.polimi.it/it/progetti/publicazioni/>.
- [2] S. C. BRENNER AND L. R. SCOTT, *The Mathematical Theory of Finite Element Methods*, Springer, New York, 2008.
- [3] S. C. BRENNER AND L. Y. SUNG, *C^0 interior penalty methods for fourth order elliptic boundary value problems on polygonal domains*, J. Sci. Comput., 22/23 (2005), pp. 83–118.
- [4] F. BREZZI AND M. FORTIN, *Mixed and Hybrid Finite Elements Methods*, Springer, New York, 1991.
- [5] P. G. CIARLET, *The Finite Element Method for Elliptic Problems*, SIAM, Philadelphia, 2002.
- [6] D. D. COX, *Asymptotics for m -type smoothing splines*, Ann. Statist., 11 (1983), pp. 530–551.
- [7] D. D. COX, *Multivariate smoothing spline functions*, SIAM J. Numer. Anal., 21 (1984), pp. 789–813.
- [8] F. CUCKER AND D. X. ZHOU, *Learning Theory: An Approximation Theory Viewpoint*, Cambridge Monogr. Appl. Comput. Math. 24, Cambridge University Press, Cambridge, UK, 2007.
- [9] M. D’ELIA, M. PEREGO, AND A. VENEZIANI, *A variational data assimilation procedure for the incompressible Navier-Stokes equations in hemodynamics*, J. Sci. Comput., 52 (2012), pp. 340–359.
- [10] M. D’ELIA AND A. VENEZIANI, *Uncertainty quantification for data assimilation in a steady incompressible Navier-Stokes problem*, ESAIM Math. Model. Numer. Anal., 47 (2012), pp. 1037–1057.
- [11] H. W. ENGL, M. HANKE, AND G. NEUBAUER, *Regularization of Inverse Problems*, Kluwer Academic Publishers, Dordrecht, The Netherlands, 2000.
- [12] A. ERN AND J. L. GUERMOND, *Theory and Practice of Finite Elements*, Springer, New York, 2004.
- [13] B. ETTINGER, S. PEROTTO, AND L. M. SANGALLI, *Spatial Regression Models over Two-Dimensional Manifolds*, Technical Report 54/2012, MOX - Dipartimento di Matematica, Politecnico di Milano, Milano, Italy, 2012; available online from <http://mox.polimi.it/it/progetti/publicazioni/>.
- [14] L. C. EVANS, *Partial Differential Equations*, American Mathematical Society, Providence, RI, 1998.
- [15] E. H. GEORGIOULIS, P. HOUSTON, AND J. VIRTANEN, *An a posteriori error indicator for discontinuous Galerkin approximations of fourth-order elliptic problems*, IMA J. Numer. Anal., 31 (2011), pp. 281–298.
- [16] L. GYÖRFI, M. KOHLER, A. KRZYŻAK, AND H. WALK, *A Distribution-Free Theory of Nonparametric Regression*, Springer, New York, 2002.
- [17] T. J. HASTIE, R. J. TIBSHIRANI, AND J. FRIEDMAN, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer, New York, 2009.
- [18] M. HINZE, R. PINNAU, M. ULBRICH, AND S. ULBRICH, *Optimization with PDE Constraints*, Springer, Berlin, 2009.
- [19] J.-L. LIONS, *Optimal Control of Systems Governed by Partial Differential Equations*, Springer, New York, Berlin, 1971.
- [20] A. QUARTERONI, *Primal hybrid finite element method for 4th order elliptic equations*, Calcolo, 16 (1979), pp. 21–59.
- [21] A. QUARTERONI, *Numerical Models for Differential Problems*, Springer, Milan, 2008.
- [22] J. O. RAMSAY AND B. W. SILVERMAN, *Functional Data Analysis*, Springer, New York, 2005.
- [23] T. RAMSAY, *Spline smoothing over difficult regions*, J. R. Stat. Soc. Ser. B Stat. Methodol., 64 (2002),

- pp. 307–319.
- [24] G. ROZZA, A. MANZONI, AND F. NEGRI, *Reduction strategies for PDE-constrained optimization problems in haemodynamics*, in Proceedings of the European Congress on Computational Methods in Applied Sciences and Engineering (ECCOMAS 2012), 2012.
 - [25] L. M. SANGALLI, J. O. RAMSAY, AND T. RAMSAY, *Spatial spline regression models*, J. R. Stat. Soc. Ser. B Stat. Methodol., 75 (2013), pp. 681–703.
 - [26] G. STONE, *Bivariate Splines*, Ph.D. thesis, University of Bath, Bath, UK, 1988.
 - [27] A. M. STUART, *Inverse problems: A Bayesian perspective*, Acta Numer., 19 (2010), pp. 451–559.
 - [28] E. SÜLI AND I. MOZOLEVSKI, *hp-version interior penalty DGFEMs for the biharmonic equation*, Comput. Methods Appl. Mech. Engrg., 196 (2007), pp. 1851–1863.
 - [29] C. R. VOGEL, *Computational Methods for Inverse Problems*, SIAM, Philadelphia, 2002.
 - [30] G. WAHBA, *Spline Models for Observational Data*, SIAM, Philadelphia, 1990.
 - [31] S. N. WOOD, M. V. BRAVINGTON, AND S. L. HEDLEY, *Soap film smoothing*, J. R. Stat. Soc. Ser. B Stat. Methodol., 70 (2008), pp. 931–955.