Spatial spline regression models

Author(s): Laura M. Sangalli, James O. Ramsay and Timothy O. Ramsay

Source: *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, SEPTEMBER 2013, Vol. 75, No. 4 (SEPTEMBER 2013), pp. 681-703

Published by: Wiley for the Royal Statistical Society

Stable URL: https://www.jstor.org/stable/24772451

J. R. Statist. Soc. B (2013)
75, *Part* 4, *pp.* 681–703

Spatial spline regression models

Laura M. Sangalli,

Politecnico di Milano, Italy

James O. Ramsay

McGill University, Montréal, Canada

and Timothy O. Ramsay

Ottawa Health Research Institute, Canada

[Received January 2012. Revised October 2012]

Summary. We describe a model for the analysis of data distributed over irregularly shaped spatial domains with complex boundaries, strong concavities and interior holes. Adopting an approach that is typical of functional data analysis, we propose a spatial spline regression model that is computationally efficient, allows for spatially distributed covariate information and can impose various conditions over the boundaries of the domain. Accurate surface estimation is achieved by the use of piecewise linear and quadratic finite elements.

Keywords: Finite elements; Functional data analysis; Penalized smoothing; Semiparametric model; Spatial data analysis

1. Introduction

We propose a semiparametric model for the analysis of data distributed over spatial domains, including those with complex domain boundaries, strong concavities and interior holes. Fig. 1, showing census tract locations over the Island of Montréal, Québec, Canada, but excluding the airport and rail yards in the south and an industrial park with an oil refinery tank farm in the north-east, illustrates the kind of problem that we consider. Population density, average *per capita* income and other measures are available at each census tract, and a binary variable indicating whether a tract is predominantly residential or industrial and commercial is available as covariate for estimating the distributions of census quantities. Here in particular we are interested in population density; the airport and industrial park are thus not part of the domain of interest since people cannot live in those two areas. Census quantities can vary sharply across these uninhabited parts of the city; for instance, in the south of the industrial park there is a densely populated area with medium–low income, but north-east and west of it are wealthy neighbourhoods, with low population density to the north-east, and high population density to the west. Hence, although it seems reasonable to assume that population density features a smooth spatial variation over the inhabited parts of the island, there is no reason to assume smooth spatial variation across interior boundaries. Fig. 1 also shows the island coasts as boundaries of the domain of interest; those parts of the boundary that are highlighted in

Address for correspondence: Laura M. Sangalli, Laboratorio di Modellistica e Calcolo Scientifico, Dipartimento di Matematica, Politecnico di Milano, Piazza L. da Vinci 32, Milano 20133, Italy.
E-mail: laura.sangalli@polimi.it

 1369–7412/13/75681

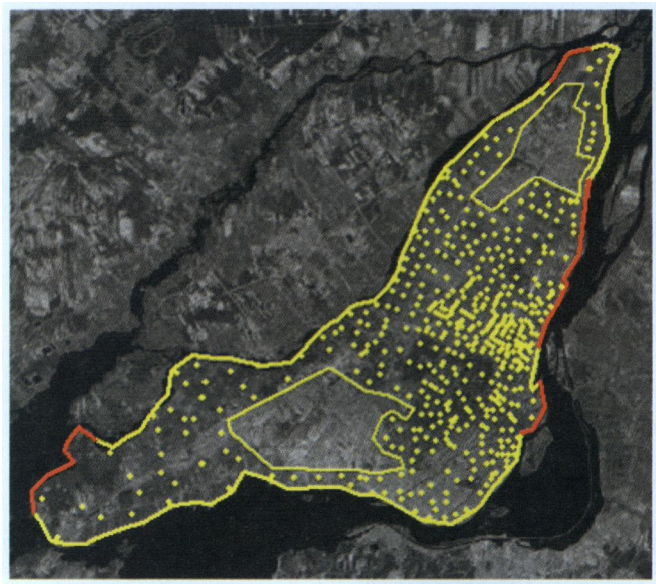

Fig. 1. Island of Montréal census data (○, centroids of census enumeration areas, for which population density and other census information are available): the two parts of the island where there are no data (encircled) are areas where people cannot live (the airport and rail yards in the south and an industrial park with an oil refinery tank farm in the north-east); the island boundary is also outlined, with the darker sections indicating the harbour and two public parks

red correspond respectively to the harbour, in the east shore, and to two public parks, in the south-west and north-east shore; and no people live by the river banks in these boundary intervals. We thus want to study population density, taking into account covariate information, being careful not to link data artificially across areas where people cannot live, and also efficiently including prior information concerning those stretches of coast where the population density should drop to zero. Well-known methods for spatial data analysis, such as kriging, kernel smoothing, wavelet-based smoothing, tensor product splines and thin plate splines, are not appropriate for these data, since they do not take into account the shape of the domain and also smooth across concave boundary regions; moreover, these methods do not allow the specification of fixed values for the surface estimate at the domain boundary, or along parts of it.

We adopt a functional data analysis approach in proposing a *spatial spline regression* (SSR) model that overcomes these limitations, being able to deal efficiently with data distributed over irregularly shaped regions. The model incorporates the penalized bivariate spline smoother that was introduced by Ramsay (2002); in this smoother, the roughness penalty consists of a Laplace operator that is integrated only over the region of interest thanks to a finite element formulation, that defines a system of local basis functions for continuous piecewise polynomial surfaces. We improve the Ramsay (2002) smoother in many respects, from both computational and modelling perspectives. The modelling generalizations include the capacity to account for covariate measures and also to comply with different conditions at the boundary of the domain; the surfaces describing the spatial variation may either exhibit free-value behaviour at the exterior or interior boundaries of the domain, with a specified gradient in the normal direction, or be constrained to desired fixed values. Moreover, these boundary conditions can vary over different intervals on the boundaries, and linear combinations of them may be applied as well. SSR estimators

turn out to be linear in the observed data values, so classical inferential tools may be readily derived. The finite element formulation is also computationally highly efficient.

SSR is compared with kriging, thin plate splines and soap film smoothing, the last introduced by Wood *et al.* (2008) and used for instance in Marra *et al.* (2012) in a tensor product smoother that also includes the time dimension. Our simulation studies show that SSR and soap film smoothing provide a large advantage over the other more classical techniques when dealing with data that are scattered over irregularly shaped domains. SSR is also similar to the spatial data analysis models that were introduced by Guillas and Lai (2010) which also penalize roughness with a partial differential operator. Finally, SSR models have also strong connections with the work of Lindgren *et al.* (2011), that links Gaussian fields and Gaussian Markov random fields via a stochastic partial differential equation that induces a Matérn covariance and is solved over irregular grids of points resorting to finite elements.

1.1. Data and model

Let $\{\mathbf{p}_i = (x_i, y_i); i = 1, \ldots, n\}$ be a set of n points on a bounded regular domain $\Omega \subset \mathbb{R}^2$. Let z_i be the value of a real-valued variable observed at point $\mathbf{p}_i$, and let $\mathbf{w}_i = (w_{i1}, \ldots, w_{iq})^{\mathrm{T}}$ be a q-vector of covariates associated with observation z_i.

The semiparametric model for these data is

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

where ε_i, $i = 1, \ldots, n$, are residuals or errors distributed independently of each other, with zero mean and variance σ^2. Vector $\boldsymbol{\beta} \in \mathbb{R}^q$ contains regression coefficients and function f is real valued and twice differentiable. We estimate the regression coefficient vector $\boldsymbol{\beta}$ and the surface or spatial field f by minimizing the penalized sum of square error functional

$$J_\lambda(\boldsymbol{\beta}, f) = \sum_{i=1}^{n} \{z_i - \mathbf{w}_i^{\mathrm{T}} \boldsymbol{\beta} - f(\mathbf{p}_i)\}^2 + \lambda \int_\Omega (\Delta f)^2. \tag{2}$$

The roughness penalty is the integral over Ω of the square of the Laplacian of f,

$$\Delta f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2},$$

which is a measure of local curvature of f that is invariant with respect to Euclidean transformations of spatial co-ordinates and therefore ensures that the concept of smoothness does not depend on the orientation of the co-ordinate system. To lighten the notation, surface integrals will be written without the integration variable $\mathbf{p}$, i.e., for any integrable function g defined over $D \subseteq \mathbb{R}^2$, $\int_D g := \int_D g(\mathbf{p})\, \mathrm{d}\mathbf{p}$.

The estimation problem (2) is tackled by means of finite element analysis, which is a methodology that is mainly developed and used in engineering applications, to solve partial differential equations. The strategy of finite element analysis is very similar in spirit to that of univariate splines and consists of partitioning the problem domain into small disjoint subdomains and defining polynomial functions on each of these subdomains in such a way that the union of these pieces closely approximates the solution. Convenient domain partitions are given for instance by triangular meshes; Fig. 10 in Section 6.3 shows for example a triangulation of the domain of interest for the Island of Montréal data. The simplified problem is made computationally tractable by the choice of the basis functions for the space of piecewise polynomials on the domain partition. Each piece of the partition, equipped with the basis functions that are defined over it, is named a *finite element*. Introductions to finite element analysis are offered, for example, by Gockenbach (2006), Quarteroni (2009) and Braess (2007).

Many methods for surface or spatial field estimation define the estimate as the minimizer of a penalized sum of square error functional, with the roughness penalty involving a partial differential operator. For instance, the regularizing term in problem (2) can be rewritten as $\int_\Omega (f_{xx}^2 + 2f_{xx}f_{yy} + f_{yy}^2)$ to facilitate comparison with thin plate spline penalty $\int_{\mathbb{R}^2}(f_{xx}^2 + 2f_{xy}^2 + f_{yy}^2)$, the latter being integrated over the entire plane $\mathbb{R}^2$. The soap film smoothing technique that was introduced by Wood *et al.* (2008) uses the same penalty as in problem (2) and in Ramsay (2002) but follows a different analytic and numerical approach to find the minimizer. The spatial models that were proposed by Guillas and Lai (2010) consider a regularizing term involving linear combinations of all partial derivatives up to a chosen order and employ bivariate splines over triangulations (see, for example, Lai and Schumaker (2007)), that, like finite elements, provide a basis for piecewise polynomial surfaces. Within the framework that is proposed in the current paper, the regularizing term in problem (2) can be generalized to deal with more complicated partial differential operators; the interesting implications of this line of research will be discussed in Section 7.

The rest of the paper is organized as follows. Section 2 presents the estimation problem in functional or variational form as a preliminary to the development of the finite element approximation to the functional solution. Section 3 describes the finite element function spaces that are used for the approximation, and Section 4 derives the finite element solution to the estimation problem and the properties of the associated estimators. Section 5 focuses on how to deal with various boundary conditions. Section 6 illustrates the performances of the proposed method via simulation studies and an application to the Island of Montréal census data. Finally, Section 7 draws some conclusive considerations and discusses extensions of the model considered and directions of future research. Detailed proofs are deferred to Appendix A.

2. The estimation problem in variational terms

Let $H^m(\Omega)$ be the set of functions in $L^2(\Omega)$ having all derivatives up to order m in $L^2(\Omega)$. Let $H_{n0}^m(\Omega) \subset H^m(\Omega)$ consist of those functions whose normal derivatives are 0 on the boundary of Ω. Denote by $\mathbf{W}$ the $n \times q$ matrix whose ith row is given by $\mathbf{w}_i^{\mathrm{T}}$, the vector of q covariates associated with observation z_i at $\mathbf{p}_i$, and assume that $\mathbf{W}$ has full rank. Let $\mathbf{P}$ be the orthogonal projection matrix that projects orthogonally on the subspace of $\mathbb{R}^n$ generated by the columns of $\mathbf{W}$, i.e. $\mathbf{P} := \mathbf{W}(\mathbf{W}^{\mathrm{T}}\mathbf{W})^{-1}\mathbf{W}^{\mathrm{T}}$, and let $\mathbf{Q} := \mathbf{I} - \mathbf{P}$, where $\mathbf{I}$ is the identity matrix. Furthermore, set $\mathbf{z} := (z_1, \ldots, z_n)^{\mathrm{T}}$ and, for a given function f on Ω, denote by $\mathbf{f}_n$ the vector of evaluations of f at the n data locations, i.e. $\mathbf{f}_n := (f(\mathbf{p}_1), \ldots, f(\mathbf{p}_n))^{\mathrm{T}}$.

The penalized sum of square error functional (2) is well defined for $\boldsymbol{\beta} \in \mathbb{R}^q$ and $f \in H^2(\Omega)$. Furthermore, as is shown in Appendix A.1, imposing a boundary condition on f, such as $f \in H_{n0}^2(\Omega)$, ensures that the estimation problem has a unique solution. In particular, $f \in H_{n0}^2(\Omega)$ corresponds to the assumption of zero flow across boundaries; see also the discussion on boundary conditions in Section 5. For this reason, we shall start by considering the estimation problem over $\boldsymbol{\beta} \in \mathbb{R}^q$ and $f \in H_{n0}^2(\Omega)$. These assumptions on functional spaces, as well as other assumptions on functional spaces that we shall make in what follows, can be relaxed; however, for simplicity of exposition, we do not pursue this here. Proposition 1 characterizes the solution of the estimation problem.

Proposition 1. The estimators $\hat{\boldsymbol{\beta}}$ and $\hat{f}$ that jointly minimize model (1), over $\boldsymbol{\beta} \in \mathbb{R}^q$ and $f \in H_{n0}^2(\Omega)$, exist unique:

(a) $\hat{\boldsymbol{\beta}} = (\mathbf{W}^{\mathrm{T}}\mathbf{W})^{-1}\mathbf{W}^{\mathrm{T}}(\mathbf{z} - \hat{\mathbf{f}}_n)$;
(b) $\hat{f}$ satisfies

$$\mathbf{u}_n^{\mathrm{T}}\mathbf{Q}\hat{\mathbf{f}}_n + \lambda \int_{\Omega} \Delta u \, \Delta \hat{f} = \mathbf{u}_n^{\mathrm{T}}\mathbf{Q}\mathbf{z} \tag{3}$$

for every $u \in H^2_{n0}(\Omega)$.

For a proof, see Appendix A.1.

The following sections show how an approximate solution to the estimation problem can be obtained by using the finite element method. In particular, the variational problem (3) is reformulated as a problem in $H^1(\Omega)$ and this problem is thus discretized by using a finite element space included in $H^1(\Omega)$. Section 3 briefly reviews the construction of this finite element space.

3. Lagrange triangular finite elements

We consider a regular triangulation $\mathcal{T}$ of Ω, where adjacent triangles share either a vertex or a complete edge. Domain Ω is hence approximated by domain $\Omega_{\mathcal{T}}$ consisting of the union of all triangles, so that the boundary $\partial\Omega$ of Ω is approximated by a polygon (or more polygons, in the case for instance of domains with interior holes). It is assumed, therefore, that the number and density of triangles in $\mathcal{T}$ are sufficient to capture sharp features in $\partial\Omega$ as well as providing a basis for adequately describing the data. We shall assume in this paper, for simplicity only, that the triangulation points in $\mathcal{T}$ are such that the first n coincide with the data location points $\mathbf{p}_i$, i.e. letting $\mathbf{v}_j, j = 1, \ldots, J$, indicate vertices of triangles, we have that $\mathbf{v}_i = \mathbf{p}_i$, $i = 1, \ldots, n$, where $n \leqslant J$. In particular, for the simulations and application in Section 6 we use Delaunay triangulations of the set of data location points, constrained within the domain of interest. Delaunay triangulations (see, for example, Hjelle and Dæhlen (2006)) of a set of points $\mathbf{V}$ are such that no point in $\mathbf{V}$ is inside the circumcircle of any triangle; they maximize the minimum angle of all the triangle angles, avoiding stretched triangles. Starting from set $\mathbf{V}$ of data locations has the advantage of providing triangulations that are naturally finer where there are more data points, and coarser where data points are fewer; the triangulation can of course be refined according to criteria concerning maximal allowed triangle edge and minimal allowed triangle angle. Our methodology can also be easily extended to allow the triangulation points and the data location points to be separate. It must be emphasized that the triangular mesh generation problem is far from trivial, is the subject of a large literature and constitutes a great amount of the technical challenge in applications of finite element analysis. Triangulation software is readily available in many free and commercial finite element packages.

The surface to be constructed over $\Omega_{\mathcal{T}}$ is assumed to be a polynomial in x and y over any triangle and is continuous across edges and vertices. In this paper we consider the two cases where the polynomial is either linear or quadratic (the latter case being that considered in Ramsay (2002)). The linear case defines the linear polynomial over each triangle as linear combination of three basis functions, each having value 1 at a single vertex and 0 at the other two. The vertices that are used in this way are also called *control points* or *nodes*, and are indicated by $\boldsymbol{\xi}_k, k = 1, \ldots, K$. In the quadratic case, six nodes are required for each triangle, each node being associated with the unique quadratic polynomial that has value 1 at a single node and 0 at the remaining five. In this quadratic case, we extend the set of vertices to include midpoints of edges in defining the nodes $\boldsymbol{\xi}_k$. To summarize, nodes and vertices coincide for piecewise linear surfaces, but quadratic surfaces involve roughly twice as many nodes as vertices.

Since adjacent triangles share an edge or a vertex, it is automatic that a piecewise linear or quadratic basis function ψ_k is associated with each node $\boldsymbol{\xi}_k$ that has value 1 at the node and value 0 at all neighbouring nodes. Each such piecewise polynomial basis function is called a *Lagrange*

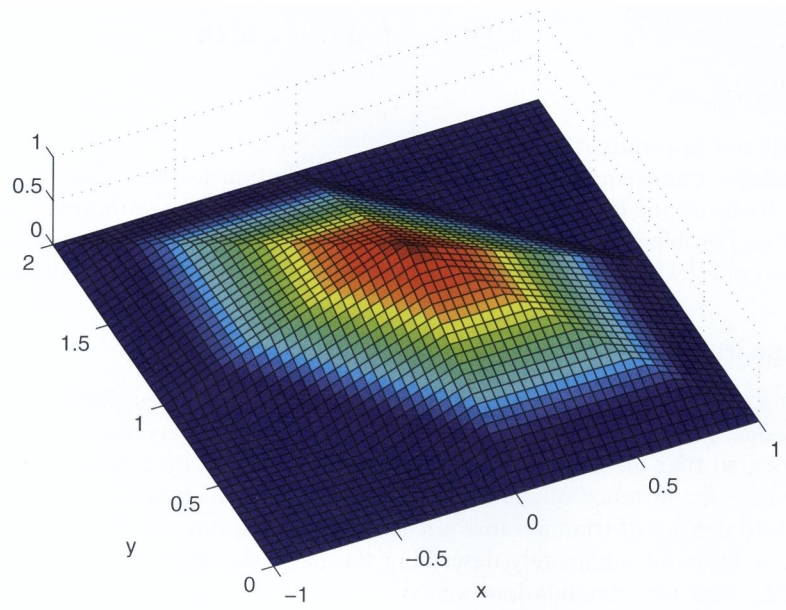

Fig. 2. Nodal basis associated with node $\xi = (0, 1)$ over a uniform triangular mesh

finite element. Fig. 2 shows for the linear case one such basis function; its shape depends on the number of triangles sharing the vertex $\boldsymbol{\xi}_k$ and on the lengths of the triangle edges. The set of K basis functions constructed in this way defines a function subspace $H^1_{\mathcal{T}}(\Omega) \subset H^1(\Omega_{\mathcal{T}})$.

Set $\boldsymbol{\psi} := (\psi_1, \ldots, \psi_K)^{\mathrm{T}}$; moreover, for a given function f on Ω, denote by $\mathbf{f}$ the K-vector having as entries the evaluations of f at the K nodes, i.e. $\mathbf{f} := (f(\boldsymbol{\xi}_1), \ldots, f(\boldsymbol{\xi}_K))^{\mathrm{T}}$. Being $\psi_k(\boldsymbol{\xi}_l) = \delta_{kl}$, the finite element space has been constructed precisely so that any function f in $H^1_{\mathcal{T}}(\Omega)$ is completely defined by its value at the K nodes:

$$f(x, y) = \sum_{k=1}^{K} c_k \psi_k(x, y) = \sum_{k=1}^{K} f(\boldsymbol{\xi}_k) \psi_k(x, y) = \mathbf{f}^{\mathrm{T}} \boldsymbol{\psi}(x, y). \tag{4}$$

This Lagrangian property of the basis is very convenient from a computational point of view and will be exploited extensively in what follows.

4. Finite element solution to the estimation problem

In Appendix A.2 we show, by introducing an auxiliary function g and using integration by parts, that the problem of finding $\hat{f} \in H^2_{n0}(\Omega)$ that satisfies condition (3) for every $u \in H^2_{n0}(\Omega)$ may be reformulated as follows: find $(\hat{f}, g) \in \{H^1_{n0}(\Omega) \cap C^0(\Omega)\} \times H^1(\Omega)$ that satisfies

$$\begin{aligned} \mathbf{u}_n^{\mathrm{T}} \mathbf{Q} \hat{\mathbf{f}}_n - \lambda \int_{\Omega} (\nabla u \cdot \nabla g) &= \mathbf{u}_n^{\mathrm{T}} \mathbf{Q} \mathbf{z}, \\ \int_{\Omega} v g + \int_{\Omega} (\nabla v \cdot \nabla \hat{f}) &= 0 \end{aligned} \tag{5}$$

for all $(u, v) \in \{H^1_{n0}(\Omega) \cap C^0(\Omega)\} \times H^1(\Omega)$; moreover, such $\hat{f}$ belongs to $H^2_{n0}(\Omega)$. The finite element method requires the solution of this weak formulation of the estimation problem within

a finite element space $H^1_{\mathcal{T}}(\Omega)$. Corollary 1 shows that, thanks to the choice for domain partition and function basis of the finite element space, finding the solution to this discrete counterpart of the estimation problem reduces to solving a linear system.

Let $\boldsymbol{\psi}_x := (\partial\psi_1/\partial x, \ldots, \partial\psi_k/\partial x)^{\mathrm{T}}$ and $\boldsymbol{\psi}_y := (\partial\psi_1/\partial y, \ldots, \partial\psi_k/\partial y)^{\mathrm{T}}$, and define the order K matrices

$$\mathbf{R}_0 := \int_{\Omega_{\mathcal{T}}} \boldsymbol{\psi}\boldsymbol{\psi}^{\mathrm{T}},$$

$$\mathbf{R}_1 := \int_{\Omega_{\mathcal{T}}} (\boldsymbol{\psi}_x\boldsymbol{\psi}_x^{\mathrm{T}} + \boldsymbol{\psi}_y\boldsymbol{\psi}_y^{\mathrm{T}}).$$

Moreover, let us also introduce the order K block matrix $\mathbf{L}$, and the $K \times n$ block matrix $\mathbf{D}$, defined by

$$\mathbf{L} := \left(\begin{array}{c|c} \mathbf{Q} & \mathbf{O}_{n\times(K-n)} \\ \hline \mathbf{O}_{(K-n)\times n} & \mathbf{O}_{(K-n)\times(K-n)} \end{array}\right),$$

$$\mathbf{D} := \left(\begin{array}{c} \mathbf{I}_n \\ \hline \mathbf{O}_{(K-n)\times n} \end{array}\right)$$

where $\mathbf{O}_{m_1\times m_2}$ is an $m_1 \times m_2$ matrix with all entries equal to 0. Finally, denote by $\mathbf{0}$ the null vector.

Corollary 1. The estimators $\hat{\boldsymbol{\beta}} \in \mathbb{R}^q$ and $\hat{f} \in H^1_{\mathcal{T}}(\Omega)$, that solve the discrete counterpart of the estimation problem, exist unique:

(a) $\hat{\boldsymbol{\beta}} = (\mathbf{W}^{\mathrm{T}}\mathbf{W})^{-1}\mathbf{W}^{\mathrm{T}}(\mathbf{z} - \hat{\mathbf{f}}_n)$;

(b) $\hat{f} = \hat{\mathbf{f}}^{\mathrm{T}}\boldsymbol{\psi}$, with $\hat{\mathbf{f}}$ satisfying

$$\begin{pmatrix} -\mathbf{L} & \lambda\mathbf{R}_1 \\ \lambda\mathbf{R}_1 & \lambda\mathbf{R}_0 \end{pmatrix} \begin{pmatrix} \mathbf{f} \\ \mathbf{g} \end{pmatrix} = \begin{pmatrix} -\mathbf{LDz} \\ \mathbf{0} \end{pmatrix}. \tag{6}$$

For a proof, see Appendix A.3.

As mentioned in Section 1, the boundary condition that the normal derivatives of f are 0 on the boundary of Ω, i.e. $f \in H^2_{n0}(\Omega)$, guarantees the uniqueness of the solution to the estimation problem (2). Although the estimate $\hat{f} \in H^1_{\mathcal{T}}(\Omega)$ is an approximation of $f \in H^2_{n0}(\Omega)$, the estimate $\hat{f}$ itself will not generally have normal derivatives equal to 0 on the domain boundary, since this has not been enforced in the construction of the finite element space. The boundary condition on normal derivatives is rather used as a so-called *natural boundary condition*, being exploited when using integration by parts to obtain reformulation (5) of the estimation problem (see Appendix A.2). See also Section 5.

Solving the linear system (6) is fast. In fact, although the system is typically large, being of order $2K$, it is highly sparse because the matrices $\mathbf{R}_0$ and $\mathbf{R}_1$ are highly sparse, since the cross-products of nodal basis functions and of their partial derivatives are mostly zero. As an example, for the Isle of Montréal census data, we used 626 nodes and only about 1% of the entries of $\mathbf{R}_0$ and 0.2% of the entries of $\mathbf{R}_1$ were non-zero.

We compute the integrals of the cross-products of nodal basis functions and of their partial derivatives in $\mathbf{R}_0$ and $\mathbf{R}_1$ exactly, instead of using quadrature approximations as in Ramsay (2002).

4.1. *Properties of the estimators*

Denote by $\mathbf{B}$ the order $2K$ matrix in system (6), and set $\mathbf{A} := -\mathbf{B}^{-1}$. Moreover, denote by $\mathbf{A}_n$

the order n matrix corresponding to the first n rows and n columns of $\mathbf{A}$, and by $\mathbf{A}_{Kn}$ the $K \times n$ matrix corresponding to the first K rows and n columns of $\mathbf{A}$. With a few simplifications we can thus write

$$\begin{aligned} \hat{\mathbf{f}}_n &= \mathbf{A}_n \mathbf{Q} \mathbf{z}, \\ \hat{\mathbf{f}} &= \mathbf{A}_{Kn} \mathbf{Q} \mathbf{z} = \mathbf{A}_{Kn} \mathbf{A}_n^{-1} \hat{\mathbf{f}}_n. \end{aligned} \tag{7}$$

This expression for $\hat{\mathbf{f}}$ highlights the fact that the finite element solution $\hat{f}$ is determined by $\hat{\mathbf{f}}_n$, the solution $\hat{f}$ at the n data points. Moreover, using the expression of $\hat{\mathbf{f}}_n$ in equations (7), we have

$$\hat{\boldsymbol{\beta}} = (\mathbf{W}^{\mathrm{T}}\mathbf{W})^{-1}\mathbf{W}^{\mathrm{T}}(\mathbf{I} - \mathbf{A}_n\mathbf{Q})\mathbf{z}.$$

Consequently the estimators $\hat{\mathbf{f}}$, $\hat{\mathbf{f}}_n$ and $\hat{\boldsymbol{\beta}}$ are linear functions of the data values; their properties are therefore straightforward to derive and classic inferential tools can be obtained. Recalling that $E[\mathbf{z}] = \mathbf{W}\boldsymbol{\beta} + \mathbf{f}_n$ and $\mathrm{var}(\mathbf{z}) = \sigma^2\mathbf{I}$, and exploiting the properties of the matrices involved (e.g. $\mathbf{Q}$ is symmetric and idempotent, $\mathbf{A}_n$ is symmetric, $\mathbf{QW} = \mathbf{O}_{n\times n}$ and $\mathbf{QW}(\mathbf{W}^{\mathrm{T}}\mathbf{W})^{-1} = (\mathbf{W}^{\mathrm{T}}\mathbf{W})^{-1}\mathbf{W}^{\mathrm{T}}\mathbf{Q} = \mathbf{O}_{n\times n}$), with a few simplifications we obtain the means and variances of the estimators $\hat{\mathbf{f}}_n$ and $\hat{\boldsymbol{\beta}}$:

$$\begin{aligned} E[\hat{\mathbf{f}}_n] &= \mathbf{A}_n\mathbf{Q}\mathbf{f}_n, \\ \mathrm{var}(\hat{\mathbf{f}}_n) &= \sigma^2\mathbf{A}_n\mathbf{Q}\mathbf{A}_n \end{aligned} \tag{8}$$

and

$$\begin{aligned} E[\hat{\boldsymbol{\beta}}] &= \boldsymbol{\beta} + (\mathbf{W}^{\mathrm{T}}\mathbf{W})^{-1}\mathbf{W}^{\mathrm{T}}(\mathbf{I} - \mathbf{A}_n\mathbf{Q})\mathbf{f}_n, \\ \mathrm{var}(\hat{\boldsymbol{\beta}}) &= \sigma^2(\mathbf{W}^{\mathrm{T}}\mathbf{W})^{-1} + \sigma^2(\mathbf{W}^{\mathrm{T}}\mathbf{W})^{-1}\mathbf{W}^{\mathrm{T}}(\mathbf{A}_n\mathbf{Q}\mathbf{A}_n)\mathbf{W}(\mathbf{W}^{\mathrm{T}}\mathbf{W})^{-1}. \end{aligned} \tag{9}$$

Consider the vector $\hat{\mathbf{z}}$ of fitted values at the n data points

$$\hat{\mathbf{z}} = \mathbf{W}\hat{\boldsymbol{\beta}} + \hat{\mathbf{f}}_n = (\mathbf{p} + \mathbf{Q}\mathbf{A}_n\mathbf{Q})\mathbf{z} = \mathbf{S}\mathbf{z}$$

where $\mathbf{S}$ denotes the smoothing matrix $\mathbf{S} := \mathbf{P} + \mathbf{Q}\mathbf{A}_n\mathbf{Q}$. The spatial spline regression estimator is thus a linear estimator, with the fitted values $\hat{\mathbf{z}}$ obtained from observations $\mathbf{z}$ via application of the linear operator $\mathbf{S}$, independent of $\mathbf{z}$. A commonly used measure of the equivalent degrees of freedom for linear estimators is given by $\mathrm{tr}(\mathbf{S})$ (see, for example, Buja *et al.* (1989), who first introduced this notion). The equivalent degrees of freedom of the SSR estimator,

$$\mathrm{tr}(\mathbf{S}) = q + \mathrm{tr}(\mathbf{A}_n\mathbf{Q}),$$

are given by the sum of the q degrees of freedom of the parametric part of the model (q being the number of covariates considered) and the equivalent degrees of freedom $\mathrm{tr}(\mathbf{A}_n\mathbf{Q})$ corresponding to the non-parametric part of the model; recall in fact that $\hat{\mathbf{f}}_n = (\mathbf{A}_n\mathbf{Q})\mathbf{z}$ as in equations (7). We can now estimate σ^2 by

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

This estimate, together with expressions (9) and (8), may be used to obtain approximate confidence intervals for $\boldsymbol{\beta}$ and approximate confidence bands for f. Furthermore, the value of the smoothing parameter λ may be selected by generalized cross-validation (see, for example, Ramsay and Silverman (2005), and references therein):

$$\mathrm{GCV}(\lambda) = \frac{1}{n\{1 - \mathrm{tr}(\mathbf{S})/n\}^2}(\mathbf{z} - \hat{\mathbf{z}})^{\mathrm{T}}(\mathbf{z} - \hat{\mathbf{z}}).$$

Finally, the value predicted for a new observation, at point $\mathbf{p}_{n+1}$ and with covariates $\mathbf{w}_{n+1}$, is given by

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

whose mean and variance can be obtained from expressions above; correspondingly, approximate prediction intervals may be also derived.

5. Boundary conditions

Often we would like the smoothing surface f to have specific values at the boundary of the domain, or in some part of the boundary domain. For instance, Azzimonti *et al.* (2012) study the blood flow velocity field in a section of a carotid artery, using data provided by echo colour dopplers. In this applied problem it is *a priori* known that blood flow velocity should be 0 at the arterial wall because of the friction between blood cells and arterial wall.

Thus, suppose we know that $f = f_{\partial\Omega}$ on $\partial\Omega$, where $f_{\partial\Omega}$ is sufficiently regular. We therefore want to estimate $\boldsymbol{\beta}$ and f by minimizing the penalized sum of square error functional (2) over $\boldsymbol{\beta} \in \mathbb{R}^q$ and $f \in H^2(\Omega)$, conditioned on $f|_{\partial\Omega} = f_{\partial\Omega}$. Denote by $H_0^m(\Omega)$ the subset of $H^m(\Omega)$ consisting of those functions vanishing on the boundary of Ω. Proposition 2 characterizes the solution of this estimation problem with fixed value boundary conditions.

Proposition 2. The estimators $\hat{\boldsymbol{\beta}}$ and $\hat{f}$ that jointly minimize expression (1), over $\boldsymbol{\beta} \in \mathbb{R}^q$ and $f \in H^2(\Omega)$ with $f|_{\partial\Omega} = f_{\partial\Omega}$, exist unique:

(a) $\hat{\boldsymbol{\beta}} = (\mathbf{W}^{\mathrm{T}}\mathbf{W})^{-1}\mathbf{W}^{\mathrm{T}}(\mathbf{z} - \hat{\mathbf{f}}_n)$;

(b) $\hat{f} = \hat{s} + \bar{f}$, where $\bar{f}$ is any fixed function in $H^2(\Omega)$ such that $\bar{f}|_{\partial\Omega} = f_{\partial\Omega}$ and with known gradient $\nabla\bar{f}$ and Laplacian $\Delta\bar{f}$, and $\hat{s} \in H_0^2(\Omega)$ satisfies

$$\mathbf{u}_n^{\mathrm{T}}\mathbf{Q}\hat{\mathbf{s}}_n + \lambda\int_\Omega \Delta u\,\Delta\hat{s} = \mathbf{u}_n^{\mathrm{T}}\mathbf{Q}(\mathbf{z} - \bar{\mathbf{f}}_n) - \lambda\int_\Omega \Delta u\,\Delta\bar{f} \tag{10}$$

for every $u \in H_0^2(\Omega)$.

For a proof, see Appendix A.4.

Likewise in Section 4 we derive a reformulation of the estimation problem that will constitute the base for the application of the finite element method. Denote by ∂_ν the normal derivative on the boundary of Ω, $\partial_\nu u := \nu \cdot \nabla u$, where ν is the outward normal to $\partial\Omega$. Appendix A.5 shows that, introducing an auxiliary function g and using integration by parts, the problem of finding $\hat{s}$ that satisfies condition (10) for every $u \in H_0^2(\Omega)$ may be formulated as follows: find $(\hat{s}, g) \in \{H_0^1(\Omega) \cap C^0(\Omega)\} \times H^1(\Omega)$ that satisfies

$$\begin{aligned} &\mathbf{u}_n^{\mathrm{T}}\mathbf{Q}\hat{\mathbf{s}}_n - \lambda\int_\Omega (\nabla u \cdot \nabla g) + \lambda\int_{\partial\Omega} (\partial_\nu u) g = \mathbf{u}_n^{\mathrm{T}}\mathbf{Q}(\mathbf{z} - \bar{\mathbf{f}}_n), \\ &\int_\Omega vg + \int_\Omega (\nabla v \cdot \nabla\hat{s}) - \int_{\partial\Omega} v(\partial_\nu\hat{s}) = -\int_\Omega (\nabla v \cdot \nabla\bar{f}) + \int_{\partial\Omega} v(\partial_\nu\bar{f}) \end{aligned} \tag{11}$$

for all $(u, v) \in \{H_0^1(\Omega) \cap C^0(\Omega)\} \times H^1(\Omega)$; moreover, such $\hat{s}$ belongs to $H_0^2(\Omega)$, and therefore $\hat{f} = \hat{s} + \bar{f} \in H^2(\Omega)$ with $\hat{f}|_{\partial\Omega} = f_{\partial\Omega}$. Corollary 2 gives the finite element solution to this problem.

Let m be the number of interior nodes. Denote by $\tilde{\mathbf{L}}$ the order m matrix that is obtained from $\mathbf{L}$ by removing the $K - m$ rows and columns corresponding to boundary nodes, and by $\tilde{\mathbf{R}}_1$ the

$m \times K$ matrix that is obtained from $\mathbf{R}_1$ by removing the $K-m$ rows corresponding to boundary nodes. Moreover, define the order K matrix $\mathbf{R}_\nu$,

$$\mathbf{R}_\nu := \int_{\partial\Omega_\mathcal{T}} (\nu_x \psi_x + \nu_y \psi_y) \psi^{\mathrm{T}},$$

and denote by $\tilde{\mathbf{R}}_\nu$ the $m \times K$ matrix that is obtained from $\mathbf{R}_\nu$ by removing the $K-m$ rows corresponding to boundary nodes. The matrix $\mathbf{R}_\nu$ is highly sparse, having non-zero entries only for couples of nodal basis associated with boundary elements.

Corollary 2. The estimators $\hat{\boldsymbol{\beta}} \in \mathbb{R}^q$ and $\hat{f} \in H^1_{\mathcal{T}}(\Omega)$ with $\hat{f} = f_{\partial\Omega}$ on all boundary nodes, that solve the discrete counterpart of the estimation problem with fixed value boundary conditions, exist unique:

(a) $\hat{\boldsymbol{\beta}} = (\mathbf{W}^{\mathrm{T}}\mathbf{W})^{-1}\mathbf{W}^{\mathrm{T}}(\mathbf{z} - \hat{\mathbf{f}}_n)$;

(b) $\hat{f} = \hat{s} + \bar{f}$, where $\bar{f}$ is the finite element function that coincides with $f_{\partial\Omega}$ on the $K-m$ boundary nodes and equals 0 on the m interior nodes, and $\hat{s}$ is the finite element function that equals 0 on the $K-m$ boundary nodes and takes values on the m interior nodes given by the vector $\hat{\tilde{\mathbf{s}}} \in \mathbb{R}^m$ satisfying

$$\begin{pmatrix} -\tilde{\mathbf{L}} & \lambda(\tilde{\mathbf{R}}_1 - \tilde{\mathbf{R}}_\nu) \\ \lambda(\tilde{\mathbf{R}}_1 - \tilde{\mathbf{R}}_\nu)^{\mathrm{T}} & \lambda \mathbf{R}_0 \end{pmatrix} \begin{pmatrix} \tilde{\mathbf{s}} \\ \mathbf{g} \end{pmatrix} = \begin{pmatrix} -\tilde{\mathbf{L}}\mathbf{z} \\ -\lambda(\mathbf{R}_1 - \mathbf{R}_\nu)^{\mathrm{T}} \bar{\mathbf{f}} \end{pmatrix}. \tag{12}$$

For a proof, see Appendix A.6.

We have considered the case where the values of the smoothing surface function are fixed along the whole boundary. This has been done only for clarity of explanation; with minor changes we can deal with mixed boundary conditions. For instance, we can specify fixed values over only a part of the boundary, say Γ_D, leaving the values of the solution free on the other part, say Γ_N, where we instead fix the value of normal derivatives (possibly also different from 0), as done for instance in Section 2 (here, $\Gamma_D \cup \Gamma_N = \partial\Omega$). From a computational point of view, this is a minor modification of what has been explained above. In the simulations and application that are presented in the following section, we shall look for estimates that satisfy mixed boundary conditions. In the numerical analysis terminology, fixing the values of the solution at the domain boundary is a *Dirichlet boundary condition*; fixing instead the values of the normal derivatives, controlling the flow through boundaries, is a *Neumann boundary condition*. Within our framework, besides the *mixed boundary conditions* that were mentioned above (with different conditions on different parts of the boundary), it is also straightforward to consider *Robin boundary conditions*, consisting in linear combinations of Dirichlet and Neumann conditions. As mentioned in Section 2, imposing a boundary condition on f ensures uniqueness of the solution. The soap film smoothing technique that was mentioned in Section 1 can deal with Dirichlet conditions over part of or the whole boundary.

6. Simulation studies and application

6.1. Simulations on C-domain

We shall here compare the performances of filtered kriging, method KRIG, thin plate splines, method TPS, soap film smoothing, method SOAP, and SSR models, in a simulation study on a C-shaped domain with the surface test function that was used by Wood *et al.* (2008), which is in turn a modification of that presented by Ramsay (2002). This function f, on its C-shaped domain, is shown in Fig. 3(a) (a three-dimensional image is shown in Fig. 4). With respect to Wood *et al.* (2008) we also include covariates in the simulation study. In particular, for $N = 50$

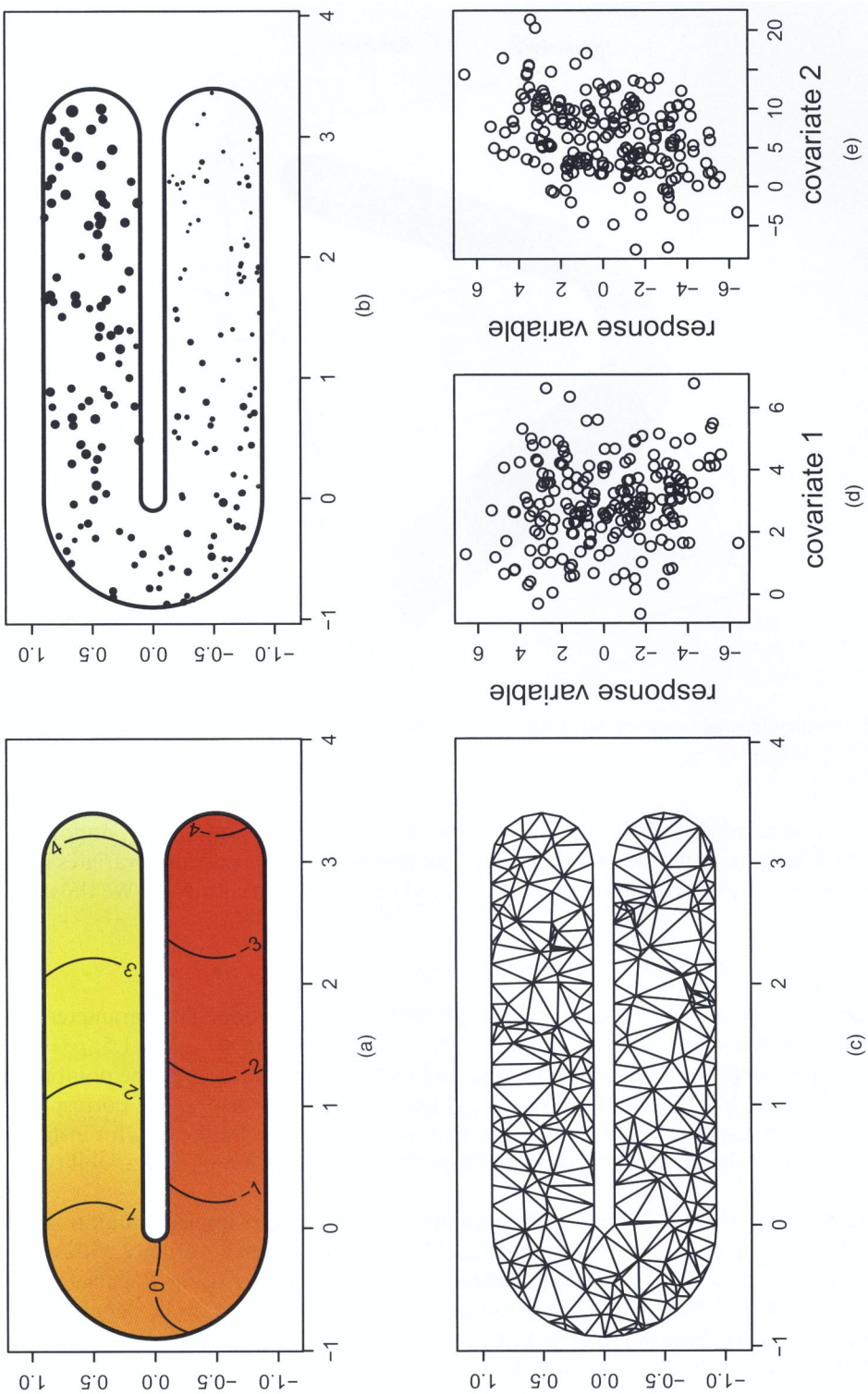

Fig. 3. (a) Colour map of the true function, (b) sampled data (the size of the point marker is proportional to the sampled data value), replicate 1, (c) domain triangulation, replicate 1, and (d), (e) scatter plots of the response variable *versus* covariates, replicate 1

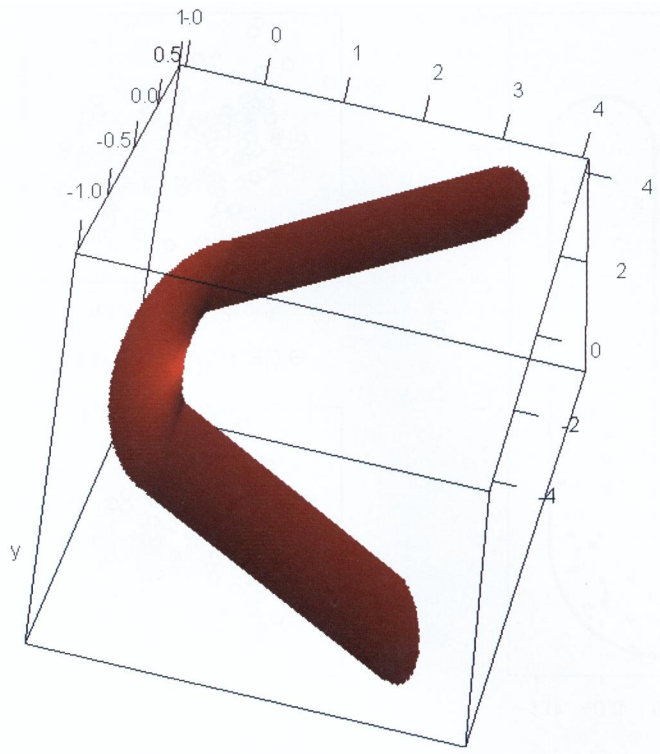

Fig. 4. Three-dimensional image of the surface function on a C-shaped domain (the corresponding colour map is shown in Fig. 3)

replicates, we simulate data as follows. We sample $n = 200$ locations, $\mathbf{p}_1, \ldots, \mathbf{p}_n$, uniformly on the C-shaped domain. Independently for each $\mathbf{p}_i$, we sample two independent covariates w_{i1} and w_{i2}, from an $N(\mu_1, \sigma_1^2)$ distribution and an $N(\mu_2, \sigma_2^2)$ distribution respectively. We thus obtain $z_1, \ldots, z_n$ from

$$z_i = \beta_1 w_{i1} + \beta_2 w_{i2} + f(\mathbf{p}_i) + \varepsilon_i, \qquad i = 1, \ldots, n, \tag{13}$$

where ε_i, $i = 1, \ldots, n$, are independent errors with $N(0, \sigma_\varepsilon^2)$ distribution. The parameter values that were used in the simulation are $\beta_1 = -0.5$, $\beta_2 = 0.2$, $\sigma_\varepsilon = 0.5$, $\mu_1 = 3$, $\sigma_1 = 1.5$, $\mu_2 = 7$ and $\sigma_2 = 5$. Fig. 3(b) shows the data sampled in the first replicate, with the size of the point marker proportional to the size of the sampled value. Figs 3(d) and 3(e) display the corresponding scatter plots of the response variable *versus* the two covariates. From these plots, for instance, it is not apparent that the first covariate is significant for the explanation of the variability of the response.

Method KRIG and TPS estimates are obtained under standard implementation settings, using the R package `fields` (see Furrer *et al.* (2010)), with the Matérn covariance with smoothness parameter $\nu = 1$ for KRIG. Method SOAP is implemented by using R package `soap` (see Wood (2010)) and uses 32 interior knots and a rank 39 (40-knot) cyclic penalized cubic regression spline as the boundary curve. Fig. 3(c) illustrates the triangulation that was used for SSR computations for the first replicate. The values of the smoothing parameters for the four methods are selected, for each replicate and method, by generalized cross-validation.

Table 1 shows the root-mean-square errors, over the 50 simulation replicates, of the estimate of the parameters β_1, β_2 and σ_ε provided by methods KRIG, TPS and SOAP and linear and quadratic SSR models (SSR1 and SSR2 respectively). The SOAP and SSR models give the best estimates of the parameters β_1 and β_2; SSR models give better estimates of σ_ε than any of the other methods considered. Fig. 5 shows the boxplots of the root-mean-squared error over the 50 simulation replicates, of the estimate of f, evaluated on a fine grid of step 0.02 over the C-domain; the boxplots highlight that the best estimates of f are provided by SSR models, followed by those given by SOAP; these two methods provide significantly better estimates than TPS and KRIG. Figs 6 and 7 illustrate the reason for this comparative advantage; they show respectively the estimated functions provided by the different methods in the first simulation replicate (Fig. 6), and the corresponding absolute residuals with respect to true f, always in the first replicate (Fig. 7). Fig. 7, in particular, shows that KRIG and TPS estimates have high absolute errors near the inner borders of the C-arms; this 'leakage effect' in KRIG and TPS estimates is because these two methods do not take into any account domain boundaries but instead smooth across them.

Table 1. Root-mean-squared error of the estimates of the trend parameters β_1 and β_2 and of the noise standard deviation σ_ε, obtained via filtered kriging, thin plate splines, soap film smoothing and linear and quadratic SSR models

Method	*Root-mean-squared errors of the following parameters:*		
	β_1	β_2	σ_ε
KRIG	0.0271	0.0098	0.0583
TPS	0.0247	0.0087	0.0612
SOAP	0.0232	0.0079	0.0314
SSR1	0.0232	0.0079	0.0289
SSR2	0.0232	0.0079	0.0289

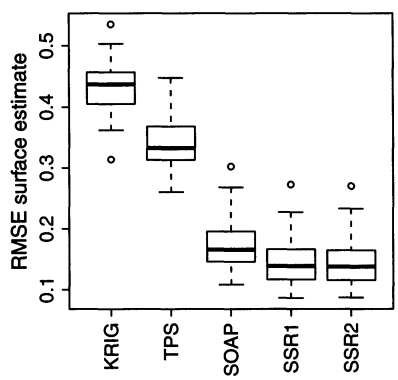

Fig. 5. Boxplots of the root-mean-squared error of the estimates of f, evaluated on a fine grid of step 0.02 over the C-domain, obtained by the various methods

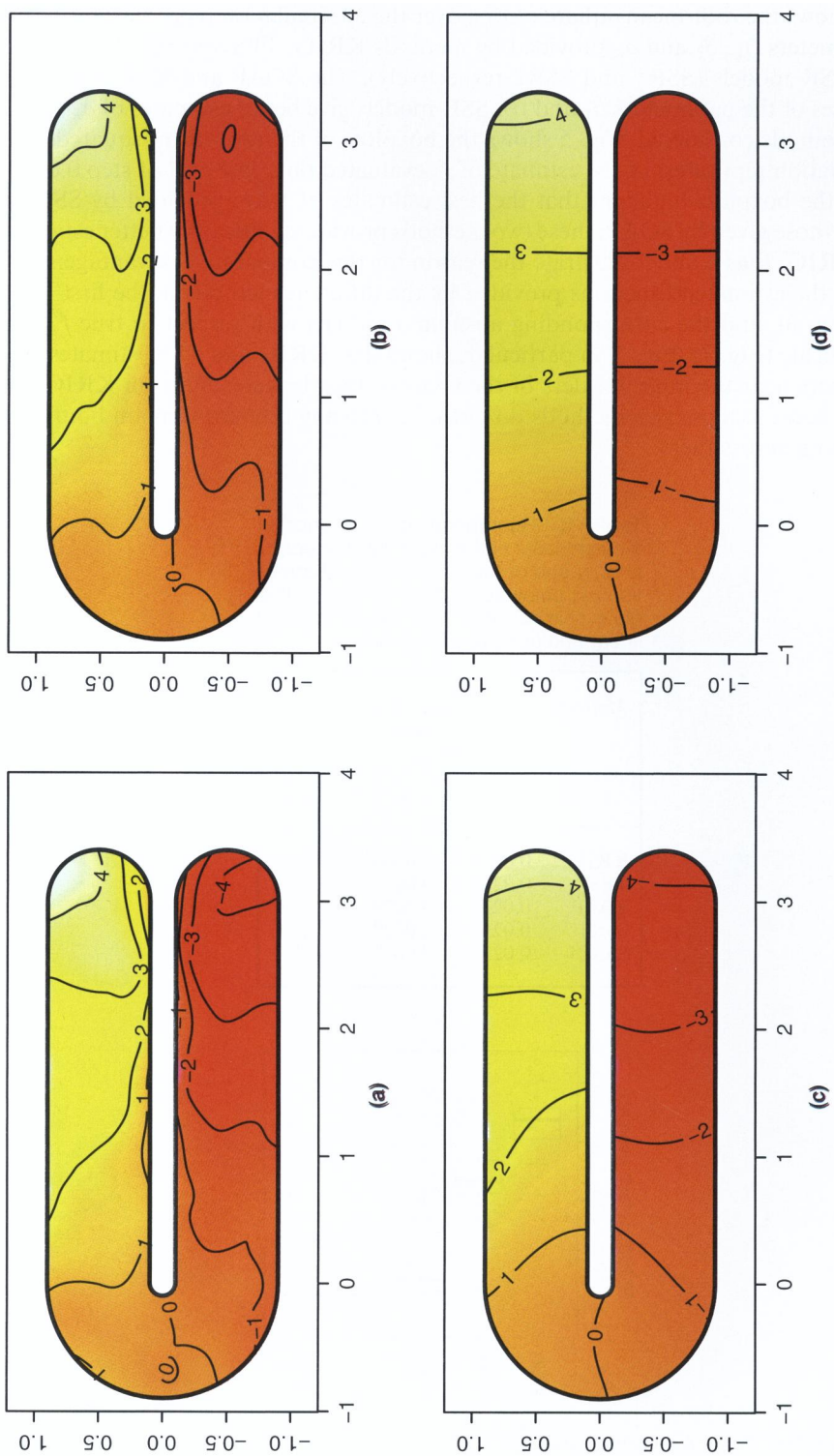

Fig. 6. Colour maps of f-estimates obtained in replicate 1 (the quadratic SSR is not displayed here as it is almost indistinguishable from the linear SSR): (a) KRIG; (b) TPS; (c) SOAP; (d) SSR1

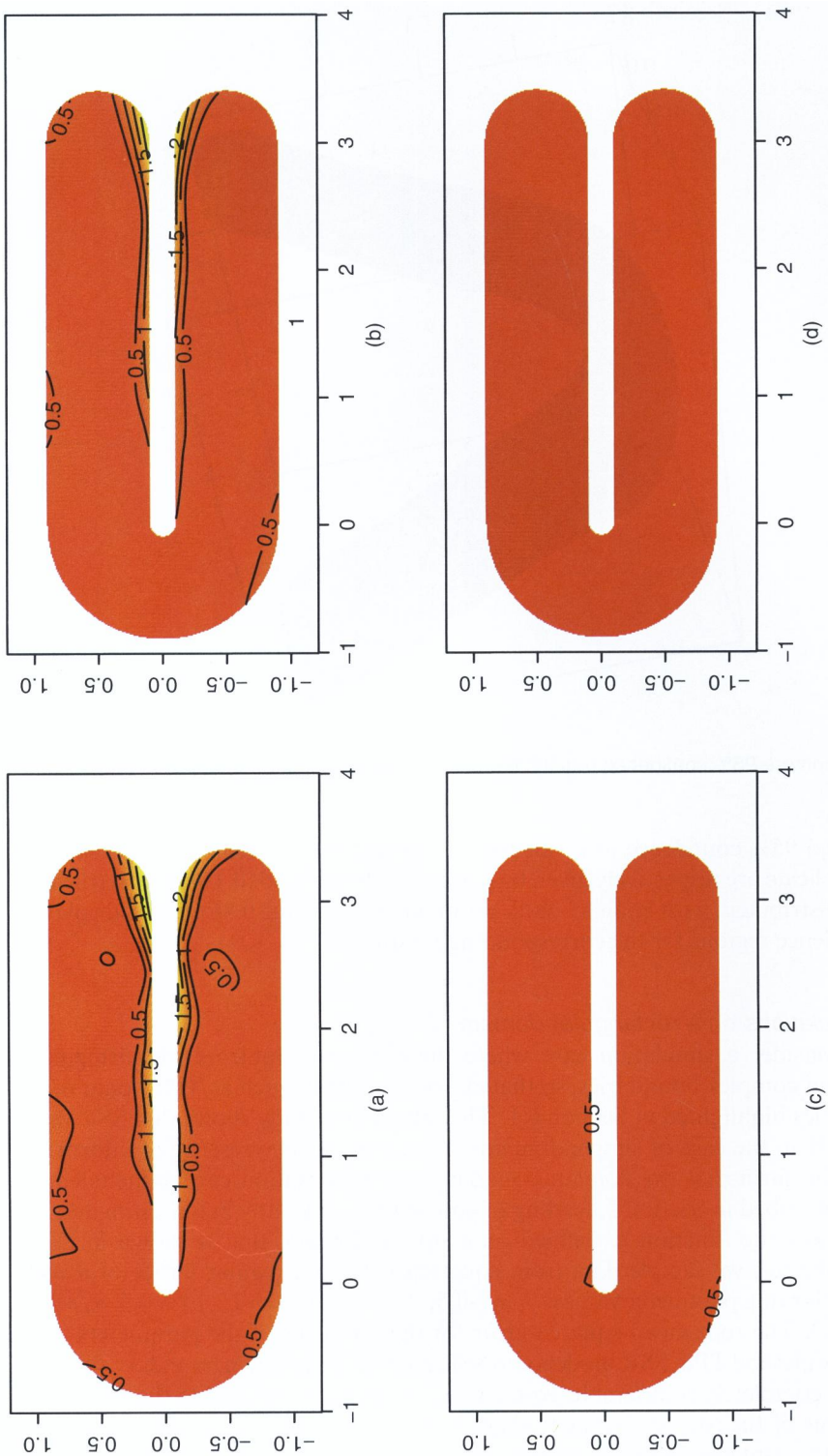

Fig. 7. Absolute residuals of f-estimates shown in Fig. 6, with respect to the true function f: (a) KRIG; (b) TPS; (c) SOAP; (d) SSR1

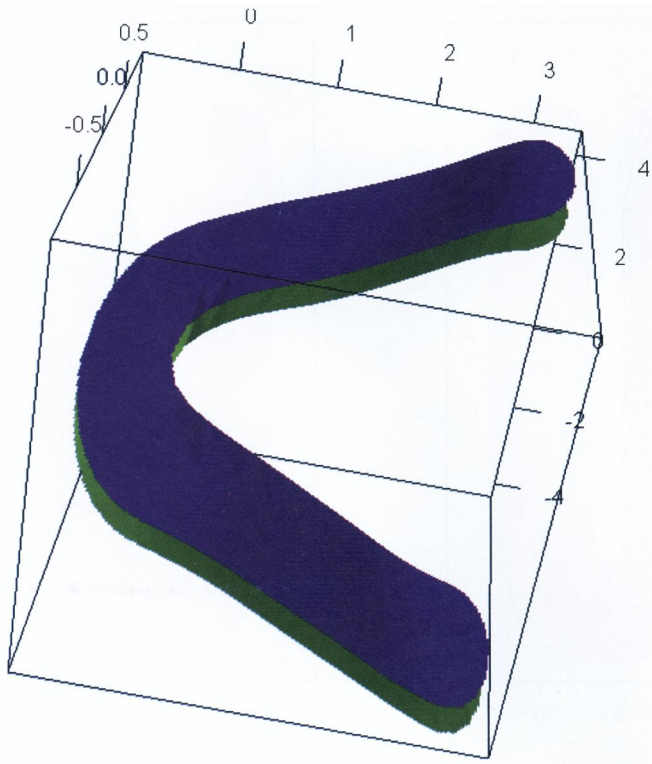

Fig. 8. Pointwise 95% confidence regions for the linear SSR estimate, replicate 1

Individual 95% confidence intervals for β_1 and β_2 for the linear SSR estimate obtained in the first replicate are respectively given by $[-0.521, -0.410]$ and $[0.171, 0.201]$ (the residuals are normally distributed, with Shapiro–Wilk normality test p-value 0.86). Fig. 8 displays pointwise 95% confidence regions for the corresponding f-estimate.

6.2. *Simulations on a rectangular domain*

We also consider a simulation case where there are no problems with irregularly shaped domains and complex boundaries, so that classical methods such as TPS would not encounter the difficulties highlighted in Section 6.1. This simulation study shows that SSR models are as good as TPS in the case of simple domains. Moreover, in the presence of information on the values of the surface at the domain boundary, this information can be included in the SSR model as described in Section 5, leading of course to significantly better estimates.

We consider the function f, defined on a square domain, that is shown in Fig. 9(a). For $N = 50$ replicates, we sample data from equation (13) on a regular lattice of $n = 400$ points, using the following parameter values: $\beta_1 = -0.5$, $\beta_2 = 0.25$, $\sigma_\varepsilon = 0.15$, $\mu_1 = 0.3$, $\sigma_1 = 0.5$, $\mu_2 = 0.7$ and $\sigma_2 = 1.3$. The root-mean-squared error for the estimates of the parameters β_1, β_2 and σ_ε obtained by method TPS, SSR models and SSR models with known boundary value conditions, SSRbc, are completely comparable, equal in fact when approximated to the third decimal digit. The boxplots of the root-mean-squared error of the estimate of f, evaluated on a fine grid of step 0.03, are displayed in Fig. 9(b): the estimates of f that are provided by TPS and SSR are

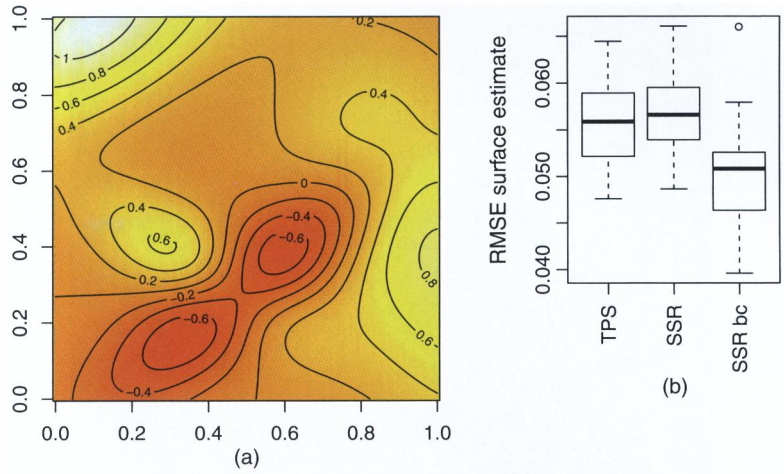

Fig. 9. (a) Test function on a simple domain and (b) boxplots of the root-mean-squared error of the estimate of the function f, evaluated on a fine grid of step 0.03 for TPS, SSR models and SSR with known boundary values (SSRbc)

comparable; SSRbc models efficiently use the extra information about known boundary values, resulting in significantly better estimates.

Finally, further simulations based on Wood *et al.* (2008) also demonstrate that SSR compares favourably with the SOAP method on functions with strong gradients at the boundary (see Sangalli *et al.* (2012)).

6.3. Application to Island of Montréal census data

We consider the problem of estimating population density over the Island of Montréal. The data are derived from the 1996 Canadian census. Fig. 1 shows the municipality of Montréal, along with 493 data points defined by the centroids of census enumeration areas. Population density is available at each census tract, measured as 1000 inhabitants per square kilometre. As covariate, we use the binary variable indicating whether a tract is predominantly residential (1) or commercial and industrial (0). Figs 1 and 10 highlight two areas that are not part of the domain of interest for the study of population density: the Montréal airport, with some surrounding services and industries, in the south-western part of the island, and the industrial park in the north-eastern tip, that includes an oil refinery tank farm. As mentioned in Section 1, it is known that no people live by the river banks in correspondence of the harbour (east shore) and the public parks (the south-west and north-west shore). We hence impose appropriate (i.e., in this case, homogeneous zero) Dirichlet boundary conditions along these stretches of coast, which are highlighted in red in Figs 1 and 10; along the remaining coasts we impose Neumann zero conditions, meaning no flow of population density across the shores. Fig. 11 shows the SSR estimate of the spatial variation structure: the non-parametric part of the model. Note that the estimate complies with the boundary conditions imposed, dropping to zero along the stretches of coast mentioned, thus efficiently including this *a priori* information. Also, the estimate has not artificially linked data points on either side of the uninhabited parts; see for instance the densely populated areas just in the south and west of the industrial park with respect to the low population density neighbourhood just north-east of it. The β-coefficient that corresponds to the binary covariate is estimated to be 1.300; this means that census tracts that are predominantly residential are on average expected to have 1300 more inhabitants per square kilometre with

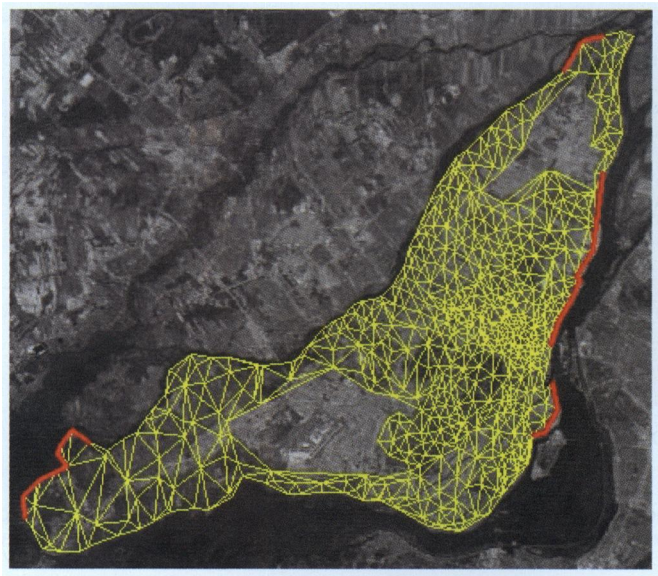

Fig. 10. Constrained Delaunay triangulation of the Island of Montréal

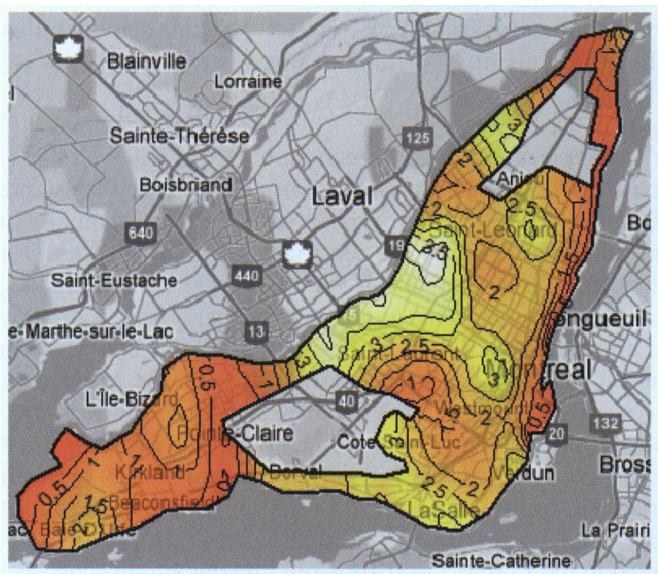

Fig. 11. SSR estimate of spatial structure for population density over the Island of Montréal

respect to those classified as mostly commercial or industrial; the approximate 95% confidence interval is given by [0.755; 1.845].

7. Discussion

The technique described can be generalized in several directions, for instance to areal data (see Azzimonti *et al.* (2012)), to general link functions such as the logit, to loss functions other

than the classical sum of squared errors, thus allowing for very large potential application. The covariates themselves, when having a spatial structure, could be modelled as surfaces, in a functional regression model setting.

An important line of investigation within the SSR framework concerns extensions of the roughness term to more complex partial differential operators; this model extension is particularly interesting for applications where some *a priori* knowledge of the problem (physical, chemical, mechanical or morphological) suggests the choice of a partial differential operator modelling to some extent the phenomenon under study. Azzimonti *et al.* (2012) investigates this research direction, generalizing SSR to penalties involving general second-order elliptic operators. The applied problem considered there concerns the estimation of the blood flow velocity field in a section of a carotid artery, using data provided by echo doppler images. In this application, physiological knowledge of the phenomenon suggests the theoretical shape of the velocity field; this knowledge is thus translated into a partial differential operator and used in the roughness penalty to regularize the estimate. This extension also constitutes a very promising line of research towards the modelling of spatiotemporal phenomena, including dynamical functional data, such as curves and surfaces deforming over time. Moreover, an approach that is analogous to that described here could also be applied to the problem of parameter estimation for partial differential equations, and we shall be exploring possible extensions in this direction of the work of Ramsay *et al.* (2007).

Finally, this model can also be extended to three dimensions, dealing with volumes and surfaces that are embedded in three-dimensional spaces. Such a model extension would have a strong influence on forefront applications, as for instance those concerning the analysis of three-dimensional images of the internal structures of a body provided by diagnostic medical scanners (angiographies, tomographies, magnetic resonance imaging devices, etc.). It is apparent that when analysing these data it would be in many contexts desirable to use techniques that take into account the boundary of the problem. For instance, organs such as the brain have complex boundaries, both external and internal, and when studying brain imaging data it would certainly be of great interest to be able to comply efficiently with those boundaries.

In future work we also intend to study the link between SSR and the model that was proposed by Lindgren *et al.* (2011), computed by using the integrated nested Laplace approximation method for direct Bayesian computations proposed by Rue *et al.* (2009).

The models proposed have been implemented in R (R Development Core Team, 2011) and MATLAB. Both versions will be released shortly.

Acknowledgements

L. Sangalli acknowledges funding by the Ministero dell'Istruzione dell'Università e della Ricerca, Fondo per gli Investimenti della Ricerca di Base *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/~sangalli/firb/`), and by the programme Dote Ricercatore Politecnico di Milano—Regione Lombardia, research project 'Functional data analysis for life sciences'. J. Ramsay acknowledges funding by a 'Discovery grant' from the Natural Sciences and Engineering Research Council of Canada. Part of this work was carried out when L. Sangalli and J. Ramsay were visiting the Statistical and Applied Mathematical Sciences Institute, Research Triangle Park, North Carolina, USA. We are grateful to Laura Azzimonti and Giancarlo Sangalli for help on the numerical analysis aspects of this work, and to Fabio Nobile and Piercesare Secchi for thoughtful and constructive comments. We also thank the Associate Editor and referees for helpful suggestions.

Appendix A

A.1. Proof of proposition 1

The functional $J_\lambda(\beta, f)$ in problem (2) can be rewritten as

$$J_\lambda(\beta, f) = (\mathbf{z} - \mathbf{W}\beta - \mathbf{f}_n)^T(\mathbf{z} - \mathbf{W}\beta - \mathbf{f}_n) + \lambda \int_\Omega (\Delta f)^2.$$

Given f, the unique minimizer $\hat{\beta}(f)$ of $J_\lambda(\beta, f)$ with respect to β is given by

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

Plugging $\hat{\beta}(f)$ into $J_\lambda(\beta, f)$ we obtain

$$J_\lambda\{\hat{\beta}(f), f\} = \mathbf{z}^T\mathbf{Q}\mathbf{z} - 2\mathbf{f}_n^T\mathbf{Q}\mathbf{z} + \mathbf{f}_n^T\mathbf{Q}\mathbf{f}_n + \lambda \int_\Omega (\Delta f)^2.$$

Our estimation problem is thus reduced to an optimization problem over f only: find $f \in H^2_{n0}(\Omega)$ that minimizes

$$\tilde{J}_\lambda(f) = \mathbf{f}_n^T\mathbf{Q}\mathbf{f}_n + \lambda \int_\Omega (\Delta f)^2 - 2\mathbf{f}_n^T\mathbf{Q}\mathbf{z} \tag{14}$$

where we have dropped the terms not depending on f.

The proof is completed by showing that $\hat{f}$, the minimizer of problem (14) over $H^2_{n0}(\Omega)$, is unique and satisfies condition (3) for every $u \in H^2_{n0}(\Omega)$. To prove this result, we exploit a characterization theorem (see, for example, Braess (2007), chapter 2) which states that if G is a symmetric, positive definite, bilinear form on a vector space V, and F is a linear functional on V, then v is the unique minimizer of

$$G(v, v) - 2F(v)$$

in V if and only if

$$G(v, u) = F(u) \qquad \text{for all } u \in V. \tag{15}$$

Moreover, there is at most one solution of problem (15).

The desired result follows immediately from application of the above theorem considering $V = H^2_{n0}(\Omega)$, the symmetric, positive definite, bilinear form

$$G(f, u) := \mathbf{u}_n^T\mathbf{Q}\mathbf{f}_n + \lambda \int_\Omega \Delta u \, \Delta f \tag{16}$$

and the linear functional $F(f) = \mathbf{f}_n^T\mathbf{Q}\mathbf{z}$.

Positive definitiveness of the form G in expression (16), on $H^2_{n0}(\Omega)$, is shown by the following argument. Suppose that $G(f, f) = 0$ for some $f \in H^2_{n0}(\Omega)$; then $\int_\Omega (\Delta f)^2 = 0$ and $\mathbf{f}_n^T\mathbf{Q}\mathbf{f}_n = 0$. The fact that $\int_\Omega (\Delta f)^2 = 0$, for $f \in H^2_{n0}(\Omega)$, implies that f is a constant function on Ω. Thus, let $f(\cdot) \equiv c$ on Ω, for some constant c. Then $\mathbf{f}_n^T\mathbf{Q}\mathbf{f}_n = c^2\mathbf{1}_n^T\mathbf{Q}\mathbf{1}_n$, where $\mathbf{1}_n$ is the n-vector having all entries equal to 1. Note that $\mathbf{1}_n^T\mathbf{Q}\mathbf{1}_n \neq 0$ because $\mathbf{1}_n$ does not belong to the subspace of $\mathbb{R}^n$ that is generated by the columns of $\mathbf{W}$, and hence the two vectors $\mathbf{1}_n$ and $\mathbf{Q}\mathbf{1}_n$ are not orthogonal. Therefore, $\mathbf{f}_n^T\mathbf{Q}\mathbf{f}_n = 0$ implies that $c = 0$, so $f(\cdot) \equiv 0$ on Ω, i.e. G is positive definite on $H^2_{n0}(\Omega)$. This concludes the proof.

A.2. Derivation of problem (5)

In this section we derive the reformulation of the estimation problem given in problem (5), that constitutes our base for the implementation of the finite element method.

The problem of finding $\hat{f} \in H^2_{n0}(\Omega)$ that satisfies condition (3) for every $u \in H^2_{n0}(\Omega)$ can be rewritten as the problem of finding $(\hat{f}, g) \in H^2_{n0}(\Omega) \times L^2(\Omega)$ that satisfy

$$\begin{aligned} \mathbf{u}_n^T\mathbf{Q}\hat{\mathbf{f}}_n + \lambda \int_\Omega g(\Delta u) &= \mathbf{u}_n^T\mathbf{Q}\mathbf{z}, \\ \int_\Omega gv - \int_\Omega (\Delta \hat{f})v &= 0 \end{aligned} \tag{17}$$

for all $(u, v) \in H^2_{n0}(\Omega) \times L^2(\Omega)$. In fact, if the pair of functions $(\hat{f}, g) \in H^2_{n0}(\Omega) \times L^2(\Omega)$ satisfies condition (17) for all $(u, v) \in H^2_{n0}(\Omega) \times L^2(\Omega)$, then $\hat{f}$ also satisfies problem (3). In contrast, if $\hat{f} \in H^2_{n0}(\Omega)$ satisfies problem (3), then the pair $(\hat{f}, \Delta\hat{f})$ automatically satisfies the two equations in problem (17). Now, asking a slightly higher regularity of the auxiliary function g and of the test functions v, namely $g, v \in H^1(\Omega)$ instead

of $g, v \in L^2(\Omega)$, this problem may in turn be reformulated as the problem of finding $(\hat{f}, g) \in \{H^1_{n0}(\Omega) \cap C^0(\Omega)\} \times H^1(\Omega)$ that satisfies condition (17) for all $(u, v) \in \{H^1_{n0}(\Omega) \cap C^0(\Omega)\} \times H^1(\Omega)$. Moreover, the theory of problems of elliptic regularity ensure that such $\hat{f}$ still belongs to $H^2_{n0}(\Omega)$ (see, for example, Lions and Magenes (1973), chapter 8). Reformulating the problem as a problem in $H^1(\Omega)$ is done here in view of its discretization via the finite element space $H^1_{\mathcal{T}}(\Omega)$. Finally, Green's theorem or integration by parts yields

$$\int_\Omega g(\Delta u) = -\int_\Omega (\nabla g \cdot \nabla u) + \int_{\partial\Omega} g(\partial_\nu u),$$
$$\int_\Omega (\Delta \hat{f}) v = -\int_\Omega (\nabla v \cdot \nabla \hat{f}) + \int_{\partial\Omega} v(\partial_\nu \hat{f})$$

where $\int_{\partial\Omega} g(\partial_\nu u) = 0$ and $\int_{\partial\Omega} v(\partial_\nu \hat{f}) = 0$ thanks to the boundary condition on the normal derivatives of f and u, hence implying that the latter reformulation is equivalent to that given in problem (5).

A.3. Proof of corollary 1

Problem (5) has the following discrete counterpart: find $(f, g) \in H^1_{\mathcal{T}}(\Omega) \times H^1_{\mathcal{T}}(\Omega)$ that satisfy problem (5) for all $(u, v) \in H^1_{\mathcal{T}}(\Omega) \times H^1_{\mathcal{T}}(\Omega)$, with the integrals now computed over $\Omega_{\mathcal{T}}$. Exploiting property (4), we have that for any functions f, g and v in $H^1_{\mathcal{T}}(\Omega)$

$$\int_{\Omega_{\mathcal{T}}} vg = \mathbf{v}^{\mathrm{T}} \mathbf{R}_0 \mathbf{g},$$
$$\int_{\Omega_{\mathcal{T}}} (\nabla v \cdot \nabla f) = \mathbf{v}^{\mathrm{T}} \mathbf{R}_1 \mathbf{f}.$$

Moreover, $\mathbf{u}_n^{\mathrm{T}} \mathbf{Q} \mathbf{f}_n = \mathbf{u}^{\mathrm{T}} \mathbf{L} \mathbf{f}$ and $\mathbf{u}_n^{\mathrm{T}} \mathbf{Q} \mathbf{z} = \mathbf{u}^{\mathrm{T}} \mathbf{L} \mathbf{D} \mathbf{z}$. Using these identities, solving the discrete counterpart of the estimation problem reduces to finding a couple of vectors $(\mathbf{f}, \mathbf{g}) \in \mathbb{R}^K \times \mathbb{R}^K$ that satisfy

$$\begin{aligned} \mathbf{u}^{\mathrm{T}} \mathbf{L} \mathbf{f} - \lambda \mathbf{u}^{\mathrm{T}} \mathbf{R}_1 \mathbf{g} &= \mathbf{u}^{\mathrm{T}} \mathbf{L} \mathbf{D} \mathbf{z}, \\ \mathbf{v}^{\mathrm{T}} \mathbf{R}_0 \mathbf{g} + \mathbf{v}^{\mathrm{T}} \mathbf{R}_1 \mathbf{f} &= 0 \end{aligned} \tag{18}$$

for all $(\mathbf{u}, \mathbf{v}) \in \mathbb{R}^K \times \mathbb{R}^K$. Since condition (18) must hold for all $(\mathbf{u}, \mathbf{v}) \in \mathbb{R}^K \times \mathbb{R}^K$, this is equivalent to finding $(\mathbf{f}, \mathbf{g}) \in \mathbb{R}^K \times \mathbb{R}^K$ that satisfies the linear system (6), that has been symmetrized for computational convenience. Solving this system (6) provides $\hat{\mathbf{f}}$ and, thanks to property (4), $\hat{\mathbf{f}}$ identifies the estimate $\hat{f} \in H^1_{\mathcal{T}}(\Omega)$. Since the system has a unique solution, the finite element solution to the estimation problem is unique. This can be shown working sequentially on the two equations composing the system. From the second equation, exploiting the positive definiteness and thus invertibility of the matrix $\mathbf{R}_0$ (which is proved below), we obtain $\mathbf{g} = -\mathbf{R}_0^{-1} \mathbf{R}_1 \mathbf{f}$. Then plugging this expression for $\mathbf{g}$ into the first equation composing the system, thanks to the fact that $\mathbf{L} + \lambda \mathbf{R}_1 \mathbf{R}_0^{-1} \mathbf{R}_1$ is also invertible (as proved below), we obtain the unique solution $\hat{\mathbf{f}} = (\mathbf{L} + \lambda \mathbf{R}_1 \mathbf{R}_0^{-1} \mathbf{R}_1)^{-1} \mathbf{L} \mathbf{D} \mathbf{z}$.

Positive definiteness of $\mathbf{R}_0$ follows immediately from the fact that $\mathbf{R}_0$ is the Gramm matrix that is associated with the set of K linearly independent vectors $(\psi_k(\boldsymbol{\xi}_1), \ldots, \psi_k(\boldsymbol{\xi}_K))^{\mathrm{T}}$, for $k = 1, \ldots, K$, in the inner product space $(H^1_{\mathcal{T}}(\Omega), \langle v_1, v_2 \rangle)$, where $\langle v_1, v_2 \rangle = \int_\Omega v_1 v_2$.

Positive definiteness of $\mathbf{L} + \lambda \mathbf{R}_1 \mathbf{R}_0^{-1} \mathbf{R}_1$ is proved by the following argument. Since $\mathbf{L}$ is positive semidefinite by construction, and $\mathbf{R}_0$, and thus $\mathbf{R}_0^{-1}$, is positive definite, then $\mathbf{L} + \lambda \mathbf{R}_1 \mathbf{R}_0^{-1} \mathbf{R}_1$ is at least positive semidefinite. Suppose now that $\mathbf{c}^{\mathrm{T}} (\mathbf{L} + \lambda \mathbf{R}_1 \mathbf{R}_0^{-1} \mathbf{R}_1) \mathbf{c} = 0$ for some K-vector $\mathbf{c}$. Since

$$0 = \mathbf{c}^{\mathrm{T}} (\mathbf{L} + \lambda \mathbf{R}_1 \mathbf{R}_0^{-1} \mathbf{R}_1) \mathbf{c} = \mathbf{c}^{\mathrm{T}} \mathbf{L} \mathbf{c} + \lambda \mathbf{c}^{\mathrm{T}} \mathbf{R}_1 \mathbf{R}_0^{-1} \mathbf{R}_1 \mathbf{c}$$

and both terms on the right-hand side are non-negative, it follows that $\mathbf{c}^{\mathrm{T}} \mathbf{L} \mathbf{c} = 0$ and $\mathbf{c}^{\mathrm{T}} \mathbf{R}_1 \mathbf{R}_0^{-1} \mathbf{R}_1 \mathbf{c} = 0$. Being $\mathbf{R}_0^{-1}$ positive definite, the latter implies that $\mathbf{R}_1 \mathbf{c} = \mathbf{0}$, and this in turn implies that $\mathbf{c}^{\mathrm{T}} \mathbf{R}_1 \mathbf{c} = 0$. Hence

$$\begin{aligned} 0 = \mathbf{c}^{\mathrm{T}} \mathbf{R}_1 \mathbf{c} &= \mathbf{c}^{\mathrm{T}} \left\{ \int_\Omega (\psi_x \psi_x^{\mathrm{T}} + \psi_y \psi_y^{\mathrm{T}}) \right\} \mathbf{c} = \int_\Omega \{ \mathbf{c}^{\mathrm{T}} (\psi_x \psi_x^{\mathrm{T}} + \psi_y \psi_y^{\mathrm{T}}) \mathbf{c} \} \\ &= \| \mathbf{c}^{\mathrm{T}} \psi_x \|_2^2 + \| \mathbf{c}^{\mathrm{T}} \psi_y \|_2^2 \end{aligned}$$

where $\| \cdot \|_2$ denotes the L^2-norm, and this implies that $\| \mathbf{c}^{\mathrm{T}} \psi_x \|_2^2 = 0$ and $\| \mathbf{c}^{\mathrm{T}} \psi_y \|_2^2 = 0$. Thus, both partial derivatives of the piecewise quadratic finite element function $\mathbf{c}^{\mathrm{T}} \psi$ vanish, and this means that $\mathbf{c}^{\mathrm{T}} \psi$ is a constant function. Since the entries of the vector $\mathbf{c}$ are the values of the function $\mathbf{c}^{\mathrm{T}} \psi$ evaluated at the K nodes, $\mathbf{c}$ must have the form $(c, \ldots, c)^{\mathrm{T}}$ for some real constant c, i.e. $\mathbf{c} = c \mathbf{1}_k$. Then, $\mathbf{c}^{\mathrm{T}} \mathbf{L} \mathbf{c} = c^2 \mathbf{1}_n \mathbf{Q} \mathbf{1}_n$.

Exploiting this identity, the fact that $\mathbf{c}^{\mathrm{T}}\mathbf{L}\,\mathbf{c}=0$ implies that $c=0$, since $\mathbf{1}_n\,\mathbf{Q}\,\mathbf{1}_n\neq 0$, as already noted in Appendix A.1. It follows that $\mathbf{c}=\mathbf{0}$, and this proves that the matrix $\mathbf{L}+\lambda\mathbf{R}_1\mathbf{R}_0^{-1}\mathbf{R}_1$ is positive definite.

A.4. Proof of proposition 2

Likewise, in the proof of proposition 1, we reduce the estimation problem to an optimization problem over f only: find f that minimizes the functional $\tilde{J}_\lambda(f)$ in expression (14) over $f\in H^2(\Omega)$ with $f|_{\partial\Omega}=f_{\partial\Omega}$. Now fix $\bar{f}$ to be any function in $H^2(\Omega)$ such that $\bar{f}|_{\partial\Omega}=f_{\partial\Omega}$ and $\bar{f}$ has known gradient $\nabla\bar{f}$ and Laplacian $\Delta\bar{f}$. If $f_{\partial\Omega}$ is sufficiently regular, e.g. $f_{\partial\Omega}\in H^{3/2}(\partial\Omega)$, this function exists. Moreover, let s be the function in $H_0^2(\Omega)$ such that $f=s+\bar{f}$. Holding $\bar{f}$ fixed, our optimization problem becomes finding $s\in H_0^2(\Omega)$ that minimizes

$$T_\lambda(s)=(\mathbf{s}_n^{\mathrm{T}}+\bar{\mathbf{f}}_n^{\mathrm{T}})\,\mathbf{Q}\,(\mathbf{s}_n+\bar{\mathbf{f}}_n)+\lambda\int_\Omega(\Delta s+\Delta\bar{f})^2-2(\mathbf{s}_n^{\mathrm{T}}+\bar{\mathbf{f}}_n^{\mathrm{T}})\mathbf{Q}\mathbf{z}.$$

This is in turn equivalent to finding $s\in H_0^2(\Omega)$ that minimizes

$$\tilde{T}_\lambda(s)=\mathbf{s}_n^{\mathrm{T}}\mathbf{Q}\mathbf{s}_n+\lambda\int_\Omega(\Delta s)^2-2\left\{\mathbf{s}_n^{\mathrm{T}}\mathbf{Q}(\mathbf{z}-\bar{\mathbf{f}}_n)-\lambda\int_\Omega\Delta s\,\Delta\bar{f}\right\} \tag{19}$$

where we have dropped the terms that are constant with respect to s. To complete the proof, we need to show that the function $\hat{s}$ is the unique minimizer over $H_0^2(\Omega)$ of $\tilde{T}_\lambda(s)$ in problem (19) if and only if it satisfies condition (10) for every $u\in H_0^2(\Omega)$. This result is proved resorting to the characterization theorem that has already been evoked in Appendix A.1 and considering the vector space $V=H_0^2(\Omega)$, the symmetric positive definite bilinear form $G(s,u):=\mathbf{u}_n^{\mathrm{T}}\,\mathbf{Q}\,\mathbf{s}_n+\lambda\int_\Omega\Delta u\,\Delta s$ and the linear functional $F(s)=\mathbf{s}_n^{\mathrm{T}}\,\mathbf{Q}\,(\mathbf{z}-\bar{\mathbf{f}}_n)-\lambda\int_\Omega\Delta s\,\Delta\bar{f}$.

A.5. Derivation of problem (11)

Likewise, in Appendix A.2, the problem of finding $s\in H_0^2(\Omega)$ that satisfies condition (10) for every $u\in H_0^2(\Omega)$ can be rewritten as follows: find $(\hat{s},g)\in H_0^2(\Omega)\times L^2(\Omega)$ that satisfies

$$\begin{aligned}\mathbf{u}_n^{\mathrm{T}}\,\mathbf{Q}\,\hat{\mathbf{s}}_n+\lambda\int_\Omega g\Delta u&=\mathbf{u}_n^{\mathrm{T}}\mathbf{Q}(\mathbf{z}-\bar{\mathbf{f}}_n),\\ \int_\Omega gv-\int_\Omega(\Delta\hat{s}+\Delta\bar{f})v&=0\end{aligned} \tag{20}$$

for all $(u,v)\in H_0^2(\Omega)\times L^2(\Omega)$. Reasoning as in Appendix A.2, we hence note that, asking a slightly higher regularity of the auxiliary function g and of the test functions v, namely $g,v\in H^1(\Omega)$ instead of $g,v\in L^2(\Omega)$, the problem above may in turn be reformulated as the problem of finding $(\hat{s},g)\in\{H_0^1(\Omega)\cap C^0(\Omega)\}\times H^1(\Omega)$ that satisfies condition (20) for all $(u,v)\in\{H_0^1(\Omega)\cap C^0(\Omega)\}\times H^1(\Omega)$; moreover, the theory of problems of elliptic regularity ensures that such $\hat{s}$ still belongs to $H_0^2(\Omega)$. Also, in this case, this is done in view of the discretization of the problem via the finite element space $H_{\mathcal{T}}^1(\Omega)$. Finally, using Green's theorem as in Appendix A.2 now implies that the latter reformulation is equivalent to that given in problem (11).

A.6. Proof of corollary 2

It is convenient to choose $\bar{f}$ to be the finite element function that coincides with $f_{\partial\Omega}$ on the boundary nodes and is 0 on the m interior nodes; recall that this is a datum in our system. Problem (11) has the following discrete counterpart: find $(\hat{s},g)\in H_{0,\mathcal{T}}^1(\Omega)\times H_{\mathcal{T}}^1(\Omega)$ that satisfies condition (11) for all $(u,v)\in H_{0,\mathcal{T}}^1(\Omega)\times H_{\mathcal{T}}^1(\Omega)$, with the integrals now computed over $\Omega_{\mathcal{T}}$. Here $H_{0,\mathcal{T}}^1(\Omega)$ is the subspace of $H_{\mathcal{T}}^1(\Omega)$ consisting of the finite element functions that equal zero on boundary nodes; this space is spanned by the m nodal basis functions corresponding to m interior nodes. The finite element functions $s,u\in H_{0,\mathcal{T}}^1(\Omega)$ are thus identified by vectors of $\mathbb{R}^m$ having as entries the values of these functions at the m interior nodes. Hence, the problem above reduces to finding a couple of vectors $(\tilde{\mathbf{s}},\mathbf{g})\in\mathbb{R}^m\times\mathbb{R}^K$ that satisfies

$$\begin{aligned}\mathbf{u}^{\mathrm{T}}\tilde{\mathbf{L}}\,\tilde{\mathbf{s}}-\lambda\,\mathbf{u}^{\mathrm{T}}\tilde{K}_1\,\mathbf{g}+\lambda\,\mathbf{u}^{\mathrm{T}}\tilde{K}_\nu\,\mathbf{g}&=\mathbf{u}^{\mathrm{T}}\tilde{\mathbf{L}}\mathbf{z},\\ \mathbf{v}^{\mathrm{T}}\mathbf{R}_0\,\mathbf{g}+\mathbf{v}^{\mathrm{T}}\tilde{K}_1^{\mathrm{T}}\,\tilde{\mathbf{s}}-\mathbf{v}^{\mathrm{T}}\tilde{K}_\nu^{\mathrm{T}}\,\tilde{\mathbf{s}}&=-\mathbf{v}^{\mathrm{T}}\mathbf{R}_1\,\bar{\mathbf{f}}+\mathbf{v}^{\mathrm{T}}\,\mathbf{R}_\nu^{\mathrm{T}}\,\bar{\mathbf{f}}\end{aligned} \tag{21}$$

for all $(\mathbf{u}, \mathbf{v}) \in \mathbb{R}^m \times \mathbb{R}^K$. Since condition (21) must hold for all $(\mathbf{u}, \mathbf{v}) \in \mathbb{R}^m \times \mathbb{R}^K$, this is equivalent to finding $(\tilde{\mathbf{s}}, \mathbf{g}) \in \mathbb{R}^m \times \mathbb{R}^K$ satisfying the linear system (12). Solving the system provides the solution $\tilde{\tilde{\mathbf{s}}} \in \mathbb{R}^m$ that identifies a finite element function $\hat{s} \in H^1_{0,\mathcal{I}}(\Omega)$. The estimate $\hat{f}$, having the required fixed values at the domain boundary, is thus given by $\hat{f} = \hat{s} + f$.

References

Azzimonti, L., Sangalli, L. M., Secchi, P., Romagnoli, S. and Domanin, M. (2012) PDE penalization for spatial fields smoothing. In *Proc. 46th Scient. Meet. Italian Statistical Society*. (Available from `http://meetings.sis-statistica.org`.)

Braess, D. (2007) Theory, fast solvers, and applications in elasticity theory. In *Finite Elements*, 3rd edn. Cambridge: Cambridge University Press.

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

Furrer, R., Nychka, D. and Sain, S. (2010) fields: tools for spatial data. *R Package Version 6.3*.

Gockenbach, M. S. (2006) *Understanding and Implementing the Finite Element Method*. Philadelphia: Society for Industrial and Applied Mathematics.

Guillas, S. and Lai, M.-J. (2010) Bivariate splines for spatial functional regression models. *J. Nonparam. Statist.*, **22**, 477–497.

Hjelle, Ø. and Dæhlen, M. (2006) *Triangulations and Applications; Mathematics and Visualization*. Berlin: Springer.

Lai, M.-J. and Schumaker, L. L. (2007) *Spline Functions on Triangulations*. Cambridge: Cambridge University Press.

Lindgren, F., Rue, H. and Lindström, J. (2011) An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach (with discussion). *J. R. Statist. Soc.* B, **73**, 423–498.

Lions, J.-L. and Magenes, E. (1973) *Non-homogeneous Boundary Value Problems and Applications*, vol. III. New York: Springer.

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

Quarteroni, A. (2009) *Numerical Models for Differential Problems, Modeling, Simulation and Applications*, vol. 2. Milan: Springer Italia.

Ramsay, T. (2002) Spline smoothing over difficult regions. *J. R. Statist. Soc.* B, **64**, 307–319.

Ramsay, J. O., Hooker, G., Campbell, D. and Cao, J. (2007) Parameter estimation for differential equations: a generalized smoothing approach (with discussion). *J. R. Statist. Soc.* B, **69**, 741–796.

Ramsay, J. O. and Silverman, B. W. (2005) *Functional Data Analysis*, 2nd edn. New York: Springer.

R Development Core Team (2011) *R: a Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Rue, H., Martino, S. and Chopin, R. (2009) Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations (with discussion). *J. R. Statist. Soc.* B, **71**, 319–392.

Sangalli, L., Ramsay, J. and Ramsay, T. (2012) Spatial spline regression models. *Technical Report 08/2012*. Dipartimento di Matematica, Politecnico di Milano, Milan. (Available from `http://mox.polimi.it/it/progetti/pubblicazioni`.)

Wood, S. (2010) soap: soap film smoothing. *R Package Version 0.1-4*.

Wood, S. N., Bravington, M. V. and Hedley, S. L. (2008) Soap film smoothing. *J. R. Statist. Soc.* B, **70**, 931–955.