

Published on *Stochastic Environmental Research and Risk Assessment* (2016). DOI: 10.1007/s00477-016-1237-3.  
 Publisher's version available at <http://link.springer.com/article/10.1007/s00477-016-1237-3>

# A penalized regression model for spatial functional data with application to the analysis of the production of waste in Venice province

Mara S. Bernardi · Laura M. Sangalli · Gabriele Mazza · James O. Ramsay

**Abstract** We propose a method for the analysis of functional data with complex dependencies, such as spatially dependent curves or time dependent surfaces, over highly textured domains. The models are based on the idea of regression with partial differential regularizations. In particular, we consider here two roughness penalties that account separately for the regularity of the field in space and in time. Among the various modelling features, the proposed method is able to deal with spatial domains featuring peninsulas, islands and other complex geometries. Space-time varying covariate information is included in the model via a semi-parametric framework. The proposed method is compared via simulation studies to other spatio-temporal techniques and it is applied to the analysis of the annual production of waste in the towns of Venice province.

**Keywords** Space-time model · Differential regularization · Finite elements

## 1 Introduction

In this work we deal with spatio-temporal data distributed over a spatial domain which presents complex geometries. That is, the irregular shape of the domain influences the phenomenon under study and there are important geographical

elements within the boundary such as islands and peninsulas that impact the distribution of the data. We refer to such domains as textured.

As an illustrative example, consider the estimation of the temporal evolution of the amount of per capita municipal waste produced in the towns of Venice province. Figure 1 shows the Venice province, with dots indicating town centers, including municipalities and other tourist localities of particular relevance. The province boundary is shown by a red line, highlighting the irregular shape of the province administrative borders and its complex coastlines, with the Venice lagoon partly enclosed by elongated peninsulas and small islands.

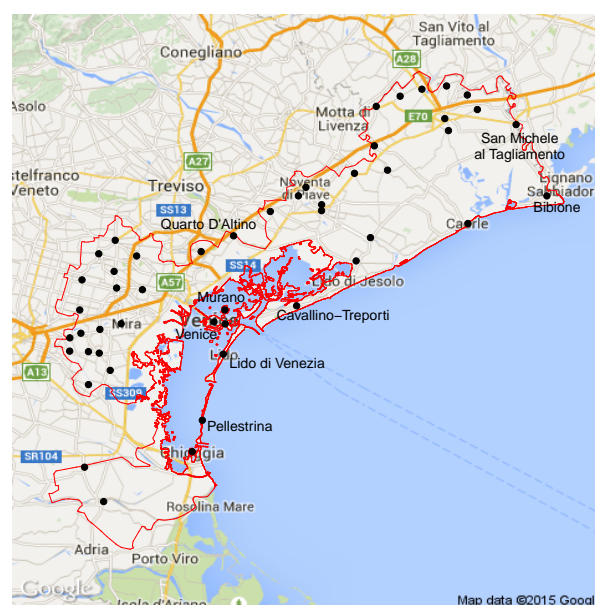


Fig. 1: Spatial domain of the Venice waste data, with a red line highlighting the province boundary and dots indicating the towns centers.

M. S. Bernardi  
 Politecnico di Milano, Milano, Italy

L. M. Sangalli (✉)  
 MOX, Dipartimento di Matematica, Politecnico di Milano,  
 Piazza L. da Vinci 32, Milano 20133, Italy.  
 E-mail: [laura.sangalli@polimi.it](mailto:laura.sangalli@polimi.it)

G. Mazza  
 Politecnico di Milano, Milano, Italy

J. Ramsay  
 McGill University, Montreal, Canada

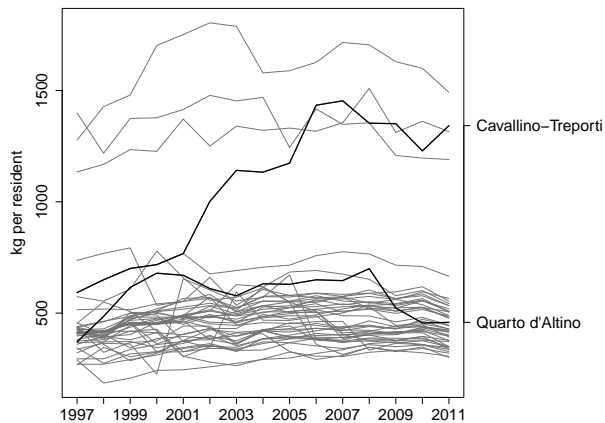


Fig. 2: Temporal evolution of the yearly per capita production (kg per resident) of municipal waste in the towns of Venice province.

The data are measurements from 1997 to 2011 of the yearly amount of per capita municipal waste (total kg divided by the number of municipality residents) and are provided by the Arpav, the Agenzia regionale per la prevenzione e protezione ambientale del Veneto.

Figure 2 shows the temporal evolution of the production of per capita waste in the towns of Venice province and Figure 3 is a bubble plot of the data at a fixed year, 2006. The phenomenon portrayed by these data is expressed differently in different parts of the domain. Consider for instance the two towns of Cavallino-Treporti (in the peninsula at the north-east of Venice) and Quarto d'Altino (north of Venice), indicated by black dots in Figure 3. The temporal evolution of the production of per capita municipal waste in the two towns, highlighted in Figure 2, is rather different, with strongly increasing and high values in the seaside and tourist town of Cavallino-Treporti, opposed to the not increasing and lower values measured in hinterland town of Quarto d'Altino. These two towns are close in terms of their geodesic distance, but they are actually separated by the Venice lagoon. Hence, appropriately accounting for the shape of the domain, characterized for instance by a strong concavity formed by the lagoon, is crucial to accurately handle these data.

When analyzing the temporal evolutions of the amount of per capita municipal waste, we shall make a strong simplification of the nature of these data, and consider them in the framework of geostatistical functional data (Delicado et al, 2010), where the datum is observable in principle in any point of the domain, instead of in the framework of functional areal data. As detailed in Section 7, this is due to the fact that we miss the information concerning the ur-

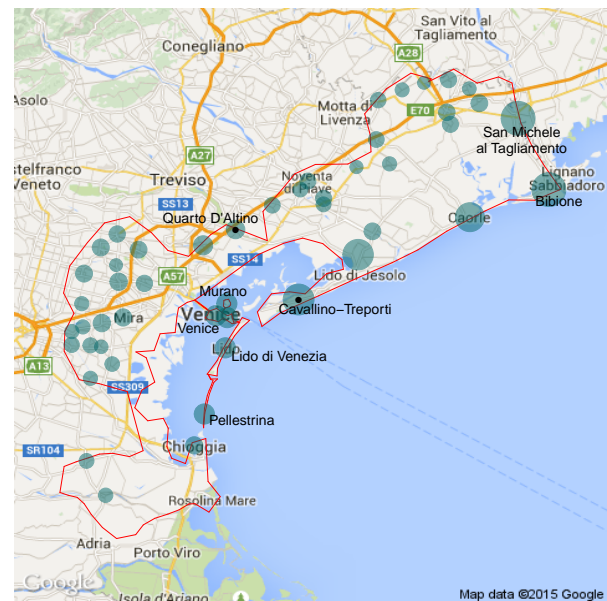


Fig. 3: Per capita production (kg per resident) of municipal waste in the towns of Venice province in 2006. The data include all municipalities of Venice province and additional four localities (Bibione, Murano, Lido di Venezia and Pellestrina), that do not constitute a municipality on their own, but have been included due to their tourist relevance and their location on the domain. For these additional four localities, the considered datum is a replicate of the datum of their corresponding municipalities (see Section 7).

banized areas of the municipalities, where the type of waste here considered (that does not include agricultural, industrial, construction/demolition and hazardous waste) is produced.

Various methods have been recently proposed for the analysis of spatially dependent functional data. Starting from the pioneering work of Goulard and Voltz (1993), kriging prediction methods for stationary spatial functional data are developed in Delicado et al (2010), Nerini et al (2010) and Giraldo et al (2011). Recent techniques developing universal kriging approaches for spatially dependent functional data are offered by Caballero et al (2013), Menafoglio et al (2013) and Menafoglio et al (2014). An extension of kriging for functional data which takes into account the presence of covariates is developed in Ignaccolo et al (2014). The same data can be also considered in a more classical space-time data framework. An extensive literature on spatio-temporal models has been produced; we refer the reader to Cressie and Wikle (2011) and references therein. On the other hand, these methods are not well suited for the context we are here considering because they do not take into account the shape of the domain; for instance these methods would smooth across concave boundary regions, thus closely linking data points that are in fact far apart by land connections.

Recent methods for the analysis of spatio-temporal data that instead specifically account for the geometry of the domain of interest are described in Augustin et al (2013) and Marra et al (2012). These models are based on the spatial smoother proposed by Wood et al (2008). Here, we extend the spatial models with differential regularization described in Ramsay (2002), Sangalli et al (2013) and Azzimonti et al (2015) to time dependent data, and propose a Spatio-Temporal regression model with Partial Differential Equations regularization (ST-PDE). The model is implemented in R (R Core Team, 2014), based on the package fdaPDE (Lila et al, 2016).

The paper is organized as follows. Section 2 describes the ST-PDE model. Section 3 shows the numerical implementation of the model. Section 4 illustrates the extension of the model for the inclusion of space-time varying covariates. Section 5 describes a variation of the ST-PDE model. Section 6 compares via simulation studies the ST-PDE model to other spatio-temporal prediction techniques. Section 7 shows the application of the proposed method to the analysis of the per capita municipal waste in the Venice province. Section 8 outlines some possible model extensions.

## 2 Data and model

Let  $\{\mathbf{p}_i = (x_i, y_i); i = 1, \dots, n\}$  be a set of  $n$  spatial points on a bounded domain  $\Omega \subset \mathbb{R}^2$ , whose boundary  $\partial\Omega$  is a curve of class  $\mathcal{C}^2$ , and  $\{t_j; j = 1, \dots, m\}$  be a set of  $m$  time instants in a time interval  $[T_1, T_2] \subset \mathbb{R}$ . Let  $z_{ij}$  be the value of a real-valued variable observed at point  $\mathbf{p}_i$  and time  $t_j$ . In our illustrative application, the spatial domain  $\Omega$  is the province of Venice, the spatial locations  $\mathbf{p}_i$  are the centers of the towns, the time instants  $t_j$  are the years between 1997 and 2011 and the variable of interest  $z_{ij}$  is the amount of per capita municipal waste produced in the town  $i$  and year  $t_j$ . The data  $z_{ij}$  are a sampling of space dependent temporal curves. Equivalently, they can be seen as a sampling of time dependent surfaces on  $\Omega$ .

We assume that  $\{z_{ij}; i = 1, \dots, n; j = 1, \dots, m\}$  are noisy observations of an underlying spatio-temporal smooth function  $f(\mathbf{p}, t)$ :

$$z_{ij} = f(\mathbf{p}_i, t_j) + \varepsilon_{ij} \quad i = 1, \dots, n, \quad j = 1, \dots, m, \quad (1)$$

where  $\{\varepsilon_{ij}; i = 1, \dots, n; j = 1, \dots, m\}$  are independently distributed residuals with mean zero and constant variance  $\sigma^2$ .

We estimate  $f(\mathbf{p}, t)$  by minimizing a penalized sum of square error functional  $J(f)$ , where the penalization takes into account separately the regularity of the function in the spatial and temporal domains. Various choices for the regularizing terms in space and in time are possible. In this work, we use simple isotropic and stationary regularizing terms in both space and time. In particular, following Ramsay (2002),

Wood et al (2008) and Sangalli et al (2013), we use the spatial roughness penalty

$$J_S(g(\mathbf{p})) = \int_{\Omega} (\Delta g(\mathbf{p}))^2 d\mathbf{p}, \quad (2)$$

where  $g: \Omega \rightarrow \mathbb{R}$  and the Laplacian  $\Delta g(\mathbf{p}) = \frac{\partial^2 g}{\partial x^2}(\mathbf{p}) + \frac{\partial^2 g}{\partial y^2}(\mathbf{p})$  provides a simple measure of the local curvature of  $g$ . Other possible choices for spatial roughness penalties are, for instance, that associated with thin plate splines, given by  $\int_{\mathbb{R}^2} (\frac{\partial^2 g}{\partial x^2}(\mathbf{p}))^2 + 2(\frac{\partial^2 g}{\partial x \partial y}(\mathbf{p}))^2 + (\frac{\partial^2 g}{\partial y^2}(\mathbf{p}))^2 d\mathbf{p}$ , or penalizations involving more complex partial differential operators describing prior knowledge on the phenomenon under study (see, e.g., Azzimonti et al (2015)). As for the temporal dimension, we here adopt the classical penalty

$$J_T(h(t)) = \int_{T_1}^{T_2} \left( \frac{d^r h(t)}{dt^r} \right)^2 dt, \quad (3)$$

where  $h: [T_1, T_2] \rightarrow \mathbb{R}$ . See, e.g., Ramsay and Silverman (2005), Chapter 5, for details.

The spatial penalty  $J_S$  is applied to the spatio-temporal function  $f(\mathbf{p}, t)$  and then integrated over the temporal domain  $[T_1, T_2]$ , and analogously the temporal penalty  $J_T$  is applied to  $f(\mathbf{p}, t)$  and then integrated over the spatial domain  $\Omega$ . The field  $f$  is thus estimated by minimizing the following penalized sum of square error criterion:

$$\begin{aligned} J(f) = & \sum_{i=1}^n \sum_{j=1}^m (z_{ij} - f(\mathbf{p}_i, t_j))^2 + \\ & + \lambda_S \int_{T_1}^{T_2} \int_{\Omega} (\Delta f(\mathbf{p}, t))^2 d\mathbf{p} dt + \\ & + \lambda_T \int_{\Omega} \int_{T_1}^{T_2} \left( \frac{\partial^r f(\mathbf{p}, t)}{\partial t^r} \right)^2 dt d\mathbf{p}, \end{aligned} \quad (4)$$

where  $\lambda_S > 0$  and  $\lambda_T > 0$  are two smoothing parameters that weight the penalizations respectively in space and time. The choice of these parameters will be discussed in Section 3.3. As detailed in the following section, the regularizing terms in (4) induce the space-time covariance structure of the estimator and different regularizations would imply different covariance structures.

## 3 Numerical implementation of the model

We represent the spatio-temporal field  $f(\mathbf{p}, t)$  as an expansion on a separable space-time basis system. Specifically, let  $\{\varphi_k(t); k = 1, \dots, M\}$  be a set of  $M$  basis functions defined on  $[T_1, T_2]$  and  $\{\psi_l(\mathbf{p}); l = 1, \dots, N\}$  a set of  $N$  basis functions defined on  $\Omega$ . Then,  $f$  is represented by the following basis expansion:

$$f(\mathbf{p}, t) = \sum_{l=1}^N \sum_{k=1}^M c_{lk} \psi_l(\mathbf{p}) \varphi_k(t), \quad (5)$$

where  $\{c_{lk}; l = 1, \dots, N; k = 1, \dots, M\}$  are the coefficients of the expansion on the separable spatio-temporal basis.



Fig. 4: Simplified boundary of the Venice province.

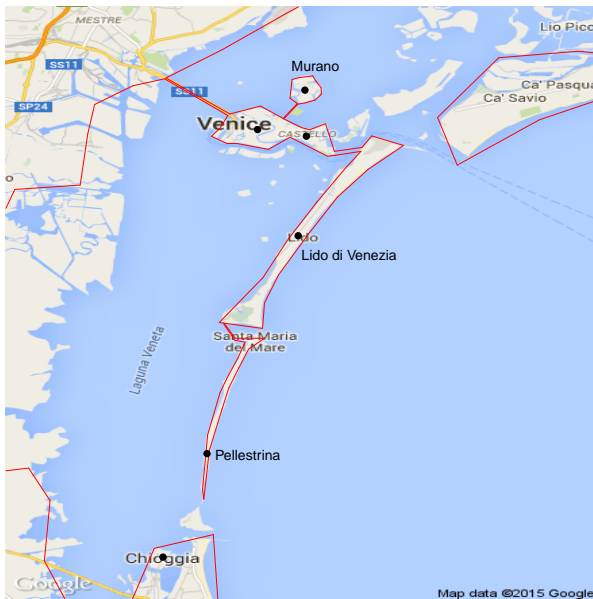


Fig. 5: Detail of the simplified boundary of the Venice province.

### 3.1 Choice of the basis systems in space and time

Various possible bases can be used for the expansions in the spatial and temporal domains. In this work, we use in space a finite element basis on a triangulation  $\Omega_\tau$  of the spatial domain  $\Omega$  of interest. This choice leads to an efficient discretization of the functional  $J$  and allows an accurate account of the shape of the spatial domain.

We illustrate the construction of such basis on Venice domain. Before building the basis, we simplify the original

spatial domain represented in Figure 1, excluding the coastal uninhabited regions and the smaller islands, and keeping in the domain of study only the four major islands: Venice, Murano (at the north-east of Venice), Lido di Venezia (at the south-east of Venice) and Pellestrina (at the south of Lido). We then smooth the boundary of the domain with regression splines. Finally, we obtain a piecewise linear boundary, subsampling from this smooth curve so that the features characterizing the domain are preserved. Figure 4 shows the simplified boundary of Venice province, while Figure 5 shows the detail around the city of Venice. This region is particularly interesting since it shows the four islands we keep in the domain. Here the domain includes four bridges: one linking Venice to the continent and the others linking some of the islands between themselves; the first one is an actual bridge with a road and a railway, while the other bridges represent regular and frequent ferries among the islands.

A triangulation of the resulting simplified domain is then obtained using the R package `fdaPDE` (Lila et al, 2016). In particular, we start from a Delaunay triangulation, constrained within the simplified boundary, where each of the town locations and each point defining the simplified boundary become a triangle vertex. A more regular mesh is then obtained with additional vertices, imposing a maximum value to the triangle areas. Figure 6 displays the resulting triangulation of Venice province. For this application, here and in Section 7, instead of using as coordinates the latitude and longitude, we employ the UTM coordinate system, which allows to compute the distance between two points on the Earth's surface by means of the Euclidean distance instead of the geodesic distance.



Fig. 6: Triangulation of the Venice province.

The finite element basis is composed by globally continuous functions that coincides with a polynomial of a certain degree on each element of the domain triangulation. In particular we use here linear finite element basis, that are piecewise linear functions. The dimension of the spatial basis is strictly related to the triangulation of the spatial domain: there is one basis function for each knot of the triangulation; for linear finite elements, each basis is associated to a vertex of the triangulation and has value 1 at that vertex and 0 at all other vertices. Figure 7 shows an example of linear basis function.

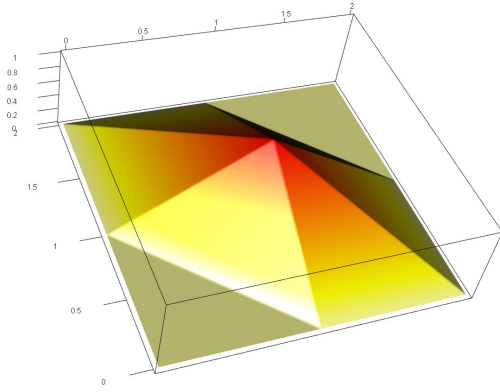


Fig. 7: Example of linear finite element basis function.

For the temporal dimension, we use here a cubic B-spline basis with penalization of the second derivative, with knots coinciding with the sampling time instants of the dataset. Other basis systems may turn out to be more appropriate in other applicative contexts. For instance, Fourier basis are well suited to the case of cyclic data, possibly with penalization of the harmonic acceleration operator, instead of the order  $r$  derivative considered in (3).

In case of dense sampling schemes, in space or time, coarser spatial or temporal grids may be preferred, for computational saving. But in general, the number of coefficients to be estimated,  $M \times N$ , may be larger than the corresponding number of observed data values  $m \times n$ . This does not create any problem from the estimation point of view, thanks to the presence of the regularizing terms. We never experienced any numerical instability of the method.

### 3.2 Discretization of the penalized sum-of-square error functional

Let  $\mathbf{z}$  be the vector of length  $nm$  of observed values at the  $n \times m$  spatio-temporal locations,  $\mathbf{f}$  the vector of length  $nm$  of evaluations of the spatio-temporal function  $f$  at the  $n \times m$  spatio-temporal locations, and  $\mathbf{c}$  the vector of length  $NM$  of

coefficients of the basis expansion (5) of the spatio-temporal field  $f$ , with entries ordered as follows

$$\mathbf{z} = \begin{bmatrix} z_{11} \\ \vdots \\ z_{1m} \\ z_{21} \\ \vdots \\ z_{2m} \\ \vdots \\ z_{nm} \end{bmatrix} \quad \mathbf{f} = \begin{bmatrix} f(\mathbf{p}_1, t_1) \\ \vdots \\ f(\mathbf{p}_1, t_m) \\ f(\mathbf{p}_2, t_1) \\ \vdots \\ f(\mathbf{p}_2, t_m) \\ \vdots \\ f(\mathbf{p}_n, t_m) \end{bmatrix} \quad \mathbf{c} = \begin{bmatrix} c_{11} \\ \vdots \\ c_{1M} \\ c_{21} \\ \vdots \\ c_{2M} \\ \vdots \\ c_{NM} \end{bmatrix}.$$

Let  $\Psi$  be the  $n \times N$  matrix of the evaluations of the  $N$  spatial basis functions in the  $n$  space locations  $\{\mathbf{p}_i; i = 1, \dots, n\}$ ,

$$\Psi = \begin{bmatrix} \psi_1(\mathbf{p}_1) & \psi_2(\mathbf{p}_1) & \dots & \psi_N(\mathbf{p}_1) \\ \psi_1(\mathbf{p}_2) & \psi_2(\mathbf{p}_2) & \dots & \psi_N(\mathbf{p}_2) \\ \vdots & \vdots & \dots & \vdots \\ \psi_1(\mathbf{p}_n) & \psi_2(\mathbf{p}_n) & \dots & \psi_N(\mathbf{p}_n) \end{bmatrix}.$$

Moreover, define the vectors of length  $N$  of the spatial basis functions  $\boldsymbol{\psi}$ , and of their first order partial derivatives  $\boldsymbol{\psi}_x$  and  $\boldsymbol{\psi}_y$ , by

$$\boldsymbol{\psi} = \begin{bmatrix} \psi_1 \\ \psi_2 \\ \vdots \\ \psi_N \end{bmatrix} \quad \boldsymbol{\psi}_x = \begin{bmatrix} \partial \psi_1 / \partial x \\ \partial \psi_2 / \partial x \\ \vdots \\ \partial \psi_N / \partial x \end{bmatrix} \quad \boldsymbol{\psi}_y = \begin{bmatrix} \partial \psi_1 / \partial y \\ \partial \psi_2 / \partial y \\ \vdots \\ \partial \psi_N / \partial y \end{bmatrix}.$$

Finally, let  $R_0$  be the  $N \times N$  matrix of the integrals over  $\Omega_\tau$  of the cross products of the  $N$  spatial basis, i.e.,

$$R_0 = \int_{\Omega_\tau} \boldsymbol{\psi} \boldsymbol{\psi}^T. \quad (6)$$

Analogously, let  $\Phi$  be the  $m \times M$  matrix of the evaluations of the  $M$  temporal basis functions in the  $m$  time instants  $\{t_j; j = 1, \dots, m\}$ :

$$\Phi = \begin{bmatrix} \varphi_1(t_1) & \varphi_2(t_1) & \dots & \varphi_M(t_1) \\ \varphi_1(t_2) & \varphi_2(t_2) & \dots & \varphi_M(t_2) \\ \vdots & \vdots & \dots & \vdots \\ \varphi_1(t_m) & \varphi_2(t_m) & \dots & \varphi_M(t_m) \end{bmatrix}.$$

Moreover, define the vectors of length  $M$  of the temporal basis functions  $\boldsymbol{\varphi}$ , and of their second order derivatives  $\boldsymbol{\varphi}_{tt}$ , by

$$\boldsymbol{\varphi} = \begin{bmatrix} \varphi_1 \\ \varphi_2 \\ \vdots \\ \varphi_M \end{bmatrix} \quad \boldsymbol{\varphi}_{tt} = \begin{bmatrix} d^2 \varphi_1 / dt^2 \\ d^2 \varphi_2 / dt^2 \\ \vdots \\ d^2 \varphi_M / dt^2 \end{bmatrix}.$$

Finally, let  $K_0$  be the  $M \times M$  matrix of the integrals over  $[T_1, T_2]$  of the cross products of the  $M$  temporal basis, i.e.,

$$K_0 = \int_{T_1}^{T_2} \boldsymbol{\varphi} \boldsymbol{\varphi}^T. \quad (7)$$

Consider now the  $nm \times NM$  matrix  $B = \Psi \otimes \Phi$ , where  $\otimes$  denotes the Kronecker product. Then  $\mathbf{f} = B\mathbf{c}$ . We may then rewrite the sum of square error functional  $J$  in (4) as

$$\begin{aligned} J &= (\mathbf{z} - B\mathbf{c})^T (\mathbf{z} - B\mathbf{c}) + \lambda_S \mathbf{c}^T (P_S \otimes K_0) \mathbf{c} + \lambda_T \mathbf{c}^T (R_0 \otimes P_T) \mathbf{c} \\ &= (\mathbf{z} - B\mathbf{c})^T (\mathbf{z} - B\mathbf{c}) + \mathbf{c}^T P \mathbf{c}, \end{aligned} \quad (8)$$

where  $P_S$  and  $P_T$  are the matrix discretizations of the spatial and temporal penalization terms, and  $P$  is the overall penalty  $P = \lambda_S (P_S \otimes K_0) + \lambda_T (R_0 \otimes P_T)$ . Specifically, the matrix  $P_T$  is obtained by direct discretization of the temporal penalty term in (3):

$$P_T = \int_{T_1}^{T_2} \boldsymbol{\varphi}_u \boldsymbol{\varphi}_u^T;$$

see Ramsay and Silverman (2005) for details. For the matrix  $P_S$ , following Ramsay (2002) and Sangalli et al (2013), we consider a computationally efficient discretization of the spatial penalty term in (2), that does not involve the computation of second order derivatives of the basis functions, but only of first order derivatives. This discretization is given by  $P_S = R_1 R_0^{-1} R_1$ , where

$$R_1 = \int_{\Omega_\tau} (\boldsymbol{\psi}_x \boldsymbol{\psi}_x^T + \boldsymbol{\psi}_y \boldsymbol{\psi}_y^T),$$

and it is based on a variational characterization of the estimation problem; see Ramsay (2002) for details. This formulation uses the Neumann condition at the boundary of the domain of interest implying zero flow across the boundary. Various other boundary conditions are possible; see Sangalli et al (2013). As shown in Azzimonti et al (2015), in the finite element space used to discretize the problem, the matrix  $P_S$  is in fact equivalent to the penalty matrix that would be obtained as direct discretization of the penalty term in (2) and (4) and involving the computation of second order derivatives.

Finally, the coefficients vector  $\hat{\mathbf{c}}$  that minimizes the functional  $J$  in (8) is computed deriving  $J$  with respect to  $\mathbf{c}$  and setting the derivative equal to 0, obtaining

$$\hat{\mathbf{c}} = (B^T B + P)^{-1} B^T \mathbf{z}.$$

### 3.3 Properties of the estimator

The estimator  $\hat{\mathbf{c}}$  is linear in the observed data values  $\mathbf{z}$ , and has a typical penalized least-square form. Denote by  $I_d$  the identity matrix of dimension  $d$ . Since  $E[\mathbf{z}] = \mathbf{f}$  and  $\text{Var}[\mathbf{z}] = \sigma^2 I_{nm}$ , we obtain

$$\begin{aligned} E[\hat{\mathbf{c}}] &= (B^T B + P)^{-1} B^T \mathbf{f}, \\ \text{Var}[\hat{\mathbf{c}}] &= \sigma^2 (B^T B + P)^{-1} B^T B (B^T B + P)^{-1}. \end{aligned}$$

Consider the vector  $\mathbf{B}(\mathbf{p}, t) = \boldsymbol{\psi}(\mathbf{p})^T \otimes \boldsymbol{\varphi}(t)^T$  of evaluations of the separable basis system at the spatio-temporal location  $(\mathbf{p}, t)$ , with  $\mathbf{p} \in \Omega$  and  $t \in [T_1, T_2]$ . The estimate of the field  $f$  at this generic location is thus given by

$$\hat{f}(\mathbf{p}, t) = \mathbf{B}(\mathbf{p}, t) \hat{\mathbf{c}} = \mathbf{B}(\mathbf{p}, t) (B^T B + P)^{-1} B^T \mathbf{z}$$

and its mean and variance are given by

$$\begin{aligned} E[\hat{f}(\mathbf{p}, t)] &= \mathbf{B}(\mathbf{p}, t) (B^T B + P)^{-1} B^T \mathbf{f} \\ \text{Var}[\hat{f}(\mathbf{p}, t)] &= \sigma^2 \mathbf{B}(\mathbf{p}, t) (B^T B + P)^{-1} B^T B (B^T B + P)^{-1} \mathbf{B}(\mathbf{p}, t)^T. \end{aligned} \quad (9)$$

The regularizing terms in (4) induce the space-time covariance structure of the estimator, given by

$$\begin{aligned} \text{Cov}[\hat{f}(\mathbf{p}_1, t_1), \hat{f}(\mathbf{p}_2, t_2)] &= \\ \sigma^2 \mathbf{B}(\mathbf{p}_1, t_1) (B^T B + P)^{-1} B^T B (B^T B + P)^{-1} \mathbf{B}(\mathbf{p}_2, t_2)^T, \end{aligned} \quad (10)$$

where  $(\mathbf{p}_1, t_1), (\mathbf{p}_2, t_2)$  are two space-time locations in the considered space-time domain and  $P$  is the discretization of the chosen regularizing terms. Different regularizations would imply different covariance structures. For instance, Azzimonti et al (2015) consider a regularized spatial regression model and show that by changing the regularizing terms and considering more complex differential operators it is possible to include in the model a priori information about the spatial variation of the phenomenon, and model also anisotropies and non-stationarities. The proposed modelling (1)-(4) jointly defines the first order structure and second order structure of the estimator.

The smoothing matrix

$$S = B (B^T B + P)^{-1} B^T$$

maps the vector of observed values  $\mathbf{z}$  to the vector of fitted values  $\hat{\mathbf{z}} = \hat{\mathbf{f}} = S\mathbf{z}$ . The trace of the smoothing matrix constitutes a commonly used measure of the equivalent degrees of freedom for linear estimators. We can thus estimate  $\sigma^2$  by

$$\hat{\sigma}^2 = \frac{1}{nm - \text{tr}(S)} (\mathbf{z} - \hat{\mathbf{z}})^T (\mathbf{z} - \hat{\mathbf{z}}). \quad (11)$$

This estimate of the error variance, plugged into (9), can be used to compute approximate pointwise confidence intervals for  $f$ . Moreover, the value of a new observation at location point  $\mathbf{p}_{n+1} \in \Omega$  and time instant  $t_{m+1} \in [T_1, T_2]$  can

be predicted by  $\hat{z}_{n+1\ m+1} = \hat{f}(\mathbf{p}_{n+1}, t_{m+1})$ , and approximate prediction intervals may be constructed.

Finally, the values of the smoothing parameters  $\lambda_S$  and  $\lambda_T$  may be chosen via Generalized Cross-Validation (GCV), searching for the values of  $\lambda_S, \lambda_T$  that minimize

$$GCV(\lambda_S, \lambda_T) = \frac{nm}{(nm - \text{tr}(S))^2} (\mathbf{z} - \hat{\mathbf{z}})^T (\mathbf{z} - \hat{\mathbf{z}}).$$

#### 4 Model with covariates

The model described above can be easily extended to include space-time varying covariates. Consider the semi-parametric generalized additive model

$$z_{ij} = \mathbf{w}_{ij}^T \boldsymbol{\beta} + f(\mathbf{p}_i, t_j) + \varepsilon_{ij} \quad i = 1, \dots, n, \quad j = 1, \dots, m, \quad (12)$$

where  $\mathbf{w}_{ij}$  is a vector of  $q$  covariates associated to the observation  $z_{ij}$ , at location  $\mathbf{p}_i$  and time instant  $t_j$ , and  $\boldsymbol{\beta}$  is a vector of  $q$  regression coefficients. We can jointly estimate the vector of regression coefficient  $\boldsymbol{\beta}$  and the spatio-temporal field  $f$  by minimizing the following penalized sum of square errors criterion

$$J(f, \boldsymbol{\beta}) = \sum_{i=1}^n \sum_{j=1}^m (z_{ij} - \mathbf{w}_{ij}^T \boldsymbol{\beta} - f(\mathbf{p}_i, t_j))^2 + \quad (13)$$

$$+ \lambda_S \int_{T_1}^{T_2} \int_{\Omega} (\Delta f(\mathbf{p}, t))^2 d\mathbf{p} dt + \quad (14)$$

$$+ \lambda_T \int_{\Omega} \int_{T_1}^{T_2} \left( \frac{\partial^2 f(\mathbf{p}, t)}{\partial t^2} \right)^2 dt d\mathbf{p}. \quad (15)$$

Let  $W$  be the  $nm \times q$  matrix containing the vectors  $\{\mathbf{w}_{ij}; i = 1, \dots, n; j = 1, \dots, m\}$ :

$$W = \begin{bmatrix} \mathbf{w}_{11}^T \\ \mathbf{w}_{12}^T \\ \vdots \\ \mathbf{w}_{1m}^T \\ \mathbf{w}_{21}^T \\ \vdots \\ \mathbf{w}_{2m}^T \\ \vdots \\ \mathbf{w}_{nm}^T \end{bmatrix}.$$

Let  $H_W$  be the matrix that projects orthogonally on the space generated by the columns of  $W$ , i.e.  $H_W = W(W^T W)^{-1} W^T$  and let  $Q = I_{nm} - H_W$ . The discretization of the functional  $J(f, \boldsymbol{\beta})$  in (13) is given by

$$J = (\mathbf{z} - W\boldsymbol{\beta} - B\mathbf{c})^T (\mathbf{z} - W\boldsymbol{\beta} - B\mathbf{c}) + \mathbf{c}^T P\mathbf{c}.$$

To compute the estimates of the vector of regression coefficients  $\boldsymbol{\beta}$  and of the vector  $\mathbf{c}$  of coefficients of the basis expansion of the spatio-temporal field  $f$ , we compute the first partial derivatives of  $J$  with respect to  $\boldsymbol{\beta}$  and  $\mathbf{c}$ , and set them

equal to zero, getting the following explicit solution of the estimation problem:

$$\hat{\boldsymbol{\beta}} = (W^T W)^{-1} W^T (\mathbf{z} - B\hat{\mathbf{c}}),$$

$$\hat{\mathbf{c}} = (B^T Q B + P)^{-1} B^T Q \mathbf{z}.$$

The estimator  $\hat{\mathbf{c}}$  has a penalized least-square form; given  $\hat{\mathbf{c}}$ , the estimator  $\hat{\boldsymbol{\beta}}$  has a least square form.

#### 4.1 Properties of the estimator

Let  $S_f = B(B^T Q B + P)^{-1} B^T Q$ , so that

$$\hat{\boldsymbol{\beta}} = (W^T W)^{-1} W^T (I_{nm} - S_f) \mathbf{z}.$$

Since  $E[\mathbf{z}] = W\boldsymbol{\beta} + \mathbf{f}$  and  $\text{Var}[\mathbf{z}] = \sigma^2 I_{nm}$ , and exploiting the fact that the matrix  $Q$  is idempotent and  $QW = 0$ , we obtain

$$E[\hat{\mathbf{c}}] = (B^T Q B + P)^{-1} B^T Q \mathbf{f},$$

$$\text{Var}[\hat{\mathbf{c}}] = \sigma^2 (B^T Q B + P)^{-1} B^T Q B (B^T Q B + P)^{-1}$$

and

$$E[\hat{\boldsymbol{\beta}}] = \boldsymbol{\beta} + (W^T W)^{-1} W^T (I_{nm} - S_f) \mathbf{f},$$

$$\text{Var}[\hat{\boldsymbol{\beta}}] = \sigma^2 (W^T W)^{-1} + \sigma^2 (W^T W)^{-1} W^T S_f S_f^T W (W^T W)^{-1}. \quad (16)$$

The estimate of the field  $f$  and its distributional properties follow as for the model without covariates. The smoothing matrix  $S$ , such that  $\hat{\mathbf{z}} = S\mathbf{z}$ , is now given by

$$S = H_W + Q S_f.$$

The trace of this matrix is given by  $\text{tr}(S) = q + \text{tr}(S_f)$  and measures the edf of this estimator, given by the sum of the  $q$  degrees of freedom corresponding to the parametric part of the model and the  $\text{tr}(S_f)$  degrees of freedom corresponding to the non-parametric part of the model. We can estimate  $\sigma^2$  as in (11). Given this estimate, it is possible to construct approximate pointwise confidence intervals for  $f$  as in the case without covariates. Moreover, using  $\hat{\sigma}^2$  in (16), it is now also possible to compute approximate confidence intervals for  $\boldsymbol{\beta}$ . Finally, the value of a new observation at location point  $\mathbf{p}_{n+1} \in \Omega$  and time instant  $t_{m+1} \in [T_1, T_2]$  and with associated covariates  $\mathbf{w}_{n+1\ m+1}$  can be predicted by  $\hat{z}_{n+1\ m+1} = \mathbf{w}_{n+1\ m+1}^T \hat{\boldsymbol{\beta}} + \hat{f}(\mathbf{p}_{n+1}, t_{m+1})$ , and approximate prediction intervals may be constructed.

#### 5 A variation of the proposed ST-PDE model

A variation of the proposed ST-PDE model can be defined as follows. Equivalently to (5), we can express the spatio-temporal field  $f$  in the following basis expansions:

$$f(\mathbf{p}, t) = \sum_{k=1}^M a_k(\mathbf{p}) \varphi_k(t) \quad (17)$$

$$f(\mathbf{p}, t) = \sum_{l=1}^N b_l(t) \psi_l(\mathbf{p}), \quad (18)$$

where  $\{a_k(\mathbf{p}); k = 1, \dots, M\}$  are the spatially varying coefficients of the expansion of the space-time field on the temporal basis and  $\{b_l(t); l = 1, \dots, N\}$  are the temporally varying coefficients of the expansion of the field on the spatial basis. Then, in analogy with Augustin et al (2013) and Marra et al (2012), we can apply the spatial penalty  $J_S$  to the  $M$  spatially varying coefficients  $a_k(\mathbf{p})$  in the basis expansion (17), and the temporal penalty  $J_T$  to the  $N$  temporally varying coefficients  $b_l(t)$  in the basis expansion (18). In this case, the field  $f$  is thus estimated by minimizing the following penalized sum of square error criterion:

$$\begin{aligned} \tilde{J}(f) = & \sum_{i=1}^n \sum_{j=1}^m (z_{ij} - f(\mathbf{p}_i, t_j))^2 + \\ & + \lambda_S \sum_{k=1}^M \int_{\Omega} (\Delta(a_k(\mathbf{p})))^2 d\mathbf{p} + \lambda_T \sum_{l=1}^N \int_{T_1}^{T_2} \left( \frac{d^r b_l(t)}{dt^r} \right)^2 dt. \end{aligned} \quad (19)$$

The numerical implementation of this method has only minor differences with the one presented in Section 3. Denote by  $I_d$  the identity matrix of dimension  $d$ . Then the discretized functional is given by

$$\tilde{J} = (\mathbf{z} - \mathbf{B}\mathbf{c})^T (\mathbf{z} - \mathbf{B}\mathbf{c}) + \lambda_S \mathbf{c}^T (P_S \otimes I_M) \mathbf{c} + \lambda_T \mathbf{c}^T (I_N \otimes P_T) \mathbf{c},$$

where the identity matrices  $I_M$  and  $I_N$  have replaced the matrices  $K_0$  in (7) and in  $R_0$  in (6), respectively. Minimizing the functional (19) is thus equivalent to minimizing the one in (4) if the spatial and the temporal basis used for the numerical discretization are orthonormal. In this work we use basis systems which are not orthonormal; nevertheless, the basis systems considered are sparse, so that the terms

$$\int_{T_1}^{T_2} \varphi_k(t) \varphi_l(t) dt \quad \text{and} \quad \int_{\Omega} \psi_l(\mathbf{p}) \psi_k(\mathbf{p}) d\mathbf{p}$$

are nonzero only for a few couples of indexes  $(l, k)$  with  $l \neq k$ . We compared the results of the two methods in all the simulation studies in Section 6 obtaining the same performances for the two methods in all cases. In the following we show only the results for the ST-PDE method.

## 6 Simulation studies

We present some simulation studies and compare the proposed model with three other approaches to spatio-temporal field estimation.

The first method is spatio-temporal kriging with a separable variogram marginally gaussian in space and exponential in time, chosen among a number of possible variogram models, with parameters estimated from the empirical variogram. (The choice among different variograms was based

on visual inspection of the resulting estimates, avoiding wiggly estimates but requesting that the spatio-temporal behavior of the field was well captured.) This method is implemented using the function `krigeST` of the R package `gstat` (Pebesma, 2004).

We then consider two space-time models presented in Augustin et al (2013) and Marra et al (2012). One model adopt a thin plate spline basis in space and a cubic spline basis in time, and minimizes a functional analogous to (19), where the spatial penalty is replaced by the thin plate spline energy recalled in Section 2. The other model uses the soap film smoothing described in Wood et al (2008) in space and a cubic spline basis in time, and minimize the same functional in (19). The two latter methods are implemented using the function `gam` of the R package `mgcv` (Wood, 2006). Finally, for these two methods, as well as for the model here proposed, the values of the smoothing parameters  $\lambda_S, \lambda_T$  are chosen via GCV.

We apply the aforementioned methods to simulated data on a C-shaped spatial domain. The test function, sampled at a few time instants, is shown in the left panels of Figure 8, and its analytical expression is detailed in the Appendix. This function is constructed starting from the spatial test function considered for instance in Ramsay (2002), Wood et al (2008) and Sangalli et al (2013), and introducing the time component in a not separable way.

We present here five simulation studies: in the first two cases we consider different sampling schemes, with fewer or more observations in space and in time; in the third case we include covariates; in the fourth case we consider correlated noise; finally, in the fifth case we start from areal data and consider the approximation consisting in assigning each datum to the area centroid.

### 6.1 First simulation study

We sample 200 spatial locations uniformly in the C-shaped domain at 9 time instants equally spaced from 0 to  $\pi$ . We simulate the data from model (1), with a gaussian noise with mean 0 and standard deviation 0.7.

Figure 8 shows in the first column the spatio-temporal test function at the 9 sampling time instants, in the second column the simulated data, and in the following columns the corresponding estimates obtained by spatio-temporal kriging (KRIG), the space-time model using thin plate spline (TPS), the space-time model using soap film smoothing (SOAP), and the space-time model here proposed (ST-PDE).



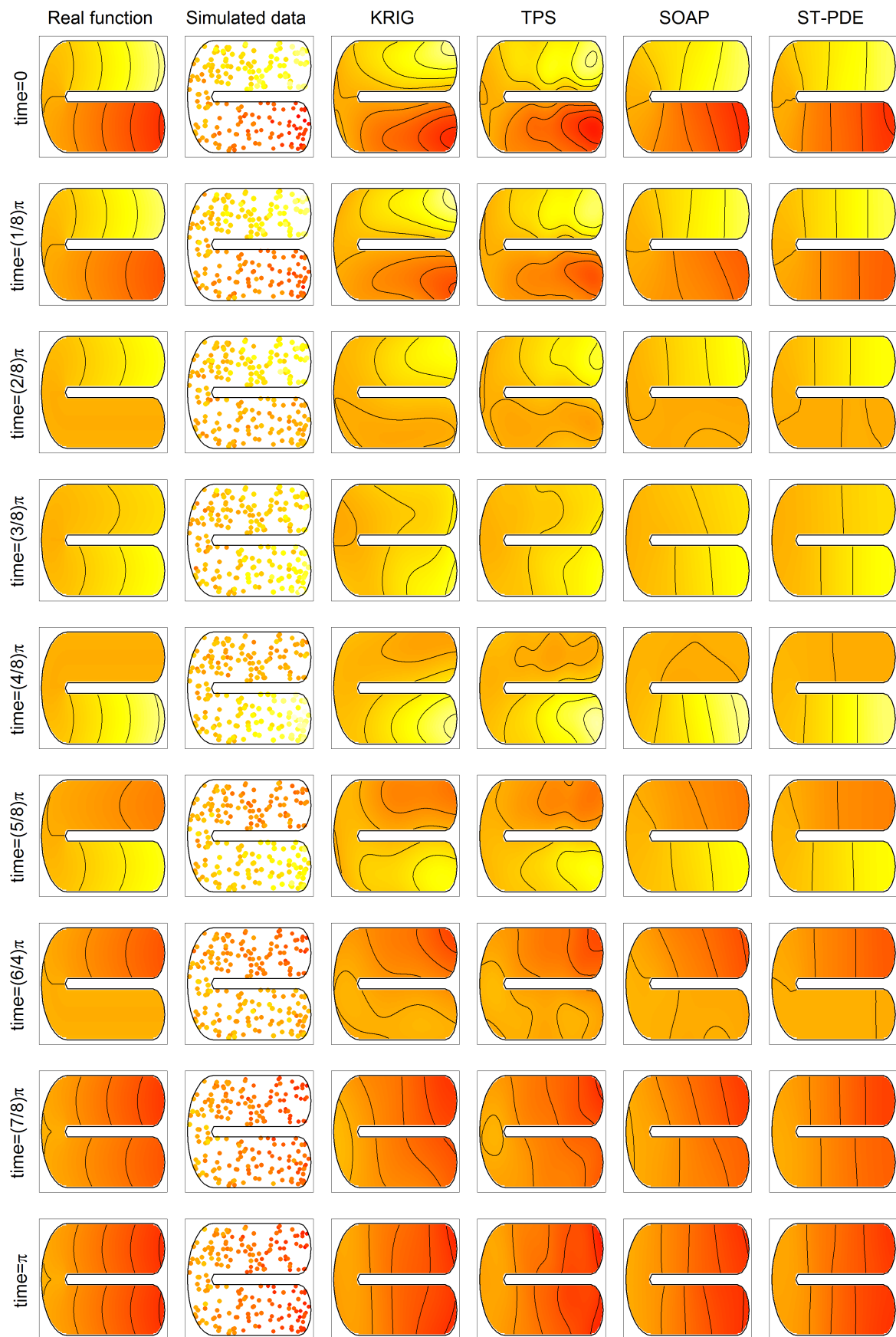


Fig. 8: The spatio-temporal test function, the simulated data and the estimated functions with spatio-temporal kriging (KRIG), space-time model using thin plate spline (TPS), space-time model using soap film smoothing (SOAP) and ST-PDE.

Figure 9 shows the boxplots of the Root Mean Square Errors (RMSE) of the space-time field estimates given by the four methods over 50 replicates of the noise generation. The RMSE is computed over a fine grid of the spatio-temporal domain (step 0.05 in space and  $\pi/24$  in time).

A visual inspection of the RMSE shows that SOAP and ST-PDE methods give better estimates than KRIG and TPS. The reason for this comparative advantage is apparent from Figure 8. In fact, the KRIG and TPS methods, that do not take into account the shape of the domain and smooth across the two arms of the C-shaped domain, provide poor estimates of the field when the true  $f$  is characterized by high values in one of the two C arms and low values in the other arm. The best estimates are provided by the ST-PDE model.

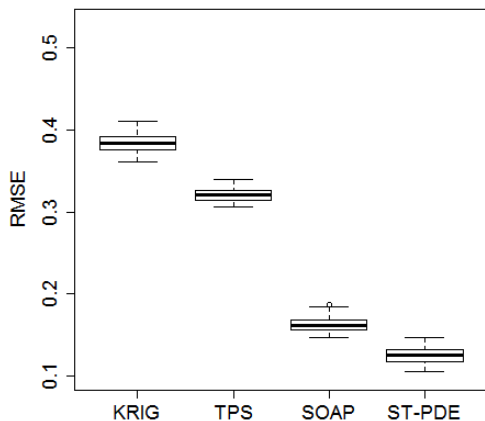


Fig. 9: First simulation study. Boxplots of the RMSE of the estimates of the spatio-temporal field obtained by the four tested methods: spatio-temporal kriging (KRIG), space-time model using thin plate spline (TPS), space-time model using soap film smoothing (SOAP) and ST-PDE.

### 6.2 Second simulation study: fewer spatial locations and more time instants

In this second simulation case we perform the same simulation study described in Section 6.1, but with fewer observations in space and more in time. Specifically, we consider 50 spatial locations uniformly distributed in the C-shaped domain and 33 time instants equally spaced in the time interval  $[0, \pi]$ .

Figure 10 shows the boxplots of the RMSE of the space-time field estimates over 50 replicates of the noise generation. The RMSE is computed over a fine grid of the spatio-temporal domain (step 0.05 in space and  $\pi/64$  in time). Also in this case, SOAP and ST-PDE provide better estimates than TPS and kriging, with ST-PDE returning the best estimates.

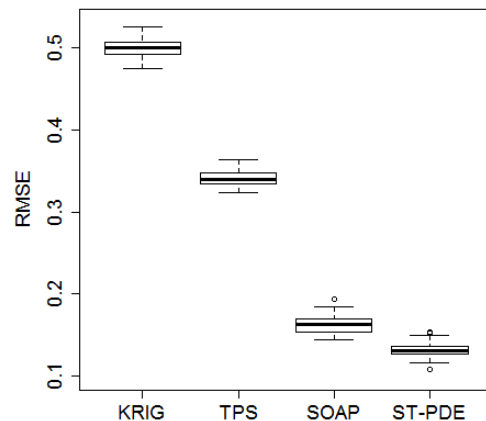


Fig. 10: Second simulation study: fewer spatial locations and more time instants. Boxplots of the RMSE of the estimates of the spatio-temporal field obtained by the four tested methods.

### 6.3 Third simulation study: covariates

Within the same simulation setting described in Section 6.1, we also perform a study including a space-time varying covariate. Specifically, the covariate used is the following space-time function:

$$w(x, y, t) = \cos(5x) \sqrt{\frac{t+1}{5}}. \quad (20)$$

We generate data from model (12), setting  $\beta = 1$ . The other simulation specifications are as in Section 6.1. We here compare the proposed ST-PDE method to the space-time models using thin plate spline and soap film smoothing. We do not instead consider the spatio-temporal kriging in this case because the function `krigeST` of the R package `gstat` cannot handle covariates.

Figure 11 shows the boxplots of the RMSE of the space-time field estimates over the 50 replicates of the noise generation. The RMSE is computed over the same fine grid of the spatio-temporal domain used in Section 6.1. Likewise in the simulation studies without covariates, SOAP and ST-PDE, that account for the shape of the domain, provide better estimates than TPS, that is instead blind to the domain structure. The best estimates are provided by the ST-PDE model.

The RMSE of the estimates of  $\beta$  over the 50 replicates are instead comparable for the three methods: 0.14 for TPS, 0.09 for both SOAP and ST-PDE. In the first simulation replicate, the approximate 95% confidence interval for the parameter  $\beta$  associated to the ST-PDE estimate is given by  $[0.88, 1.16]$ .

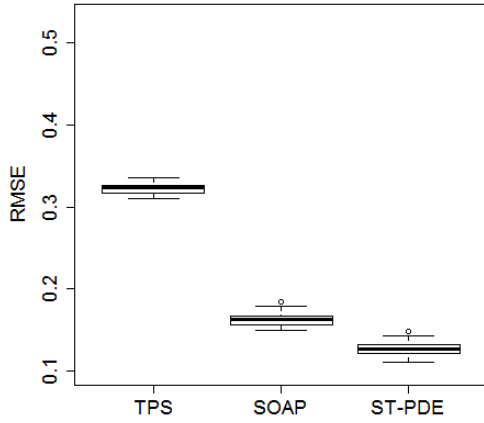


Fig. 11: Third simulation study: covariates. Boxplots of the RMSE of the estimates of the spatio-temporal field obtained by the three methods tested in this case: TPS, SOAP, ST-PDE.

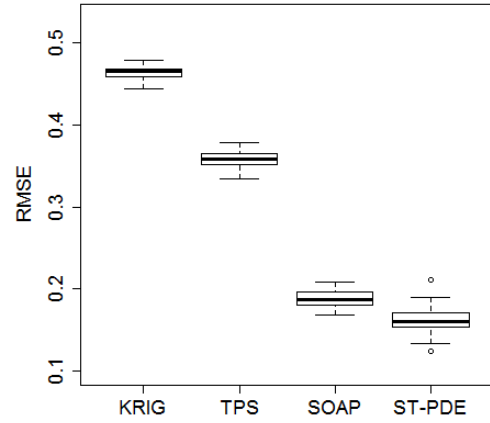


Fig. 12: Fourth simulation study: correlated noise. Boxplots of the RMSE of the estimates of the spatio-temporal field obtained by the four tested methods.

#### 6.4 Fourth simulation study: correlated noise

We performed a fourth simulation study within the same setting described in Section 6.1, but generating the additive noise with a spatio-temporal covariance structure. We use the function `RFsimulate` of the R package `RandomFields` (Schlather et al, 2015) to generate a spatio-temporal random field with mean zero and a stationary isotropic covariance model belonging to the Matern family. The corresponding covariance function, which depends on the distance  $r \geq 0$  between two points, is the following

$$C(r) = \sqrt{2r}K_1(\sqrt{2r}), \quad (21)$$

where  $K_1$  is the modified Bessel function of second kind; moreover we consider here an anisotropy matrix with stretch arguments equal to 40 for the space components and equal to 80 for the time component.

Figure 12 shows the boxplots of the RMSE of the space-time field estimates over the 50 replicates of the noise generation. The RMSE is computed over the same fine grid of the spatio-temporal domain used in Section 6.1.

The same observations made in the previous simulation studies still hold in this case.

#### 6.5 Fifth simulation study: areal data assigned to area centroid

In this simulation study we generate areal data over subdomains of the spatial domain and assign each datum to a point representing the center of the subdomain. This simulation study is devised to mimic our applicative context, where, as detailed in the following section, we are lead to approximate areal data by assigning them to point locations, due to the unavailability of information concerning the subdomains shapes. We do not in general suggest to make such an approximation, unless it is unavoidable. As detailed in Section 8, in future research we intend to develop a model extension to handle areal data. In any case, we show that also in this approximated and simplified data setting, the model proposed outperforms the competitor methods.

We sample 50 square ( $0.2 \times 0.2$ ) spatial subdomains  $D_i$  uniformly distributed over the spatial C-shaped domain and we compute their centers  $\mathbf{p}_i$ . Figure 13 represents the spatial subdomains and their centers. We consider 9 time instants  $\{t_1, \dots, t_9\}$  equally spaced from 0 to  $\pi$ . We compute the data in the following way:

$$z_{ij} = \frac{1}{|D_i|} \int_{D_i} f(\mathbf{p}, t_j) d\mathbf{p} + \varepsilon_{ij} \quad i = 1, \dots, 50, j = 1, \dots, 9,$$

and assign each value  $z_{ij}$  to the spatio-temporal location  $(\mathbf{p}_i, t_j)$ . The noise has the same distribution as in the previous simulation studies.

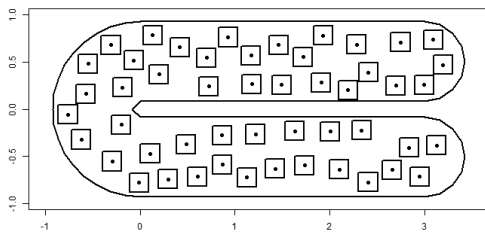


Fig. 13: Spatial subdomains.

Figure 14 shows the boxplots of the RMSE of the space-time field estimates given by the four methods over 50 replicates of the noise generation. The RMSE is computed over the same fine grid of the spatio-temporal domain used in Section 6.1.

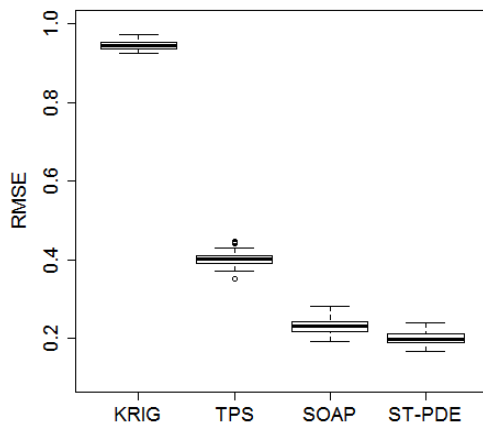


Fig. 14: Fifth simulation study: areal data assigned to area centroid. Boxplots of the RMSE of the estimates of the spatio-temporal field obtained by the four tested methods: spatio-temporal kriging (KRIG), space-time model using thin plate spline (TPS), space-time model using soap film smoothing (SOAP) and ST-PDE.

The results are analogous to the ones obtained in the previous simulation studies: SOAP and ST-PDE provide better estimates than TPS and kriging, with kriging performing worse than in the previous simulation cases considered. The best estimates are also in this case provided by the ST-PDE model.

## 7 Application to the analysis of the production of waste in Venice province

We apply the ST-PDE method to the dataset of annual amount of per capita municipal waste produced in the Venice province.

### 7.1 The Venice waste dataset

Open Data Veneto<sup>1</sup> provides the gross and per capita annual amount of municipal waste produced in each municipality of the Venice province in the period from 1997 to 2011. We here consider for the analysis the annual per capita municipal waste, in kg per municipality resident.

Municipal waste includes that produced in houses and public areas, but does not include special waste, i.e. industrial, agricultural, construction and demolition waste, or hazardous waste, for which there are special disposal programs. Therefore, the data refer only to the urban area of the municipality, whilst they do not refer to the agricultural or industrial areas in the municipality territories. Since no data identifying the urbanized areas of the municipalities is available, we face here two possible simplifications of the problem. We can either partition the Venice province in the municipality territories and attribute each datum to the whole territory of its municipality, or assign each datum to a point representing the center of the municipality. We here adopt the second simplification. The spatial coordinates of the town centers are available online<sup>2</sup>. As mentioned in Section 3, latitude and longitude are converted into UTM coordinate system.

In some cases there are localities which do not constitute a municipality on their own, but are under the jurisdiction of another town. In this case, there are two or more main urbanized areas in the municipality territory. Some of these localities are not negligible for the problem under analysis due to their tourist relevance and their location on the domain; for this reason we add them to the data. Specifically, we include the seaside town of Bibione, the eastern most village indicated in Figure 1. This popular vacation destination falls under the jurisdiction of the municipality of San Michele al Tagliamento, north west of Bibione; the waste data considered for Bibione are a replicate of the data of San Michele al Tagliamento. Moreover, we replicate the data of Venice in the islands of Murano, Lido di Venezia and Pellestrina, because of their tourist relevance and the particular shape of the domain.

Since intuition suggests that the production of waste is affected by tourism, we include in the model a covariate which accounts for it. Specifically we consider the number of beds in accommodation facilities (such as hotels, bed and breakfast, guest houses, campings, etc.) divided by the number of residents. This ratio may be as large as 7 in some tourist towns by the sea. The number of beds in accommodation facilities is provided by Istat<sup>3</sup>, the Italian national institute for statistics.

<sup>1</sup> <http://dati.veneto.it/dataset/produzione-annua-di-rifiuti-urbani-totale-e-pro-capite-1997-2011>

<sup>2</sup> <http://www.dossier.net/utilities/coordinate-geografiche/>

<sup>3</sup> <http://www.istat.it/it/archivio/113712>

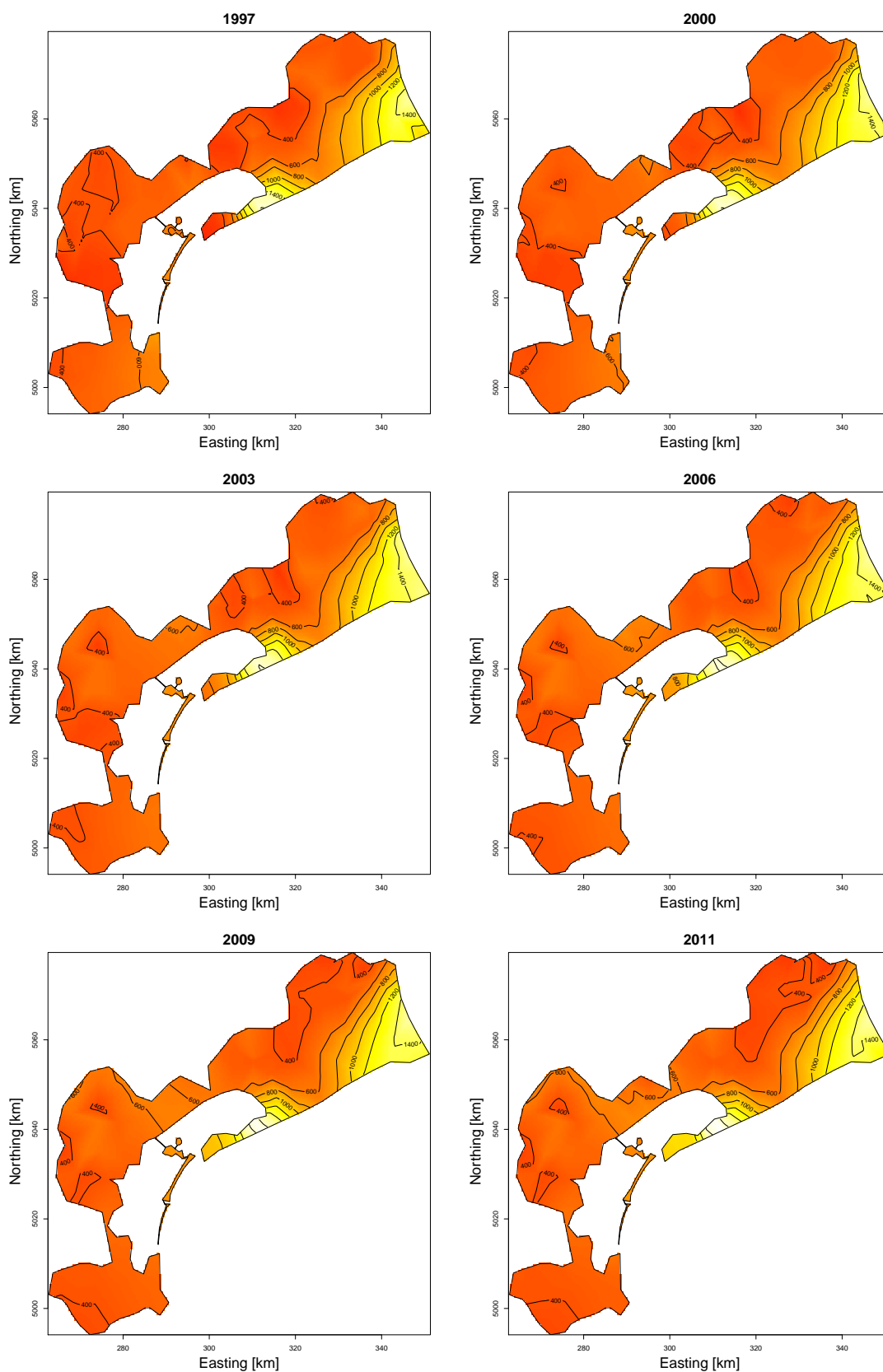


Fig. 15: Estimated spatio-temporal field for the Venice waste data (yearly per capita production) at fixed time instants.

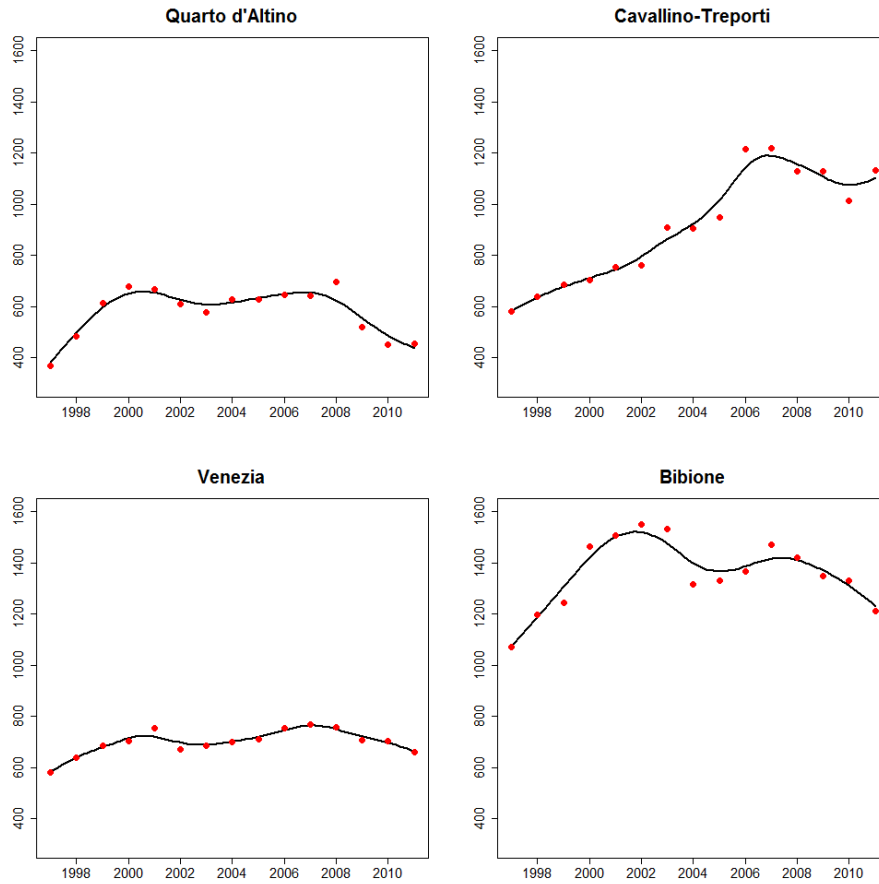


Fig. 16: Temporal evolution of the estimated spatio-temporal field for the Venice waste data (yearly per capita production) at fixed spatial locations.

## 7.2 Analysis of Venice waste data by ST-PDE

Figure 15 shows the estimated spatio-temporal field at fixed time instants. The estimate for the coefficient  $\beta$  is 39.7 meaning that one more unit in the ratio between the number of beds in accommodation facilities and the number of residents is estimated to increase the yearly per capita production of waste by residents by about 40kg. The estimated spatial field  $f$  shows the highest values, across the years, in correspondence of the coastline, around the towns of Bibione, Lido di Jesolo and Cavallino-Treporti. These higher values may be due to a type of tourism that is not captured by the available covariate, such as daily tourists who do not stay overnight, and vacationers who either own or rent vacation houses. The higher values of the field are also probably due to the presence of many seasonal workers, working in accommodation facilities, restaurants, cafés, shops, beach resorts and other services, who are not residents of these towns.

Although Venice is one of the most visited cities in Italy, and this tourism is active all year round, the production of

per capita waste in Venice appears to be lower than in other nearby tourist localities by the seaside. This might be partly explained by the fact that the tourist activities in Venice are not so highly characterized by seasonality as in the smaller seaside villages, and people working in tourist activities in Venice are more likely to be themselves residents of this large city.

It is significant to notice how the estimated function does not smooth across concave boundaries. For example, the area of the city of Quarto d'Altino and the one around the city of Cavallino-Treporti show different ranges of values. Indeed, even though the two towns are geographically close, they are separated by the Venetian lagoon. This difference is evident also from the first two panels of Figure 16, which shows the estimated spatio-temporal field at fixed localities: Quarto d'Altino, Cavallino-Treporti, Venice and Bibione. In these plots the red dots are obtained subtracting from the data the estimated contribution by the covariate, i.e.  $\hat{\beta}w_{ij}$ .

The temporal evolution plots in Figure 16 show the ability of the method to capture the temporal trend of the phenomenon. The method provides good estimates also for the

municipality of Cavallino-Treporti, which presents a strong variation of per capita waste over the year. The large increase of the per capita waste of Cavallino-Treporti is partly explained by the fact that, during the first years of this study, this town was under the jurisdiction of Venice, while the data for this new municipality are available only from 2002. In particular, the data for Cavallino-Treporti for years 1997 to 2001 are a replicate of the data of the municipality of Venice. Nevertheless, the strong variation in the data is well captured by the estimated function.

## 8 Model extensions

Various extensions of the proposed model are possible. A first generalization consists in modelling data that are areal in space and integral in time, and estimating an underlying spatio-temporal intensity function. In the application to Venice waste data, if information about the urbanized areas of each municipality would become available, such a model extension would for instance allow to appropriately refer the waste datum to the area and year where it is produced, estimating a spatio-temporal intensity of waste production.

Extending the work of Azzimonti et al (2015) it is also possible to include a priori information available on the phenomenon under study, using more complex differential regularizations modelling the space and/or time behavior of the phenomenon. This also allows to account for non-stationarities and anisotropies in space and/or time. Along the same lines, if a priori information about the interaction between space and time was available, then it would make sense to consider a unique space/time regularizing term based on a time-dependent PDE that governs the phenomenon behavior. Azzimonti et al (2015) for instance analyze the blood flow velocity in a section of the carotid artery at a fixed time instant corresponding to the systolic peak, starting from Echo-Color Doppler data, and including a priori information on the problem under study. By introducing the time dimension, we could study how the blood flow velocity field varies during the time of the heart-beat. PDEs are commonly used to describe complex phenomena behavior in many fields of engineering and sciences, including bio-sciences, geo-sciences, and physical sciences. Potential applications of particular interest of this space-time technique in the environmental sciences would for example concern the study of the dispersion of pollutant released in water or in air and transported by streams or winds, and the study of the propagation of earthquakes, tsunamis, and other wave phenomena. If one wishes instead to consider simpler isotropic and stationary regularizations, then a possibility to allow for stronger interactions in space/time, with respect to the model here presented, would consist in defining a unique regularizing term based on a heat equation.

Finally, data distributed over curved domains, instead of over planar domains, could be handled by extending the model proposed in Ettinger et al (2016). Considering the same application presented by Ettinger et al (2016) and Dassi et al (2015), this would permit the study of time-dependent hemodynamic forces exerted by blood-flow over the wall of inner carotid arteries affected by aneurysms, taking into account the complex morphology of these vessels. Another fascinating field of application of this modelling extension would be in the neurosciences, studying signals associated to neuronal activity over the cortical surface, a highly convoluted thin sheet of neural tissue that constitutes the outermost part of the brain. In the geo-sciences, this would permit the study of data distributed over regions with complex orographies. Moreover, generalizations to time-dependent data of the spatial regression model introduced by Wilhelm et al (2016) would also further broaden the applicability of the proposed model to important engineering applications, especially in the automotive, naval, aircraft and space sectors, where space-time varying quantities of interest are observed over the surface of a designed 3D object, such as the pressure over the surface of a shuttle winglet.

**Acknowledgements** We thank Alessandra Menafoglio for comments on this work. We are also grateful to the Associate Editor and three anonymous referees, whose suggestions greatly improved the presentation of this work. L.M. Sangalli acknowledges funding by MIUR Ministero dell'Istruzione dell'Università e della Ricerca, FIRB Futuro in Ricerca Starting Grant 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>.

## References

- Augustin NH, Trenkel VM, Wood SN, Lorance P (2013) Space-time modelling of blue ling for fisheries stock management. *Environmetrics* 24(2):109–119
- Azzimonti L, Sangalli LM, Secchi P, Domanin M, Nobile F (2015) Blood flow velocity field estimation via spatial regression with PDE penalization. *Journal of the American Statistical Association* 110(511):1057–1071
- Caballero W, Giraldo R, Mateu J (2013) A universal kriging approach for spatial functional data. *Stochastic Environmental Research and Risk Assessment* 27(7):1553–1563
- Cressie N, Wikle CK (2011) *Statistics for spatio-temporal data*. John Wiley & Sons
- Dassi F, Ettinger B, Perotto S, Sangalli LM (2015) A mesh simplification strategy for a spatial regression analysis over the cortical surface of the brain. *Applied Numerical Mathematics* 90:111–131
- Delicado P, Giraldo R, Comas C, Mateu J (2010) Statistics for spatial functional data: some recent contributions. *Environmetrics* 21(3-4):224–239

- Ettinger B, Perotto S, Sangalli LM (2016) Spatial regression models over two-dimensional manifolds. *Biometrika* 103(1):71–88
- Giraldo R, Delicado P, Mateu J (2011) Ordinary kriging for function-valued spatial data. *Environmental and Ecological Statistics* 18(3):411–426
- Goulard M, Voltz M (1993) Geostatistical interpolation of curves: A case study in soil science. In: *Geostatistics Tróia'92*, Springer, pp 805–816
- Ignaccolo R, Mateu J, Giraldo R (2014) Kriging with external drift for functional data for air quality monitoring. *Stochastic Environmental Research and Risk Assessment* 28(5):1171–1186
- Lila E, Sangalli LM, Ramsay J, Formaggia L (2016) *fdapDE: Regression with Partial Differential Regularizations, using the Finite Element Method*. URL <https://CRAN.R-project.org/package=fdapDE>, R package version 0.1-1
- Marra G, Miller DL, Zanin L (2012) Modelling the spatiotemporal distribution of the incidence of resident foreign population. *Statistica Neerlandica* 66(2):133–160
- Menafoglio A, Secchi P, Dalla Rosa M (2013) A Universal Kriging predictor for spatially dependent functional data of a Hilbert Space. *Electronic Journal of Statistics* 7:2209–2240
- Menafoglio A, Guadagnini A, Secchi P (2014) A kriging approach based on Aitchison geometry for the characterization of particle-size curves in heterogeneous aquifers. *Stochastic Environmental Research and Risk Assessment* 28(7):1835–1851
- Nerini D, Monestiez P, Manté C (2010) Cokriging for spatial functional data. *Journal of Multivariate Analysis* 101(2):409–418
- Pebesma EJ (2004) Multivariable geostatistics in S: the *gstat* package. *Computers & Geosciences* 30:683–691
- R Core Team (2014) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, URL <http://www.R-project.org/>
- Ramsay JO, Silverman BW (2005) *Functional Data Analysis*. Springer
- Ramsay T (2002) Spline smoothing over difficult regions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 64(2):307–319
- Sangalli LM, Ramsay JO, Ramsay TO (2013) Spatial spline regression models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 75(4):681–703
- Schlather M, Malinowski A, Menck PJ, Oesting M, Strokorb K (2015) Analysis, simulation and prediction of multivariate random fields with package *RandomFields*. *Journal of Statistical Software* 63(8):1–25
- Wilhelm M, Dedè L, Sangalli LM, Wilhelm P (2016) IGS: an IsoGeometric approach for Smoothing on surfaces.

*Computer Methods in Applied Mechanics and Engineering* DOI 10.1016/j.cma.2015.12.028

Wood S (2006) *Generalized additive models: an introduction with R*. CRC press

Wood SN, Bravington MV, Hedley SL (2008) Soap film smoothing. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 70(5):931–955

## A Spatio-temporal test function

The spatio-temporal test function  $f(x, y, t)$ , defined over the C-shaped domain, used in the simulation studies, is constructed as:

$$\begin{aligned} & \cos(t)(q+x) + (y-r)^2 && \text{if } x \geq 0 \text{ \& } y > 0 \\ & \cos(2t)(-q-x) + (-y-r)^2 && \text{if } x \geq 0 \text{ \& } y \leq 0 \\ & \cos(t)(-\arctan(\frac{y}{x})r) + (\sqrt{x^2+y^2}-r)^2 K(x,y) && \text{if } x < 0 \text{ \& } y > 0 \\ & \cos(2t)(-\arctan(\frac{y}{x})r) + (\sqrt{x^2+y^2}-r)^2 K(x,y) && \text{if } x < 0 \text{ \& } y \leq 0 \end{aligned}$$

where  $K(x, y) = (\frac{y}{r_0} \mathbb{1}_{|y| \leq r_0 \& x > -r} + \mathbb{1}_{|y| > r_0 \& |x| \leq -r})^2$ ,  $\mathbb{1}_A$  denotes the indicator function of the subset  $A$ ,  $r_0 = 0.1$ ,  $r = 0.5$  and  $q = \frac{\pi r}{2}$ .