Politecnico di Milano
Department of Mathematics
Doctoral Program In Mathematical Models
and Methods in Engineering

Modelling spatial and spatio-temporal fields via regression with PDE penalization

Doctoral Dissertation of:
Mara Sabina Bernardi

Supervisor:
Prof. Piercesare Secchi

Co-supervisor:
Prof. Laura Maria Sangalli

Tutor:
Prof. Roberto Lucchetti

The Chair of the Doctoral Program:
Prof. Irene Maria Sabadini

Year 2017 – XXIX Cycle

Contents

Introduction

In this thesis we develop a class of models for the analysis of spatially distributed data with complex dependencies, such as space-time data over spatial domains with a complex shape, or data characterized by stationary or non-stationary spatial anisotropy. Appropriately taking into account the structure of the complex dependencies present in the data is fundamental to provide accurate estimation for data arising from many fields of sciences, such as physics, biology, meteorology and geology.

The proposed method is based on spatial regression with partial differential penalization. The idea of regularization with differential operators is common in functional data analysis for instance to tackle the problem of univariate or multivariate curve fitting (Ramsay and Silverman, 2005). Similar approaches can be used for the estimation of surfaces or spatial fields. Some examples are given by thin plate splines (Wahba, 1990), soap-film smoothing (Wood et al., 2008) and bivariate splines (Guillas and Lai, 2010).

In this work, we consider the spatial regression with partial differential regularization method introduced by Ramsay (2002), Sangalli et al. (2013) and Azzimonti et al. (2015). This method interfaces advanced statistical methodology and numerical analysis techniques. In particular, it makes use of the finite element method.

This technique is able to efficiently deal with data distributed over complex domains, including irregularly shaped domains, featuring strong concavities and interior holes (Sangalli et al., 2013), or non-planar domains (Ettinger et al., 2016; Dassi et al., 2015). Furthermore, the method can take into account specific information on the behavior of the phenomenon under study, which is fundamental in many applications to obtain meaningful estimates. Indeed, the model has the capacity to incorporate problem-specific prior information about the spatial structure of the phenomenon under study, formalized in terms of a governing partial differential equation (Azzimonti et al., 2014, 2012) and can provide estimated fields which satisfy known boundary conditions (Sangalli et al., 2013; Azzimonti et al., 2014, 2012). Moreover, if space-varying covariates are available, they can be accounted for in the model via a semi-parametric framework.

This thesis shows that spatial regression with partial differential regulariza-

tion can be extended to deal with more complex types of spatial or spatio-temporal dependences, including space-time data observed over complex spatial domains (Chapter 1), spatial data characterized by unknown spatial stationary anisotropy (Chapter 2), and spatial data characterized by spatial non-stationary anisotropy driven by the texture of the domain (Chapter 3).

In Chapter 1 we extend the spatial model presented in Ramsay (2002) and Sangalli et al. (2013) to include the time dimension. This chapter is based on Bernardi et al. (2016). In particular, we consider spatio-temporal data defined on a spatial domain featuring strong concavities which affect the phenomenon under study. In this work we introduce a penalization which separately takes into account the spatial dimension and the temporal dimension. Simulation studies show the good performances of the proposed model with respect to other spatio-temporal techniques. In particular, the proposed technique provides more accurate estimates when the phenomenon under study is influenced by the complex geometry of the spatial domain. Indeed, the classical methods for spatio-temporal data are not well suited for the analysis of this kind of data since they do not take into account the geometry of the domain. The method is applied to the analysis of the waste production in Venice province. Also in this applied case, the proposed method can properly take into account the complex shape of the domain and is able to capture the temporal evolution of the phenomenon.

In Chapter 2, we consider data characterized by stationary spatial anisotropy. In this case, the approach proposed by Ramsay (2002) and Sangalli et al. (2013) is not applicable since the Laplacian is an isotropic operator and therefore the resulting smoothing effect would not capture the spatial structure present in the data. The method we propose makes use of an anisotropic diffusion operator in the regularization as to properly take into account the spatial anisotropy present in the data. Differently from Azzimonti et al. (2015) and Azzimonti et al. (2014), we do not assume that the anisotropic diffusion operator in the regularizing term is suggested by a priori knowledge of the problem under analysis; rather, we consider it as unknown and we estimate it from the data. To this end, we adopt a profiling estimation approach, adapting to our setting the parameter cascading technique proposed by Ramsay et al. (2007), Cao and Ramsay (2007), Cao and Ramsay (2009), Cao and Ramsay (2010) and Xun et al. (2013). Simulation studies compare the proposed method to both isotropic and anisotropic kriging in various simulation settings. The method is illustrated via an application to a benchmark dataset concerning rainfall data over Switzerland, that was used for the Spatial Interpolation Comparison 97 (Dubois et al., 2003). Also in this applied case, the technique is able to capture the right anisotropy, thus providing a good estimated spatial field. This chapter is based on the submitted work Bernardi et al. (2017), developed in collaboration with Prof. James O. Ramsay and Dr. Michelle Carey.

Chapter 3, presents the first steps for the development of the method for the analysis of data characterized by non-stationary spatial anisotropy induced by the texture of the domain. In this case, the domain of the phenomenon is not homogeneous, but presents regions with different characteristics which determine a non-homogeneous spatial dependence of the data. The motivating applied problem concerns the estimation of the population mobility in the metropolitan area of Milan from mobile phone data (Manfredini et al., 2015; Secchi et al., 2015; Zanini et al., 2016; Passamonti, 2016). The domain under study is strongly characterized by the presence of roads, which affect the distribution of the data. Indeed, the phenomenon is characterized by strong anisotropy along the main roads, while it is more homogeneous in the regions without main roads. The method we propose accounts for the non-homogeneity in the spatial dependence of the data by considering, in the penalization term, a non-stationary anisotropic diffusion operator whose anisotropy is estimated from the geometry of the road network characterizing the domain. The estimation of the anisotropy is performed with the technique presented in Della Rossa et al. (2010), where an algorithm is developed to model traffic flows in complex networks such as large urbanized areas. This work is in collaboration with Prof. Fabio Della Rossa.

Chapter 1

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

This chapter is based on Bernardi et al. (2016).

1.1 Introduction

In this chapter 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 1.1 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.

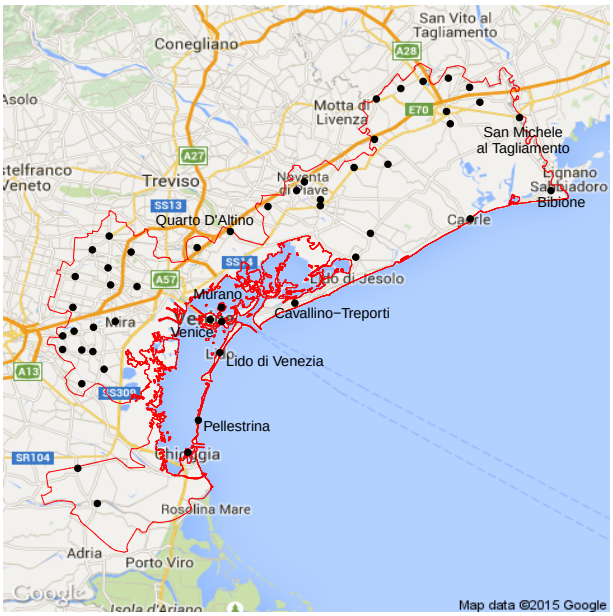

Figure 1.1: Spatial domain of the Venice waste data, with a red line highlighting the province boundary and dots indicating the towns centers.

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 1.2 shows the temporal evolution of the production of per capita waste in the towns of Venice province and Figure 1.3 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 1.3. The temporal evolution of the production of per capita municipal waste in the two towns, highlighted in Figure 1.2, 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 1.7,

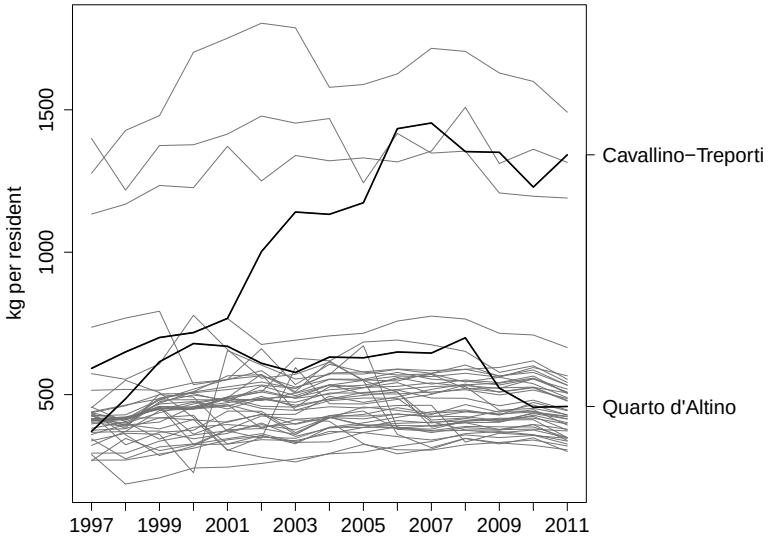

Figure 1.2: Temporal evolution of the yearly per capita production (kg per resident) of municipal waste in the towns of Venice province.

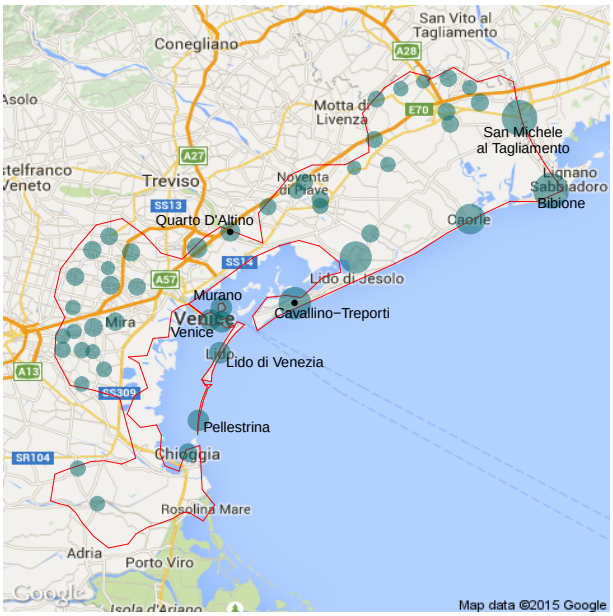

Figure 1.3: Per capita production (kg per resident) of municipal waste in the towns of Venice province in 2006. The data include all municipalities of Venice province and additional four localities (Bibione, Murano, Lido di Venezia and Pellestrina), that do not constitute a municipality on their own, 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 1.7).

this is due to the fact that we miss the information concerning the urbanized 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, 2016), based on the package fdaPDE (Lila et al., 2016).

This chapter is organized as follows. Section 1.2 describes the ST-PDE model. Section 1.3 shows the numerical implementation of the model. Section 1.4 illustrates the extension of the model for the inclusion of space-time varying covariates. Section 1.5 describes a variation of the ST-PDE model. Section 1.6 compares via simulation studies the ST-PDE model to other spatio-temporal prediction techniques. Section 1.7 shows the application of the proposed method to the analysis of the per capita municipal waste in the Venice province. Section 1.8 outlines some possible model extensions.

1.2 Data and model

Let $\{\mathbf{p}_i = (x_i, y_i); i = 1, ..., n\}$ 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 $\{t_j; j = 1, ..., m\}$ 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 Ω 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 Ω.

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

$$z_{ij} = f(\mathbf{p}_i, t_j) + \epsilon_{ij} \qquad i = 1, ..., n,\ j = 1, ..., m, \tag{1.1}$$

where $\{\epsilon_{ij}; i = 1, ..., n; j = 1, ...m\}$ are independently distributed residuals with mean zero and constant variance σ^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. Here, 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

$$J_S\left(g(\mathbf{p})\right) = \int_\Omega \Big(\Delta g(\mathbf{p})\Big)^2 d\mathbf{p}, \tag{1.2}$$

where $g : \Omega \to \mathbb{R}$ 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

$$J_T\left(h(t)\right) = \int_{T_1}^{T_2} \Big(\frac{d^r h(t)}{dt^r}\Big)^2 dt\,, \tag{1.3}$$

where $h : [T_1, T_2] \to \mathbb{R}$. 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 Ω. The field f is thus estimated by minimizing the following penalized sum of square

error criterion:

$$J(f) = \sum_{i=1}^{n}\sum_{j=1}^{m}\big(z_{ij} - f(\mathbf{p}_i, t_j)\big)^2 + \\ + \lambda_S \int_{T_1}^{T_2}\int_{\Omega}\Big(\Delta f(\mathbf{p},t)\Big)^2 d\mathbf{p}dt + \lambda_T \int_{\Omega}\int_{T_1}^{T_2}\Big(\frac{\partial^r f(\mathbf{p},t)}{\partial t^r}\Big)^2 dtd\mathbf{p}, \quad (1.4)$$

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 1.3.3. As detailed in the following section, the regularizing terms in (1.4) induce the space-time covariance structure of the estimator and different regularizations would imply different covariance structures.

1.3 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 $\{\varphi_k(t); k = 1, ..., M\}$ be a set of M basis functions defined on $[T_1, T_2]$ and $\{\psi_l(\mathbf{p}); l = 1, ..., N\}$ a set of N basis functions defined on Ω. Then, f is represented by the following basis expansion:

$$f(\mathbf{p}, t) = \sum_{l=1}^{N}\sum_{k=1}^{M} c_{lk}\ \psi_l(\mathbf{p})\ \varphi_k(t), \quad (1.5)$$

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

1.3.1 Choice of the basis systems in space and time

Various possible bases can be used for the expansions in the spatial and temporal domains. Here, we use in space a finite element basis on a triangulation Ω_τ of the spatial domain Ω 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 1.1, 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, sub-sampling from this smooth curve so that the features characterizing the domain are preserved. Figure 1.4 shows the simplified boundary of Venice province, while Figure 1.5 shows the detail around the city of Venice. This region is particularly interesting

Figure 1.4: Simplified boundary of the Venice province.

Figure 1.5: Detail of the simplified boundary of the Venice province.

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 1.6 displays the resulting triangulation of Venice province. For this application, here and in Section 1.7, 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.

Figure 1.6: 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 1.7 shows an example of linear basis function.

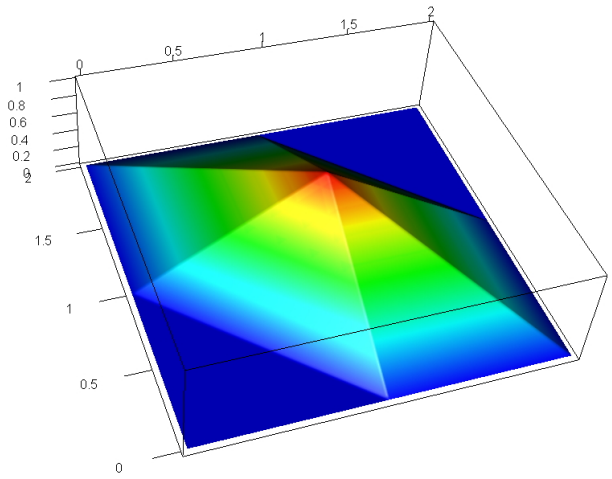

Figure 1.7: Example of linear finite element basis function.

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 (1.3).

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.

1.3.2 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 (1.5) of the spatio-temporal field f, with entries

ordered as follows

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

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

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

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 Ω_τ of the cross products of the N spatial basis, i.e.,

$$R_0 = \int_{\Omega_\tau} \boldsymbol{\psi}\boldsymbol{\psi}^T. \tag{1.6}$$

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

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

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.,

$$K_0 = \int_{T_1}^{T_2} \boldsymbol{\varphi}\boldsymbol{\varphi}^T . \tag{1.7}$$

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 (1.4) as

$$\begin{aligned} J &= (\mathbf{z} - B\mathbf{c})^T(\mathbf{z} - B\mathbf{c}) + \lambda_S \mathbf{c}^T(P_S \otimes K_0)\mathbf{c} + \lambda_T \mathbf{c}^T(R_0 \otimes P_T)\mathbf{c} \\ &= (\mathbf{z} - B\mathbf{c})^T(\mathbf{z} - B\mathbf{c}) + \mathbf{c}^T P \mathbf{c} , \end{aligned} \tag{1.8}$$

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 (1.3):

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

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 (1.2), 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 (1.2) and (1.4) and involving the computation of second order derivatives.

Finally, the coefficients vector $\hat{\mathbf{c}}$ that minimizes the functional J in (1.8) is computed deriving J with respect to $\mathbf{c}$ and setting the derivative equal to 0, obtaining

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

1.3.3 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 $Var[\mathbf{z}] = \sigma^2 I_{nm}$, we obtain

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

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

$$\begin{aligned} E[\hat{f}(\mathbf{p}, t)] &= \mathbf{B}(\mathbf{p}, t)(B^T B + P)^{-1} B^T \mathbf{f} \\ Var[\hat{f}(\mathbf{p}, t)] &= \sigma^2 \mathbf{B}(\mathbf{p}, t)(B^T B + P)^{-1} B^T B (B^T B + P)^{-1} \mathbf{B}(\mathbf{p}, t)^T. \end{aligned} \tag{1.9}$$

The regularizing terms in (1.4) induce the space-time covariance structure of the estimator, given by

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

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 (1.1)-(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 σ^2 by

$$\hat{\sigma}^2 = \frac{1}{nm - \mathrm{tr}(S)} (\mathbf{z} - \hat{\mathbf{z}})^T (\mathbf{z} - \hat{\mathbf{z}}). \tag{1.10}$$

This estimate of the error variance, plugged into (2.14), 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 λ_S and λ_T may be chosen via Generalized Cross-Validation (GCV), searching for the values of λ_S, λ_T that minimize

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

1.4 Model with covariates

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

$$z_{ij} = \mathbf{w}_{ij}^T\, \boldsymbol{\beta} \;+\; f(\mathbf{p}_i, t_j) \;+\; \epsilon_{ij} \qquad i = 1, ..., n, \quad j = 1, ..., m, \tag{1.11}$$

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

$$\begin{aligned} J(f, \boldsymbol{\beta}) = &\sum_{i=1}^{n}\sum_{j=1}^{m}\big(z_{ij} - \mathbf{w}_{ij}^T\, \boldsymbol{\beta} - f(\mathbf{p}_i, t_j)\big)^2 + \\ &+ \lambda_S \int_{T_1}^{T_2}\int_{\Omega}\Big(\Delta f(\mathbf{p}, t)\Big)^2 d\mathbf{p}dt + \lambda_T \int_{\Omega}\int_{T_1}^{T_2}\Big(\frac{\partial^2 f(\mathbf{p}, t)}{\partial t^2}\Big)^2 dtd\mathbf{p}\,. \end{aligned} \tag{1.12}$$

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

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

Let H_W be the matrix that projects orthogonally on the space generated by the columns of W, i.e. $H_W = W(W^TW)^{-1}W^T$ and let $Q = I_{nm} - H_W$. The discretization of the functional $J(f, \boldsymbol{\beta})$ in (1.12) 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:

$$\begin{aligned} \hat{\boldsymbol{\beta}} &= (W^T W)^{-1} W^T (\mathbf{z} - B\hat{\mathbf{c}}), \\ \hat{\mathbf{c}} &= (B^T Q B + P)^{-1} B^T Q \mathbf{z}. \end{aligned}$$

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.

1.4.1 Properties of the estimator

Let $S_{\mathbf{f}} = B(B^T Q B + P)^{-1} B^T Q$, so that

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

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

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

and

$$\begin{aligned} E[\hat{\boldsymbol{\beta}}] &= \boldsymbol{\beta} + (W^T W)^{-1} W^T (I_{nm} - S_{\mathbf{f}})\mathbf{f}, \\ Var[\hat{\boldsymbol{\beta}}] &= \sigma^2 (W^T W)^{-1} + \sigma^2 (W^T W)^{-1} W^T S_{\mathbf{f}} S_{\mathbf{f}}^T W (W^T W)^{-1}. \end{aligned} \tag{1.13}$$

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

$$S = H_W + QS_{\mathbf{f}}.$$

The trace of this matrix is given by $tr(S) = q + tr(S_f)$ 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 $tr(S_f)$ degrees of freedom corresponding to the non-parametric part of the model. We can estimate σ^2 as in (1.10). 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 (1.13), 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.

1.5 A variation of the proposed ST-PDE model

A variation of the proposed ST-PDE model can be defined as follows. Equivalently to (1.5), we can express the spatio-temporal field f in the following basis expansions:

$$f(\mathbf{p},t) = \sum_{k=1}^{M} a_k(\mathbf{p})\ \varphi_k(t) \tag{1.14}$$

$$f(\mathbf{p},t) = \sum_{l=1}^{N} b_l(t)\ \psi_l(\mathbf{p}), \tag{1.15}$$

where $\{a_k(\mathbf{p}); k = 1, ..., M\}$ are the spatially varying coefficients of the expansion of the space-time field on the temporal basis and $\{b_l(t); l = 1, ..., N\}$ 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 (1.14), and the temporal penalty J_T to the N temporally varying coefficients $b_l(t)$ in the basis expansion (1.15). In this case, the field f is thus estimated by minimizing the following penalized sum of square error criterion:

$$\tilde{J}(f) = \sum_{i=1}^{n}\sum_{j=1}^{m}\big(z_{ij} - f(\mathbf{p}_i,t_j)\big)^2 + \\ + \lambda_S \sum_{k=1}^{M}\int_\Omega \Big(\Delta(a_k(\mathbf{p}))\Big)^2 d\mathbf{p} + \lambda_T \sum_{l=1}^{N}\int_{T_1}^{T_2}\Big(\frac{d^r b_l(t)}{dt^r}\Big)^2 dt\,. \tag{1.16}$$

The numerical implementation of this method has only minor differences with the one presented in Section 1.3. Denote by I_d the identity matrix of dimension d. Then the discretized functional is given by

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

where the identity matrices I_M and I_N have replaced the matrices K_0 in (1.7) and in R_0 in (1.6), respectively. Minimizing the functional (1.16) is thus equivalent to minimizing the one in (1.4) if the spatial and the temporal basis used for the numerical discretization are orthonormal. Here, we use basis systems which are not orthonormal; nevertheless, the basis systems considered are sparse, so that the terms

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

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 1.6 obtaining the same performances for the two methods in all cases. In the following we show only the results for the ST-PDE method.

1.6 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 (1.16), where the spatial penalty is replaced by the thin plate spline energy recalled in Section 1.2. 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 (1.16). 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 λ_S, λ_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 1.8, 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.

1.6.1 First simulation study

We sample 200 spatial locations uniformly in the C-shaped domain at 9 time instants equally spaced from 0 to π. We simulate the data from model (1.1), with a gaussian noise with mean 0 and standard deviation 0.7.

Figure 1.8 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).

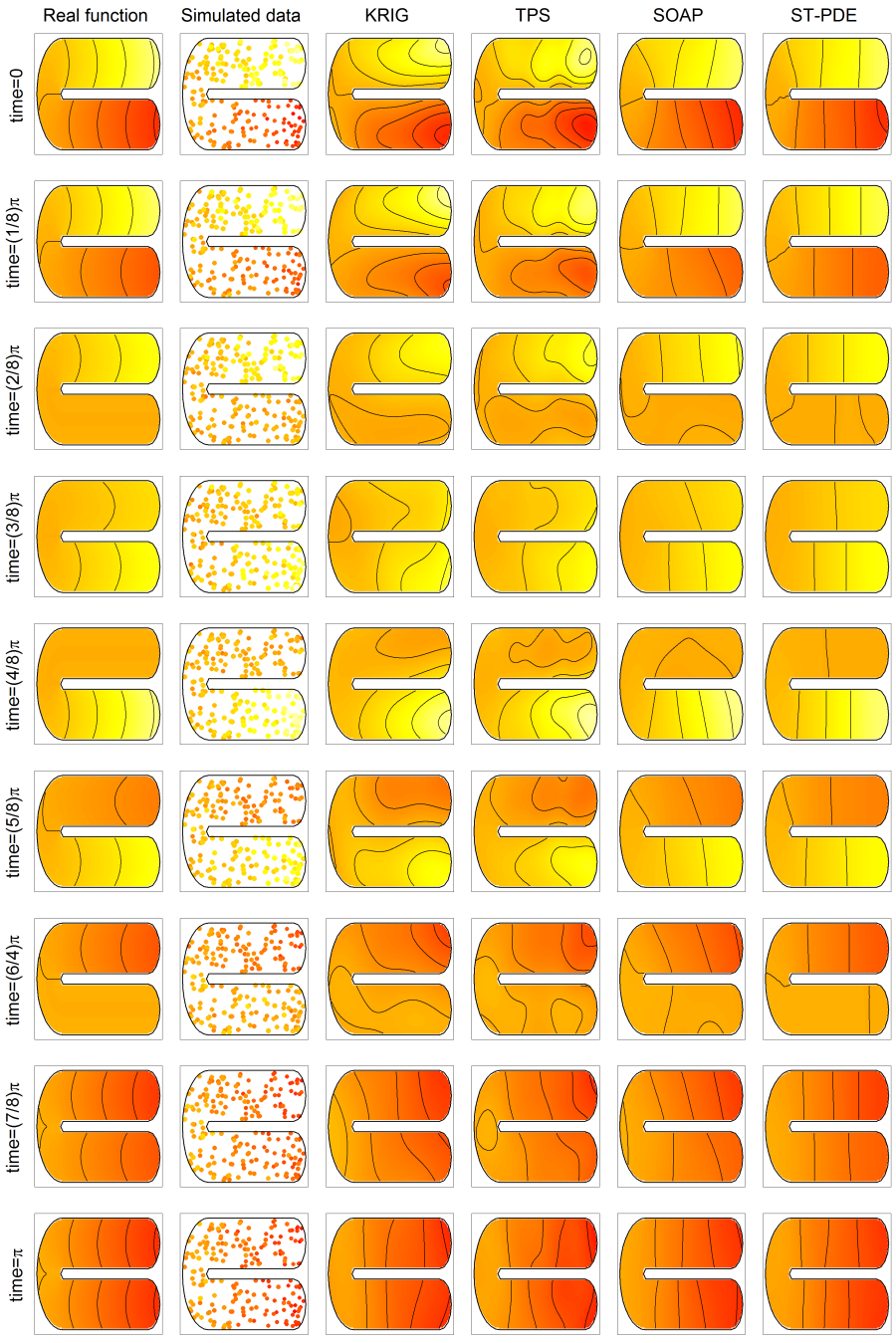

Figure 1.8: The spatio-temporal test function, the simulated data and the estimated functions with spatio-temporal kriging (KRIG), space-time model using thin plate spline (TPS), space-time model using soap film smoothing (SOAP) and ST-PDE.

Figure 1.9 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 1.8. 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.

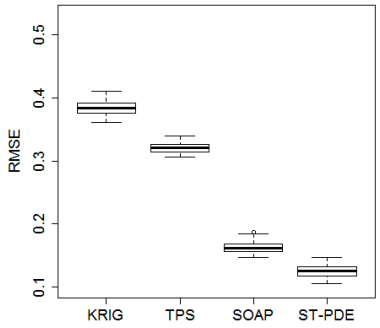

Figure 1.9: First simulation study. Boxplots of the RMSE of the estimates of the spatio-temporal field obtained by the four tested methods: spatio-temporal kriging (KRIG), space-time model using thin plate spline (TPS), space-time model using soap film smoothing (SOAP) and ST-PDE.

1.6.2 Second simulation study: fewer spatial locations and more time instants

In this second simulation case we perform the same simulation study described in Section 1.6.1, 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 1.10 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.

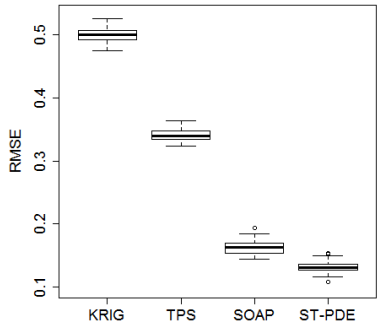

Figure 1.10: Second simulation study: fewer spatial locations and more time instants. Boxplots of the RMSE of the estimates of the spatio-temporal field obtained by the four tested methods.

1.6.3 Third simulation study: covariates

Within the same simulation setting described in Section 1.6.1, we also perform a study including a space-time varying covariate. Specifically, the covariate used is the following space-time function:

$$w(x, y, t) = \cos(5x)\sqrt{\frac{t+1}{5}}. \tag{1.17}$$

We generate data from model (1.11), setting $\beta = 1$. The other simulation specifications are as in Section 1.6.1. 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 1.11 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 1.6.1. 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 β 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 β associated to the ST-PDE estimate is given by $[0.88, 1.16]$.

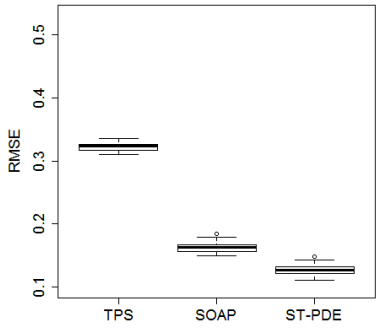

Figure 1.11: Third simulation study: covariates. Boxplots of the RMSE of the estimates of the spatio-temporal field obtained by the three methods tested in this case: TPS, SOAP, ST-PDE.

1.6.4 Fourth simulation study: correlated noise

We performed a fourth simulation study within the same setting described in Section 1.6.1, but generating the additive noise with a spatio-temporal covariance structure. We use the function `RFsimulate` of the R package RandomFields (Schlather et al., 2016) 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

$$C(r) = \sqrt{2}rK_1\left(\sqrt{2}r\right), \tag{1.18}$$

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 1.12 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 1.6.1.

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

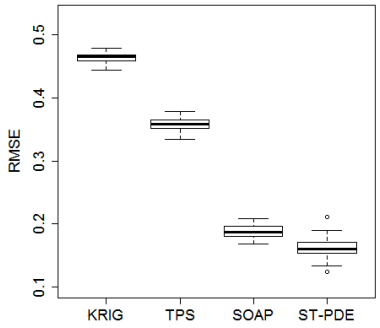

Figure 1.12: Fourth simulation study: correlated noise. Boxplots of the RMSE of the estimates of the spatio-temporal field obtained by the four tested methods.

1.6.5 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 1.8, 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 (0.2×0.2) spatial subdomains D_i uniformly distributed over the spatial C-shaped domain and we compute their centers $\mathbf{p}_i$. Figure 1.13 represents the spatial subdomains and their centers. We consider 9 time instants $\{t_1, ..., t_9\}$ equally spaced from 0 to π. We compute the data in the following way:

$$z_{ij} = \frac{1}{|D_i|} \int_{D_i} f(\mathbf{p}, t_j) d\mathbf{p} + \epsilon_{ij} \qquad i = 1, ..., 50, \; j = 1, ..., 9,$$

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.

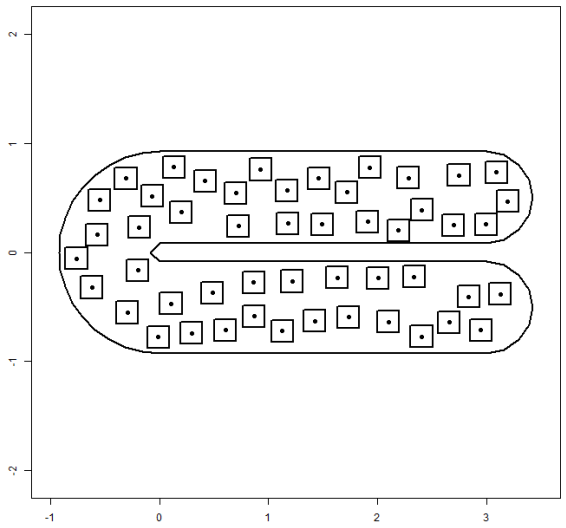

Figure 1.13: Spatial subdomains.

Figure 1.14 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 1.6.1.

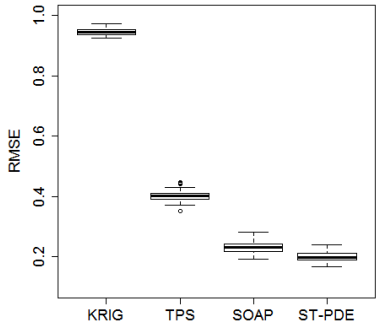

Figure 1.14: Fifth simulation study: areal data assigned to area centroid. Boxplots of the RMSE of the estimates of the spatio-temporal field obtained by the four tested methods: spatio-temporal kriging (KRIG), space-time model using thin plate spline (TPS), space-time model using soap film smoothing (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.

1.7 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.

1.7.1 The Venice waste dataset

Open Data Veneto[1] 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

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

spatial coordinates of the town centers are available online[2]. As mentioned in Section 1.3, 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 1.1. 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[3], the Italian national institute for statistics.

1.7.2 Analysis of Venice waste data by ST-PDE

Figure 1.15 shows the estimated spatio-temporal field at fixed time instants. The estimate for the coefficient β 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

[2]http://www.dossier.net/utilities/coordinate-geografiche/
[3]http://www.istat.it/it/archivio/113712

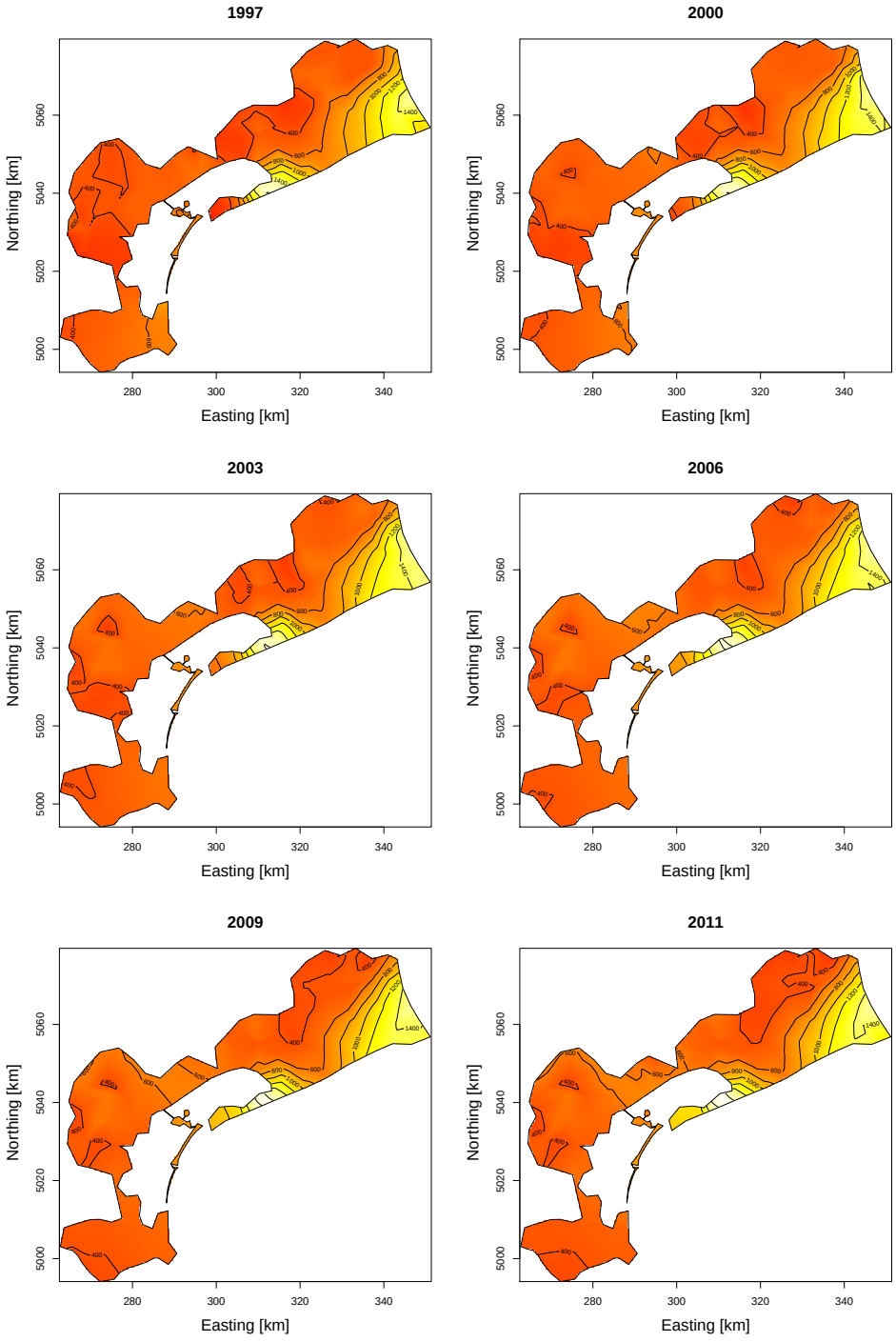

Figure 1.15: Estimated spatio-temporal field for the Venice waste data (yearly per capita production) at fixed time instants.

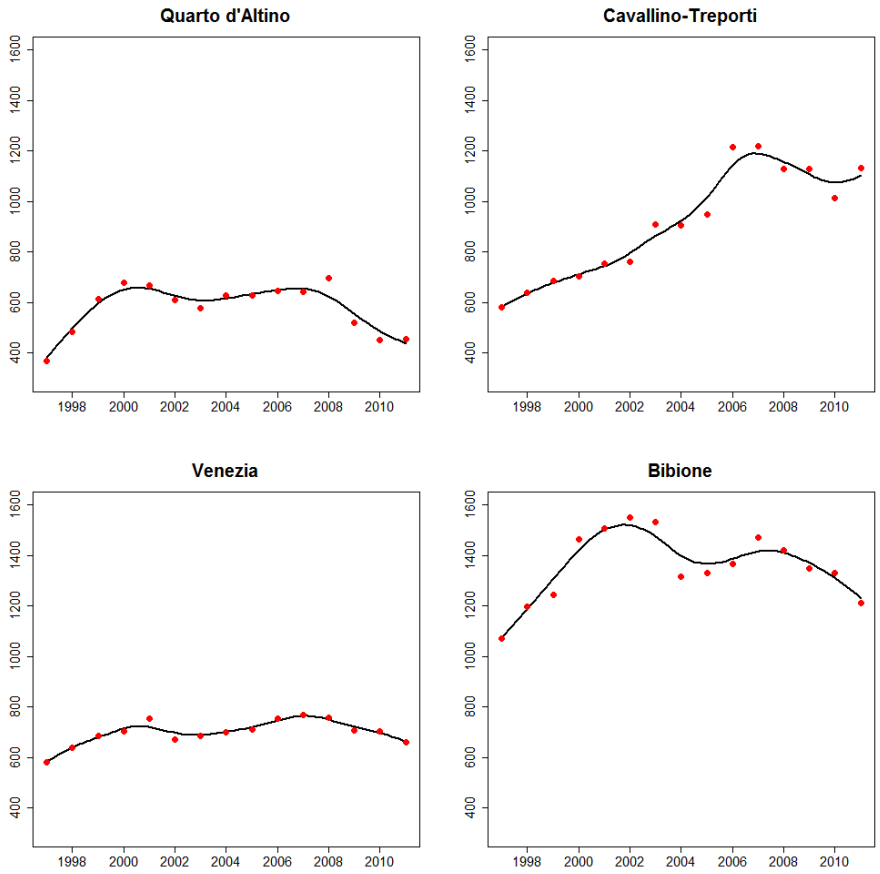

Figure 1.16: Temporal evolution of the estimated spatio-temporal field for the Venice waste data (yearly per capita production) at fixed spatial locations.

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 1.16, 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 1.16 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.

1.8 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 consists 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 and Sangalli (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.

Appendix

1.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:

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

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 $q = \frac{\pi r}{2}$.

Chapter 2

Modelling spatial anisotropy via regression with partial differential regularization

This chapter is based on the submitted work Bernardi et al. (2017).

2.1 Introduction

Many, if not most, spatial phenomena are characterized by spatial anisotropy. Just to name a few examples, in biology, anisotropy is naturally induced by the arrangement and orientation of fibers and cells in a tissue, or by the morphology of the organs; in meteorology, it may be caused by the presence of winds and sea streams, or by the orography of the region under study; in geology, by the process of sedimentation. A simple illustrative example is displayed in Figure 2.1, that pictures a dataset of 467 daily rainfall measurements made in Switzerland on the 8th of May 1986; this dataset was used for the Spatial Interpolation Comparison 97 (Dubois et al., 2003). The size and color of point markers represent the value of the rainfall at each location, highlighting a strong spatial anisotropy, with higher rainfall values alternating with lower rainfall values along elongated regions oriented in the northeast-southwest direction.

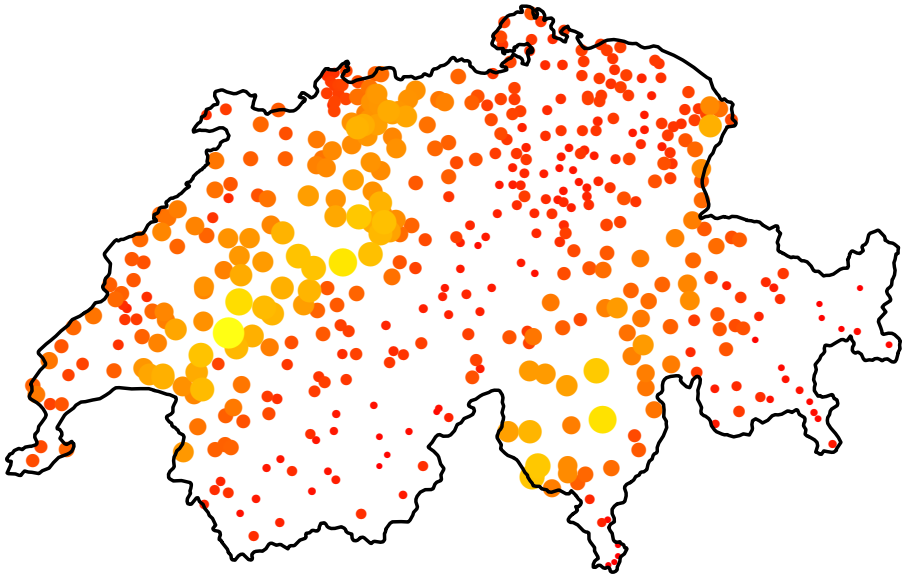

Figure 2.1: Switzerland rainfall data.

Accurately modelling and estimating the anisotropy in the variable of interest, when this cannot be well described by the available covariates, is not trivial and most statistical analyses of spatially distributed data are carried out assuming the isotropy of the spatial field. For some examples of models which take into account anisotropy see for example Ecker and Gelfand (2003) Budrikaité and Dučinskas (2005) Blanchet and Davison (2011).

Here we adopt a functional data analysis approach and propose to model spatial anisotropy via regression with partial differential regularization. Ramsay (2002), Wood et al. (2008) and Sangalli et al. (2013) consider spatial regression with a roughness penalty that involves the Laplacian of the spatial field: this partial differential operator provides a simple and isotropic measure of the curvature of the spatial field, and its use in the regularizing term induces an isotropic smoothing effect. Thin plate splines (Wahba, 1990) and bivariate splines over triangulations (Guillas and Lai, 2010) offer other classical and recent proposals of spatial isotropic smoothing defined as penalized regression with differential regularization. Azzimonti et al. (2015) extend the method in Ramsay (2002) and Sangalli et al. (2013) to the case where the regularizing term involves a more general partial differential equation (PDE) that induces an anisotropic and non-stationary smoothing. Azzimonti et al. (2015) assume that the PDE is suggested by prior knowledge of the phenomenon under study, coming for instance from the physics or morphology of the problem; the parameters of the PDE are consequently fixed considering their physical meaning.

We here instead assume no prior knowledge of the spatial variation of the

considered problem, nor do we assume the existence of a physical law governing the system, but we use the PDE in the regularizing term to model the spatial variation of the phenomenon, learning the anisotropy from the data. Specifically, the PDE in the regularizing term induces an anisotropic stationary smoothing effect; the parameters in the PDE determine the direction and the intensity of the anisotropy; these parameters are here considered unknown and are estimated from data. As in Ramsay (2002), Sangalli et al. (2013) and Azzimonti et al. (2015), we here represent the spatial field via finite elements, which provide a basis for piecewise polynomial surfaces.

Moreover, we estimate the parameters of the PDE in the regularizing term by parameter cascading. This generalized profiling estimation procedure was originally introduced by (Ramsay et al., 2007) to retrieve the parameters of an ordinary differential equation (ODE), starting from noisy measurements of the ODE solution. This technique has been successfully applied in other contexts, such as penalized smoothing (Cao and Ramsay, 2007, 2009), dynamical models (Cao et al., 2008) and linear mixed-effects modeling (Cao and Ramsay, 2010). Xun et al. (2013) extended this technique to estimate the parameters of a PDE, again starting from noisy measurements of the PDE solution. In our case, we use parameter cascading to obtain the parameters of a PDE in a more general setting, where the data do not come from the solution of the PDE itself. The estimated PDE does not model the phenomenon which generated the data, but rather is used as a regularizing term to characterize the spatial distribution of the data. Moreover, spatially varying covariates can be included in the model.

We test the performance of anisotropic spatial regression with PDE regularization (anisotropic SR-PDE) via extensive simulation studies, comparing it to both isotropic and anisotropic kriging, with various variogram models, and to the isotropic SR-PDE method described in Ramsay (2002) and Sangalli et al. (2013). The comparative studies show that, when simulating from an anisotropic Matérn field, anisotropic SR-PDE has significant lower root mean square error (RMSE) than isotropic SR-PDE, and it also has significantly lower RMSE than isotropic and anisotropic kriging. In particular, anisotropic SR-PDE provides better estimates with respect to anisotropic Matérn kriging, that should constitute the best possible model as it assumes the same space covariance structure used to generate the data. Furthermore, also when space-varying covariates are added, the proposed anisotropic SR-PDE provides significantly better estimates than all other considered methods, including anisotropic universal Matérn kriging. Moreover, anisotropic SR-PDE also outperforms all other methods when generating from a field defined over an irregularly shaped domain.

The chapter is organized as follows. Section 2.2 describes the model. Section 2.3 provides implementation details. Section 2.4 extends the model for the inclusion of space-varying covariates. Section 2.5 is devoted to simulation studies. Section 2.6 shows the application of the proposed method to the analysis of the

Switzerland rainfall data. Finally, Section 2.7 draws some directions for future research.

2.2 Model

Let $\{\mathbf{p}_i = (x_i, y_i); i = 1, \ldots, n\}$ be a set of n points on a bounded domain $\Omega \in \mathbb{R}^2$, whose boundary $\partial\Omega$ is a curve of class $\mathcal{C}^2$. We denote by $|\Omega|$ the area of the domain Ω. Let $z_i \in \mathbb{R}$ be the value of a variable of interest observed at point $\mathbf{p}_i$. We assume that $\{z_i; i = 1, \ldots, n\}$ are noisy observations of an underlying smooth function $f : \Omega \to \mathbb{R}$:

$$z_i = f(\mathbf{p}_i) + \epsilon_i \quad i = 1, \ldots, n, \tag{2.1}$$

where $\{\epsilon_i; i = 1, \ldots, n\}$ are independently distributed residuals, with mean zero and constant variance σ^2.

We want to estimate the spatial field f by minimizing the following penalized sum-of-square-error functional:

$$J_\rho(f, K) = (1 - \rho)\frac{1}{n}\sum_{i=1}^{n}\left(z_i - f(\mathbf{p}_i)\right)^2 + \rho\frac{1}{|\Omega|}\int_\Omega \left(\nabla \cdot (K\nabla f)\right)^2, \tag{2.2}$$

where the operator ∇ is defined as $\nabla = (\frac{\partial}{\partial x}, \frac{\partial}{\partial y})^T$ and K is a symmetric and positive definite matrix. The inclusion of the anisotropic diffusion operator $\nabla \cdot (K\nabla f)$ in the regularizing term provides an anisotropic smoothing effect, where the anisotropy is determined by the matrix K. Indeed, the eigenvectors of K fix the directions of maximum and minimum smoothing, while the corresponding eigenvalues set the intensity of the smoothing in each direction, with the ratio between the two eigenvalues determining the intensity of the anisotropy. Figure 2.2 illustrates the spatial regularization implied by three different matrices K, represented via ellipses whose axes are oriented according to the eigenvectors of K and have length proportional to the corresponding eigenvalues. In the case considered in the left panel of Figure 2.2, K is the identity matrix, and the model reduces to the one presented in Ramsay (2002) and Sangalli et al. (2013), where the penalty involves the operator $\Delta f = \nabla \cdot \nabla f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2}$, that is the Laplacian operator. In this case, the smoothing is isotropic: the penalization equally weights the curvature of the spatial field along every direction. The cases considered in the central and right panels of Figure 2.2 provide examples of anisotropic penalizations with different angles and intensities. The smoothing along the direction of the major axis has a longer range, while the smoothing along the minor axis has a shorter range. This effect is more pronounced in case represented in the right panel.

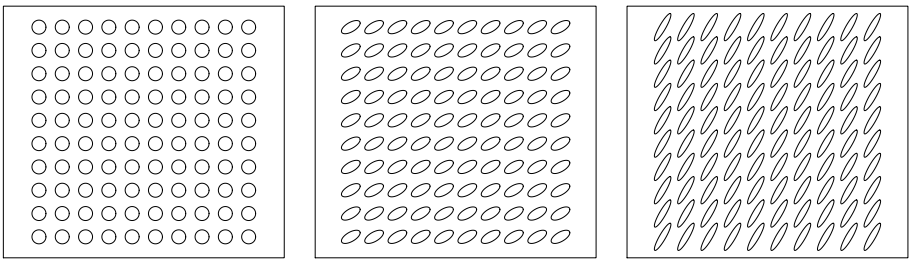

Figure 2.2: Graphical representation of the smoothing effect implied by stationary isotropic (left) and anisotropic (center and right) penalizations. The anisotropy matrix K is represented by ellipses: its eigenvectors (the ellipses axes) identify the direction of the anisotropy, while the eigenvalues (proportional to the axes lengths) determine the intensity of the anisotropy.

The smoothing parameter $\rho \in (0, 1)$ weights the contribution of the data fitting term (the least squares) against the regularization term, in the estimation of the spatial field. The higher ρ is, the more we control the roughness of the spatial field f through the anisotropic diffusion operator, forcing the estimated spatial field to be characterized by the anisotropy defined by K; the smaller the ρ, the more we locally adapt to the data. The way we parametrize the smoothing parameter in this work is different with respect to the one considered in Ramsay (2002), Sangalli et al. (2013) and Azzimonti et al. (2015). The new parametrization of the smoothing parameter in the functional (2.2) is here introduced to more efficiently balance the least squares term and the regularizing term, and consequently improve the optimization of the functional.

We want to solve the non-convex optimization problem of minimizing the functional $J_\rho(f, K)$ with respect to f and K, with a value of ρ selected to properly weight the effect of the regularization and to accurately estimate the anisotropy matrix K. We here propose a two-step algorithm which estimates separately K and f. The anisotropy matrix K is estimated using a parameter cascading approach, which is described in Section 2.2.1. The estimation of the spatial field f is conditional on K. The problem of estimating f, minimizing the functional $J_\rho(f, K)$ for a fixed value of K, has a unique solution (see Section 2.2.2); this infinite-dimensional problem can be discretized via finite elements (see Sections 2.2.3 and 2.2.4).

2.2.1 Estimation of the anisotropy matrix K via parameter cascading

We here show how we use parameter cascading for the estimation of the anisotropy matrix K. As mentioned in the Introduction, Ramsay et al. (2007) and Xun et al. (2013) use parameter cascading to respectively estimate the parameters of an ordinary differential equation and of a partial differential equation, starting from noisy observations of the solution of the differential equation itself. In our case,

the data do not come from the solution of the PDE, but the PDE is used as a convenient model to describe the spatial anisotropy characterizing the distribution of the data. Our final goal is the accurate fit of the spatial field. The different focus and the different setting considered here require some modification of the implementation of parameter cascading with respect to (Ramsay et al., 2007) and Xun et al. (2013). This will be further detailed in Section 2.3.2.

The parameter cascading algorithm distinguishes two classes of parameters to estimate: the *structural parameters*, which are the parameters of direct interest for the analysis, and the *nuisance parameters*, which are essential for fitting the data, but are not of direct interest. In our setting, when estimating the anisotropy matrix K, the structural parameter is K and the nuisance parameters are the coefficients of the expansion used to represent f. The estimates of the nuisance parameters are obtained minimizing the functional $J_\rho(f, K)$ with respect to f for a fixed K. The solution is called $\hat{f}_K$ and it is an implicit function of the structural parameter: each time K is changed, the functional $J_\rho(f, K)$ is re-optimized with respect to f alone. The structural parameter is then obtained by the optimization of the functional

$$H(K) = \sum_{i=1}^{n} \left(z_i - \hat{f}_K(\mathbf{p}_i) \right)^2, \tag{2.3}$$

which depends on K implicitly, through $\hat{f}_K$. Since $\hat{f}_K$ is already regularized, H does not include the regularization term and is simply the data fitting criterion.

The minimization of the criterion $H(K)$ does not have a closed-form solution, so its optimization is performed numerically. The minimization of the criterion $J_\rho(f, K)$ with respect to f, for a fixed K, can instead be characterized similarly to Sangalli et al. (2013) and Azzimonti et al. (2015), as detailed in the following sections.

2.2.2 Estimation of the spatial field f given the anisotropy matrix K

Let $H^m(\Omega)$ be the Hilbert space of functions belonging to $L^2(\Omega)$ along with all their distributional derivatives up to order m. The functional $J_\rho(f, K)$ in (2.2) is well posed for $f \in H^2(\Omega)$ with appropriate boundary conditions. Given a symmetric and positive definite matrix K, the minimizer of $J_\rho(f, K)$ exist and is unique; boundary conditions ensure the uniqueness of the solution. Various boundary conditions may be considered; for example, Dirichlet conditions impose the value of f at the boundary $\partial\Omega$, while Neumann conditions impose the flux of the function through the boundary. In the following, we consider homogeneous Neumann boundary conditions (i.e., null value of the normal derivative of f at the boundary, meaning null flux through the boundary). See Azzimonti et al. (2015) for the case of general boundary conditions.

We denote by $V(\Omega)$ the subspace of $H^2(\Omega)$ characterized by the chosen boundary conditions. We set $\mathbf{z} = (z_1, \ldots, z_n)^T$ and, for any function $h : \Omega \to \mathbb{R}$, we set $\mathbf{h}_n = (h(p_1), \ldots, h(p_n))^T$. The estimator is characterized by the following proposition.

Proposition 2.1. *Given a symmetric and positive definite matrix K, there exists a unique estimator $\hat{f} \in V(\Omega)$ which minimizes (2.2). Moreover, $\hat{f}$ satisfies*

$$(1-\rho)\frac{1}{n}\mathbf{h}_n^T\hat{\mathbf{f}}_n + \rho\frac{1}{|\Omega|}\int_\Omega (\nabla\cdot(K\nabla h))(\nabla\cdot(K\nabla\hat{f})) = (1-\rho)\frac{1}{n}\mathbf{h}_n^T\mathbf{z} \tag{2.4}$$

for every $h \in V(\Omega)$.

Proof. The result follows from *Proposition 1* in Azzimonti et al. (2015), by appropriate reparametrization of the smoothing parameter, and by setting $L = \nabla \cdot K\nabla f$ and $u = 0$. □

The fourth-order problem (2.4) can be rewritten as a coupled system of second order problems. Then, using Green's theorem (integration by parts), it is possible to obtain an equivalent reformulation of the problem that involves only first order derivatives. See the Appendix for details. This reformulation of the problem can thus be discretized via finite elements. This enables to approximate the infinite dimensional estimation problem by a finite dimensional one, and to reduce the estimation problem to the solution of a linear system.

2.2.3 Finite elements

To construct a finite element space, we first obtain a regular triangulation $\mathcal{T}$ of the domain Ω where adjacent triangles share either a vertex or a complete edge. Thus, the domain Ω is approximated by the union of all triangles $\Omega_{\mathcal{T}}$, and the boundary $\partial\Omega$ is approximated by a polygon (or multiple polygons in case of a domain with holes).

As an example, Figure 2.3 shows the triangulation for the Switzerland rainfall data. As a preprocessing step, we obtain a simplified border for the data in order to have a small number of points defining the boundary of the domain and, therefore, a simpler triangulation. We can discard the precise definition of the boundary since, in this application, the border has no influence on the phenomenon under study: rainfall is not correlated with the political borders of Switzerland. The simplification of the boundary is shown in the left panel of Figure 2.3, where the blue line represents the actual border of Switzerland, the red line is the simplified boundary we use in our analysis and the black dots represent the data locations. The final triangulation is shown in the right panel of Figure 2.3.

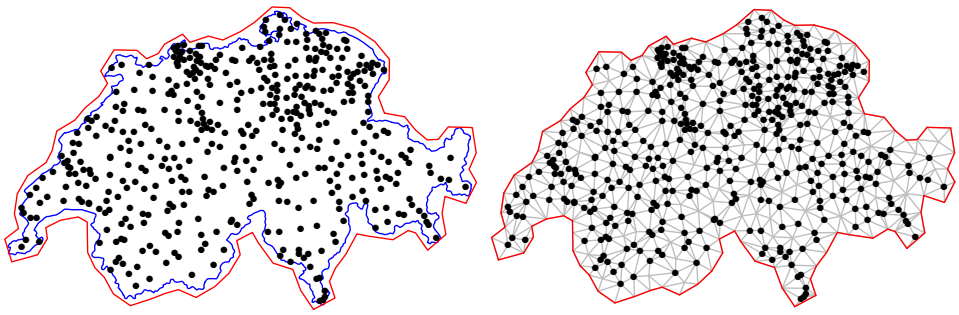

Figure 2.3: Left panel: Switzerland border (blue line) and simplified domain boundary (red line), with data locations represented by black dots. Right panel: triangulation of the simplified domain.

On the triangulation $\mathcal{T}$, we define the finite element space $V_{\mathcal{T}}^r(\Omega)$, with $r = 1, 2, \ldots,$ as the space of continuous surfaces over $\Omega_{\mathcal{T}}$ that are polynomials of degree r when restricted to any triangle in $\mathcal{T}$. This space is spanned by a set of finite elements basis functions $\{\psi_j(\mathbf{p}); j = 1, \ldots, N\}$, where each basis function is a locally supported piecewise polynomial function associated with a knot of the triangulation, and that takes value 1 at the associated knot and 0 on all other knots. For the simulation studies and for the application here presented, we use linear finite elements, as done in the Chapter 1. In this case, the knots of the triangulation coincide with the triangle vertices $\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_N$.

Let $\boldsymbol{\psi} = (\psi_1, \ldots, \psi_N)^T$ be the N-vector of the basis functions. Then we can represent a function h as an expansion on the basis system:

$$h(\mathbf{p}) = \mathbf{h}^T \boldsymbol{\psi}(\mathbf{p}), \tag{2.5}$$

where $\mathbf{h} = (h_1 \ldots h_N)$ is the vector of coefficients of the basis expansion. It turns out that $\mathbf{h}$ coincides with the evaluations of the function h at the N mesh nodes, i.e.,

$$\mathbf{h} = (h(\boldsymbol{\xi}_1), \ldots, h(\boldsymbol{\xi}_N))^T .$$

In fact, $h(\boldsymbol{\xi}_i) = \sum_{j=1}^N h_j \psi_j(\boldsymbol{\xi}_i) = \sum_{j=1}^N h_j \delta_{ij} = h_i$.

2.2.4 Discretization of the spatial field f via finite elements

Let Ψ be the $n \times N$-matrix of the evaluations of the N basis functions at the n locations $\{\mathbf{p}_i; i = 1, \ldots, n\}$:

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

Moreover, let R_0 and R_1 be the following $N \times N$-matrices:

$$R_0 = \int_{\Omega_\mathcal{T}} \boldsymbol{\psi}\boldsymbol{\psi}^T, \qquad R_1 = \int_{\Omega_\mathcal{T}} \nabla\boldsymbol{\psi}(K\nabla\boldsymbol{\psi})^T.$$

The following Proposition states that, once re-casted in the finite element space, finding the estimator $\hat{f}$, for a fixed K, reduces to solving a linear system. See the Appendix for details.

Proposition 2.2. *Given a symmetric and positive definite matrix K, the estimator $\hat{f} \in V^r_\mathcal{T}(\Omega)$ exists and is unique and is given by $\hat{f} = \hat{\mathbf{f}}\boldsymbol{\psi}$, where*

$$\begin{bmatrix} -(1-\rho)\frac{1}{n}\Psi^T\Psi & \rho\frac{1}{|\Omega|}R_1 \\ R_1 & R_0 \end{bmatrix} \begin{bmatrix} \hat{\mathbf{f}} \\ \mathbf{h} \end{bmatrix} = \begin{bmatrix} -(1-\rho)\frac{1}{n}\Psi^T\mathbf{z} \\ 0 \end{bmatrix}.$$

Proof. The existence and uniqueness of the solution are ensured by the invertibility of the matrices R_0 and $\left((1-\rho)\frac{1}{n}\Psi^T\Psi + \rho\frac{1}{|\Omega|}R_1R_0^{-1}R_1\right)$. □

From Proposition 2.2, we obtain the following expression for $\hat{\mathbf{f}}$, the vector of coefficients of the basis expansion for the spatial field estimate $\hat{f}$:

$$\hat{\mathbf{f}} = \left((1-\rho)\frac{1}{n}\Psi^T\Psi + \rho\frac{1}{|\Omega|}P\right)^{-1}(1-\rho)\frac{1}{n}\Psi^T\mathbf{z}, \tag{2.6}$$

where $P = R_1R_0^{-1}R_1$ represents the discretization of the penalty term in Equation (2.2). The discretization of the penalty term only involves first order derivatives thanks to the weak formulation (2.13) of the estimation problem, derived in the Appendix.

2.3 Implementation details

2.3.1 Parametrization of the anisotropy matrix K

Various choices are possible for the parametrization of the anisotropy matrix. In our work, we parametrize the matrix K with the parameter $\boldsymbol{\theta} = (\alpha, \gamma)$, where α represents the direction of the anisotropy and γ its intensity. The matrix $K(\boldsymbol{\theta})$ is then uniquely defined by setting its determinant equal to 1. Given α and γ, the matrix K is constructed exploiting its eigendecomposition, as follows:

$$K(\boldsymbol{\theta}) = Q(\alpha)\Sigma(\gamma)Q(\alpha)^{-1},$$

where

$$Q(\alpha) = \begin{bmatrix} \cos(\alpha) & -\sin(\alpha) \\ \sin(\alpha) & \cos(\alpha) \end{bmatrix} \qquad \text{and} \qquad \Sigma(\gamma) = \frac{1}{\sqrt{\gamma}}\begin{bmatrix} 1 & 0 \\ 0 & \gamma \end{bmatrix}.$$

The choice of this parametrization implies a periodicity of the functional H in (2.3) with respect to the parameter α; the numerical implementation of the estimation method appropriately takes this point into account.

2.3.2 Implementation of the algorithm and selection of the optimal smoothing parameter ρ

We here discuss the automatic choice of the smoothing parameter ρ. In that respect, it must be noticed that the optimal ρ for the estimation of the anisotropy matrix K differs from the optimal ρ for the estimation of the spatial field f conditional on K. In fact, when estimating the anisotropy matrix K, ρ should be large enough to imply a significant effect of the anisotropy in the estimated field; instead, when estimating the spatial field f for a given K, the smoothing parameter ρ should be selected to properly balance the effect of the penalization and the adherence to the data, and this is typically found in correspondence of rather small values of ρ. Indeed, in all the simulation studies and in the application we carried out, the optimal ρ for the estimation of K was at least one order of magnitude larger than the optimal ρ for the estimation of f given K. Specifically, we proceed as follows.

Having no knowledge of the right amount of regularization to impose for the optimal estimation of the anisotropy matrix K, we consider a grid of d possible values for the smoothing parameter $\{\rho_1, \dots, \rho_d\}$ spanning regularly the interval (0,1). For instance, the grid we use in the simulations and in the analysis of the Switzerland rainfall data is: (0.01,0.05,0.1,0.2,0.3,0.4,0.5,0.6,0.7,0.8,0.9). For each value of the smoothing parameter in this grid, the optimal anisotropy matrix is found, thus leading to the estimation of d anisotropy matrices $\{\hat{K}_1, ..., \hat{K}_d\}$. At this point, the choice of the optimal K among the d estimated ones is lead by the goal of our analysis: the optimal estimation of the spatial field. Indeed, we want to choose the anisotropy matrix which leads to the best estimated spatial field. Thus, we perform the estimation of f given the d possible anisotropy matrices $\{\hat{K}_1, \dots, \hat{K}_d\}$, where we now select in each case the optimal smoothing parameter ρ via generalized cross validation (GCV). We then select, among all of the d estimated anisotropy matrices and corresponding estimated spatial fields, the ones corresponding to the smallest GCV value.

The GCV index is available in closed-form for the model here considered, and given by

$$\mathrm{GCV}(\rho) = \frac{n}{(n - \mathrm{tr}(S))^2}(\mathbf{z} - \hat{\mathbf{z}})^T(\mathbf{z} - \hat{\mathbf{z}}), \tag{2.7}$$

where S is the smoothing matrix, which maps the vector of observed values $\mathbf{z}$ to the vector of fitted values $\hat{\mathbf{z}}$: $\hat{\mathbf{z}} = S\mathbf{z}$, i.e.

$$S = \Psi \left((1-\rho)\frac{1}{n}\Psi^T\Psi + \rho\frac{1}{|\Omega|}P \right)^{-1} (1-\rho)\frac{1}{n}\Psi^T.$$

The estimation of the anisotropy matrix K for a grid of possible smoothing parameters ρ also helps the numerical optimization of the functional H. Indeed, the functional H as a function of $\boldsymbol{\theta}$ shows different levels of regularity for different values of ρ: for lower values of ρ, the surface is smooth and has

a quadratic behavior, while, for higher values of ρ, it is almost flat in some regions and has a narrow ridge in correspondence of its minimum. As an example, Figure 2.4 shows the surfaces $H(\boldsymbol{\theta})$ for two extremes values of ρ (0.01 and 0.9) for the Switzerland rainfall data. The optimization algorithm can easily identify the right minimum for the first surface, but needs a proper starting point to reach the global minimum in the case of the second surface. In order to improve the performances of the optimization algorithm and to avoid local minima, for each value ρ_j for $j = 2, \ldots, d$, we provide, as starting point of the optimization of $H(\boldsymbol{\theta})$, the optimum obtained from the optimization for the value ρ_{j-1}. This technique allows the optimization method to start its search in the basin of attraction of the global minimum. In our implementation, we use the `optim` function of R (R Core Team, 2016) for the minimization of the functional H. In the numerical tests we performed, we never experienced any problem of convergence of the optimization algorithm.

The implementation of the model is based on the R package fdaPDE (Lila et al., 2016). The same package allows to create and refine the triangulations.

To make a comparison between the parametrization used in this Chapter (parameter ρ) and the one used in Chapter 1 (parameter λ), the grid of ρ values selected here corresponds to a quasi-linear (sinusoidal) grid on the parameter $\log(\lambda)$. For example, in the case of 200 data distributed on a domain of area 100, the grid of ρ values used here corresponds approximately to the following grid of $\log(\lambda)$ values: (-4, -2.25, -1.5, -0.7, -0.15, 0.3, 0.7, 1, 1.5, 2, 3).

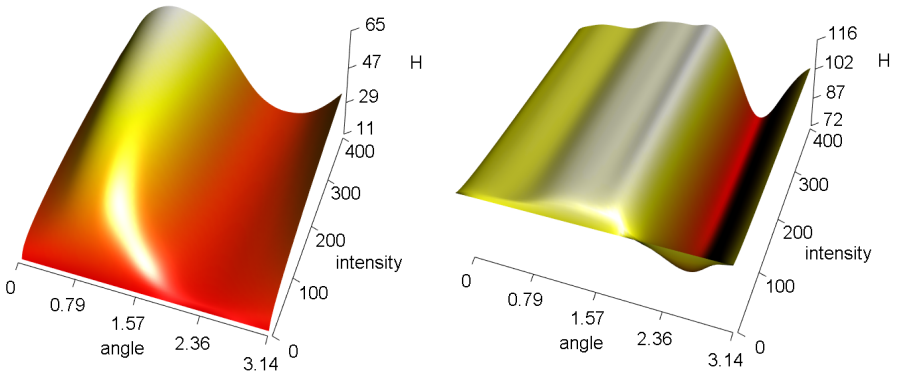

Figure 2.4: Surface $H(\boldsymbol{\theta})$ for two values of the smoothing parameter: ρ=0.01 (left panel) and ρ=0.9 (right panel).

2.4 Inclusion of space-varying covariates

The method described in the previous sections can be extended to handle space-varying covariates. Let $\mathbf{w}_i = (w_{i1}, \ldots, w_{iq})^T$ be a q-vector of covariates observed at $\mathbf{p}_i$. The model in Equation (2.1) can be modified to include an additive

term which takes into account the contribution of the covariates. Specifically, we consider the following semi-parametric generalized additive model:

$$z_i = \mathbf{w}_i^T \boldsymbol{\beta} + f(\mathbf{p}_i) + \epsilon_i \quad i = 1, \dots, n, \tag{2.8}$$

where $\boldsymbol{\beta} \in \mathbb{R}^q$ contains regression coefficients. The penalized sum-of-square-error functional becomes:

$$J_\rho(\boldsymbol{\beta}, f, K) = (1-\rho)\frac{1}{n}\sum_{i=1}^{n}\left(z_i - \mathbf{w}_i^T\boldsymbol{\beta} - f(\mathbf{p}_i)\right)^2 + \rho\frac{1}{|\Omega|}\int_\Omega \left(\nabla\cdot(K\nabla f)\right)^2. \tag{2.9}$$

The proposed two-step algorithm estimates separately the matrix K and the couple $\boldsymbol{\beta}$, f. The anisotropy matrix K is estimated using a parameter cascading approach. The estimation of $\boldsymbol{\beta}$ and f is conditional on K.

To estimate the anisotropy matrix K, we minimize the functional:

$$H(K) = \sum_{i=1}^{n}\left(z_i - \mathbf{w}_i^T\hat{\boldsymbol{\beta}}_K - \hat{f}_K(\mathbf{p}_i)\right)^2, \tag{2.10}$$

where $\hat{\boldsymbol{\beta}}_K$ and $\hat{f}_K$ are implicit functions of the structural parameter K, and are obtained minimizing the functional $J_\rho(\boldsymbol{\beta}, f, K)$ with respect to $\boldsymbol{\beta}$ and f for the given K.

As for the estimation of f and $\boldsymbol{\beta}$, given the estimated matrix K, we can prove, in a similar way to the one described in Section 2.2 and in the Appendix, that the estimation problem is well posed: the solution exists and is unique. Let W be a $n \times q$-matrix whose i-th row is the vector $\mathbf{w}_i$ of covariates at location $\mathbf{p}_i$. Let $P = W(W^TW)^{-1}W^T$ be the matrix that projects orthogonally on the subspace of $\mathbb{R}^n$ generated by the columns of W and let $Q = I - P$, where I is the identity matrix.

Proposition 2.3. *There exists a unique pair* $\left(\hat{\boldsymbol{\beta}} \in \mathbb{R}^q, \hat{f} \in V(\Omega)\right)$ *which minimize (2.9) for a fixed symmetric and positive definite matrix* K*. Moreover,*

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

$$(1-\rho)\frac{1}{n}\mathbf{h}_n^TQ\hat{\mathbf{f}}_n + \rho\frac{1}{|\Omega|}\int_\Omega(\nabla\cdot(K\nabla h))(\nabla\cdot(K\nabla\hat{f})) = (1-\rho)\frac{1}{n}\mathbf{h}_n^TQ\mathbf{z}$$

for every $h \in V(\Omega)$*.*

Proof. The estimator $\hat{\boldsymbol{\beta}}$ is obtained by differentiation the functional $J_\rho(\boldsymbol{\beta}, f, K)$ in (2.9) with respect to $\boldsymbol{\beta}$; then, by plugging $\hat{\boldsymbol{\beta}}$ in (2.9), the result is proven as in the proof of Proposition 2.1. □

Analogously to the case without covariates, after introducing the discretization with finite elements, we obtain the following result.

Proposition 2.4. *Given a fixed symmetric and positive definite matrix K, the estimators $\left(\hat{\boldsymbol{\beta}} \in \mathbb{R}^q, \hat{f} \in V_{\mathcal{T}}^r(\Omega)\right)$ that solve the discrete counterpart of the estimation problem exist and are unique. Moreover,*

- $\hat{\boldsymbol{\beta}} = (W^T W)^{-1} W^T (\mathbf{z} - \hat{\mathbf{f}}_n)$;
- $\hat{f} = \hat{\mathbf{f}}\boldsymbol{\psi}$, *where* $\hat{\mathbf{f}}$ *satisfies*

$$\begin{bmatrix} -(1-\rho)\frac{1}{n}\Psi^T Q \Psi & \rho\frac{1}{|\Omega|}R_1 \\ R_1 & R_0 \end{bmatrix} \begin{bmatrix} \hat{\mathbf{f}} \\ \mathbf{h} \end{bmatrix} = \begin{bmatrix} -(1-\rho)\frac{1}{n}\Psi^T Q \mathbf{z} \\ 0 \end{bmatrix}.$$

Proof. The existence and uniqueness of the solution are ensured by the invertibility of the matrices R_0 and $\left((1-\rho)\frac{1}{n}\Psi^T Q \Psi + \rho\frac{1}{|\Omega|}P\right)$. □

Solving the system for $\hat{\mathbf{f}}$ leads to the following expression for the vector of coefficients of the basis expansion for the spatial field estimate $\hat{f}$:

$$\hat{\mathbf{f}} = \left((1-\rho)\frac{1}{n}\Psi^T Q \Psi + \rho\frac{1}{|\Omega|}P\right)^{-1} (1-\rho)\frac{1}{n}\Psi^T Q \mathbf{z}. \tag{2.11}$$

For simplicity of exposition, we here organized the parameter cascading step, used for the estimation of the anisotropy matrix K, defining $\boldsymbol{\beta}$ as a nuisance parameter. On the other hand, one could as equivalently define $\boldsymbol{\beta}$ as a structural parameter. In fact, the minimizer of the sum of square error functional in (2.10), with respect to $\boldsymbol{\beta}$, leads to the same form of the estimator $\hat{\boldsymbol{\beta}}$ that is given in Propositions 2.3 and 2.4.

2.5 Simulation studies

In this section, we present the results of three simulation studies that show the performances of the proposed method (anisotropic SR-PDE) and compare it to isotropic smoothing with penalization of the L^2-norm of the Laplacian as described in Sangalli et al. (2013) (isotropic SR-PDE), and to isotropic and anisotropic kriging. The isotropic kriging estimates are obtained with the function `krige` of the R package gstat (Pebesma, 2004) estimating the parameters of the variogram with the function `fit.variogram` of the same package. The anisotropic kriging estimates are obtained with the function `krige.conv` of the R package geoR (Ribeiro Jr and Diggle, 2016) estimating the parameters of the variogram with the function `likfit` of the same package. We use four classical variogram models: Matérn, Gaussian, Spherical and Exponential. For some simulation replicates, we observed numerical instability with Matérn and Gaussian kriging: the estimation of the variogram parameters performed with the function `likfit` failed, not returning any valid result; in these cases, we run again the function trying different initial values for the parameters, until finite estimates

were obtained. The few replicates when the instability could not be avoided were excluded from the analysis. For all the simulation studies, we also performed isotropic and anisotropic filtered kriging with the function `Krig` of the R package fields (Nychka et al., 2015) for the variogram models implemented in the package (Matérn and Exponential), but we did not find any significant difference with respect to standard kriging (as confirmed by pairwise Wilcoxon test on the RMSE). Hence, to save space, the results for filtered kriging are not included.

2.5.1 First simulation study: Matérn field with varying anisotropy

In the first simulation study, we generate 200 Matérn fields characterized by varying anisotropy, using the function `RFsimulate` of the R package RandomFields (Schlather et al., 2016). The anisotropy matrices are randomly generated by sampling the angle and the first element of the diagonal from two uniform distributions with ranges $[0, \pi]$ and $[0.2, 0.5]$ respectively; the second element of the diagonal is fixed to 1. The smoothness parameter ν is fixed to 2. Figure 2.5 shows three fields generated for this first simulation study.

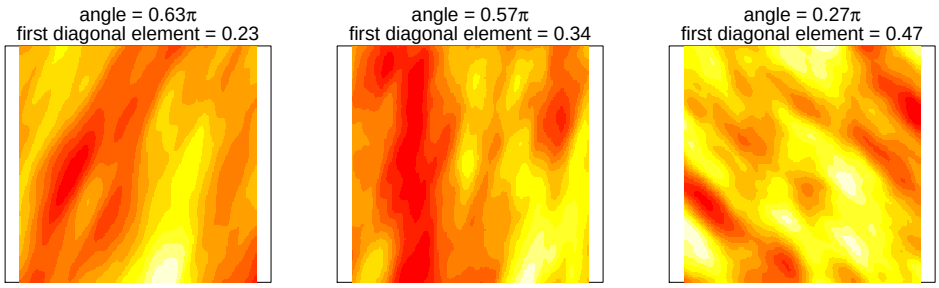

Figure 2.5: First simulation study. Three generated Matérn fields with different anisotropy angles and intensities (determined by the values of the first diagonal element of the anisotropy matrix, the second being fixed to 1).

Within a squared domain of side length 10, we uniformly sample 200 points, imposing a minimum distance of 0.1 among them. Each generated random field is then sampled at these 200 locations, adding a gaussian noise with mean 0 and standard deviation 0.2 (which corresponds approximately to 5% of the range of the data).

The top right panel of Figure 2.6 shows the data obtained in the first simulation replicate, corresponding to the true field represented in the top left panel of the same figure. The central and bottom rows of the same figure compare the field estimates obtained by isotropic and anisotropic SR-PDE and isotropic and anisotropic Matérn kriging. Superimposed to the true field and to the corresponding estimates are ellipses representing respectively the true and estimated anisotropy matrices: in the case of anisotropic SR-PDE, to use the same

parametrization of the anisotropy matrix considered in the data generation and in the kriging estimation, we plot here the squared root of K, normalized in order to have first eigenvalue equal to 1. To help the interpretation of this figure, we also superimpose circles to isotropic estimates. Anisotropic estimates clearly better capture the behavior of the true field with respect to the corresponding isotropic estimates.

To evaluate the goodness of the fit, we compute the RMSE over the 200 simulation replicates of the estimated field with respect to the true one, on a fine regular square grid spanning the whole domain with spacing 0.05. Figure 2.7 shows the boxplots of the RMSE of the estimates obtained with the different methods. The gray boxplot corresponds to the proposed method. Anisotropic SR-PDE has better performances with respect to all isotropic techniques. Indeed, pairwise Wilcoxon tests confirm that the RMSE of the estimates obtained with anisotropic SR-PDE are significantly smaller than those associated with the isotropic methods, with p-values of the order of 10^{-16}. Moreover, anisotropic SR-PDE has also significantly smaller RMSE than anisotropic kriging, with p-values of pairwise Wilcoxon tests of the order of 10^{-3} for the comparison with the Matérn and Spherical kriging, and of the order of 10^{-5} for the comparison with Exponential and Gaussian kriging. In particular, it should be noticed that the RMSE of the estimates obtained with the proposed method are significantly smaller than the RMSE of the estimates obtained by anisotropic Matérn kriging, which should, in this case, be the best possible model, since it assumes the same covariance structure used to generate the data.

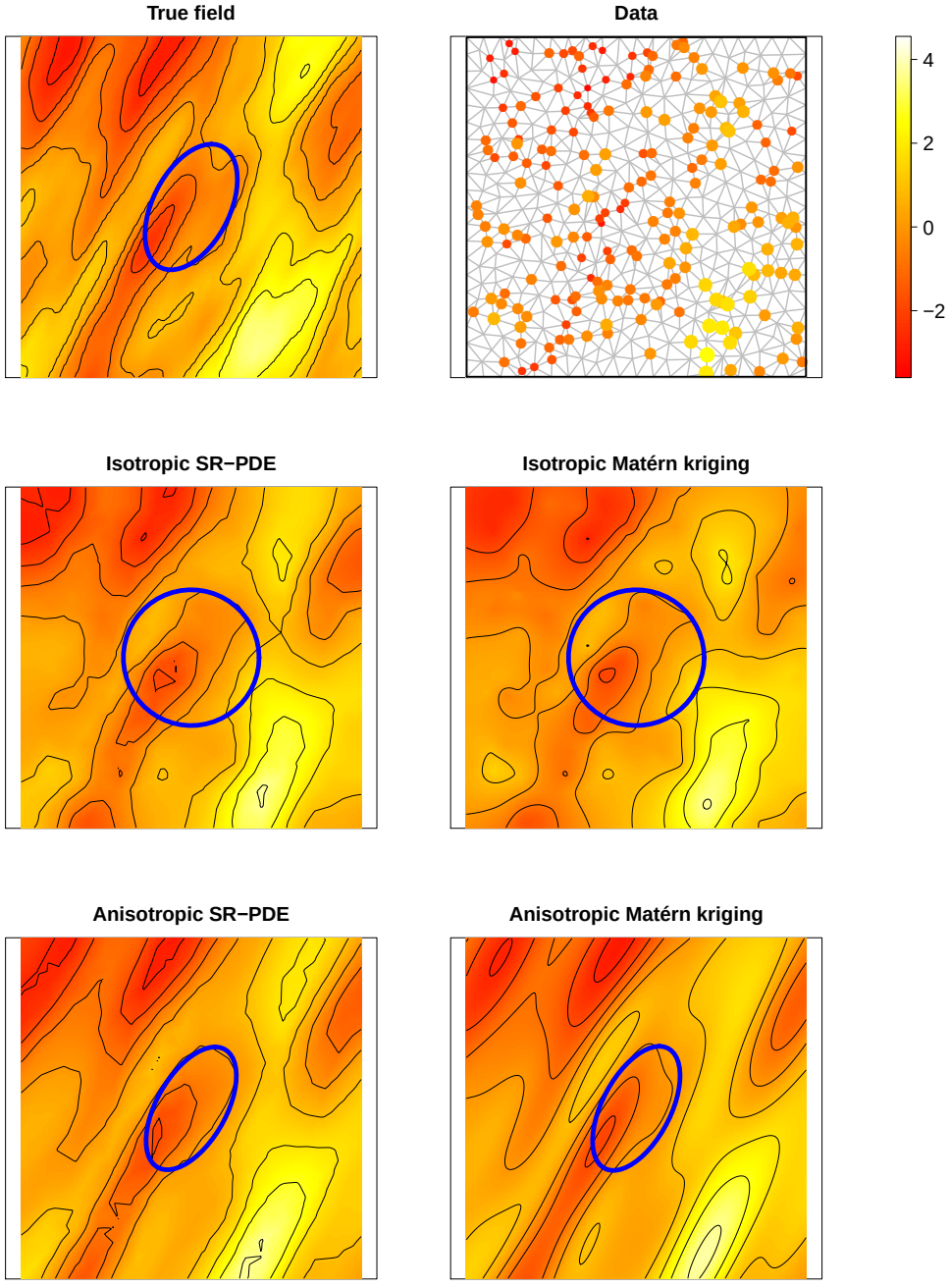

Figure 2.6: First simulation study. First row: true field (left) and sampled data (right), with the triangulation used for SR-PDE estimates in gray. Second row: estimates provided by isotropic SR-PDE and isotropic Matérn kriging. Third row: estimates provided by anisotropic SR-PDE and anisotropic Matérn kriging.

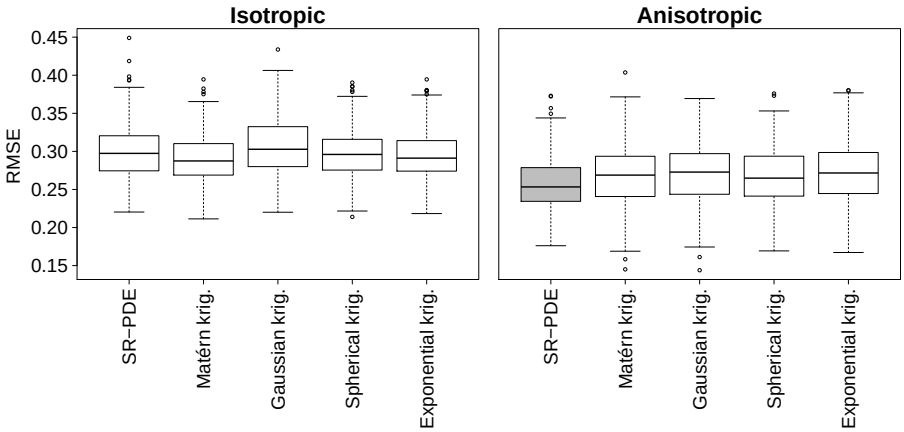

Figure 2.7: First simulation study. Boxplots of the RMSE over the 200 simulation replicates of the estimates obtained with isotropic and anisotropic SR-PDE and isotropic and anisotropic kriging with different variogram models (Matérn, Gaussian, Spherical and Exponential).

The anisotropy is correctly identified by the proposed method. Indeed, the right angle of anisotropy is estimated, as shown in Figure 2.8, which reports a scatterplot of the estimated angle against the true angle.

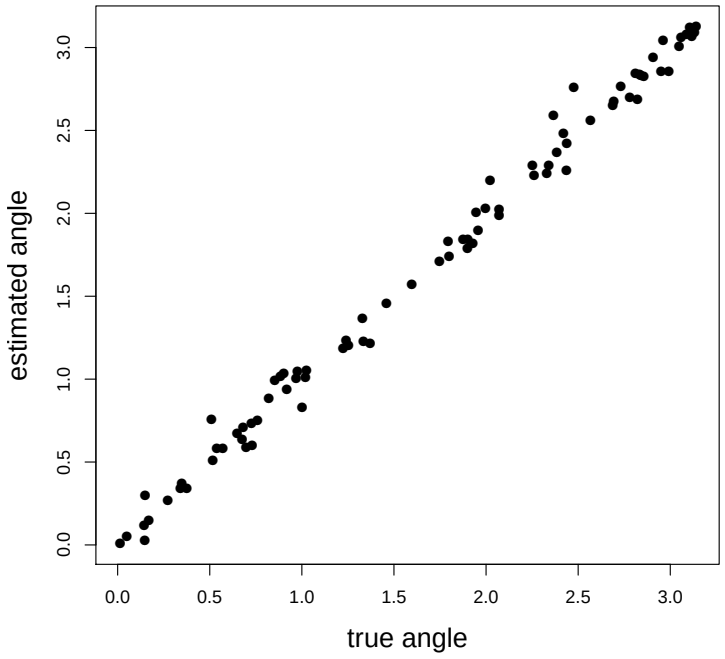

Figure 2.8: Second simulation study. Scatter plot of the angles of anisotropy estimated by the proposed method versus the true angles.

2.5.2 Second simulation study: Matérn field with varying anisotropy and inclusion of covariates

In this simulation study, we consider the same simulation setting described in Section 2.5.1, but we add a space-varying covariate. Specifically, for 200 simulation replicates, we generate from the semiparametric model (2.8), with f equal to one of the 200 Matérn fields generated for the first simulation study, $\beta = 1$ and $w_i = 2\sin((x_i/5 - 1)1.5\pi)\cos((y_i/5 - 1)1.5\pi)$.

Figure 2.9 shows, in the first row, the Matérn field in the first simulation replicate and the space-varying covariate; the total field, obtained adding the Matérn field and the covariate, is represented in the first panel of the second row. The following rows of the same figure compare the corresponding estimates provided by isotropic and anisotropic SR-PDE and isotropic and anisotropic universal Matérn kriging. Also in this case, taking into account the anisotropy significantly improves the fitting.

Figure 2.10 shows the boxplots of the RMSE of the estimates obtained with the different methods over the 200 replicates. The RMSE of the fields estimated with anisotropic SR-PDE are significantly lower than the ones obtained with the isotropic techniques, with p-values of pairwise Wilcoxon tests of the order of 10^{-14} for the comparison with isotropic SR-PDE, and of the order of 10^{-16} for the comparison with isotropic universal kriging. Furthermore, anisotropic SR-PDE has also better performances with respect to anisotropic universal kriging with the different variogram models, with p-values of pairwise Wilcoxon tests of the order of 10^{-16}. It should here be remarked that surprisingly anisotropic SR-PDE does significantly better even than universal Matérn kriging, which exactly assumes the same covariance structure used to generate the data.

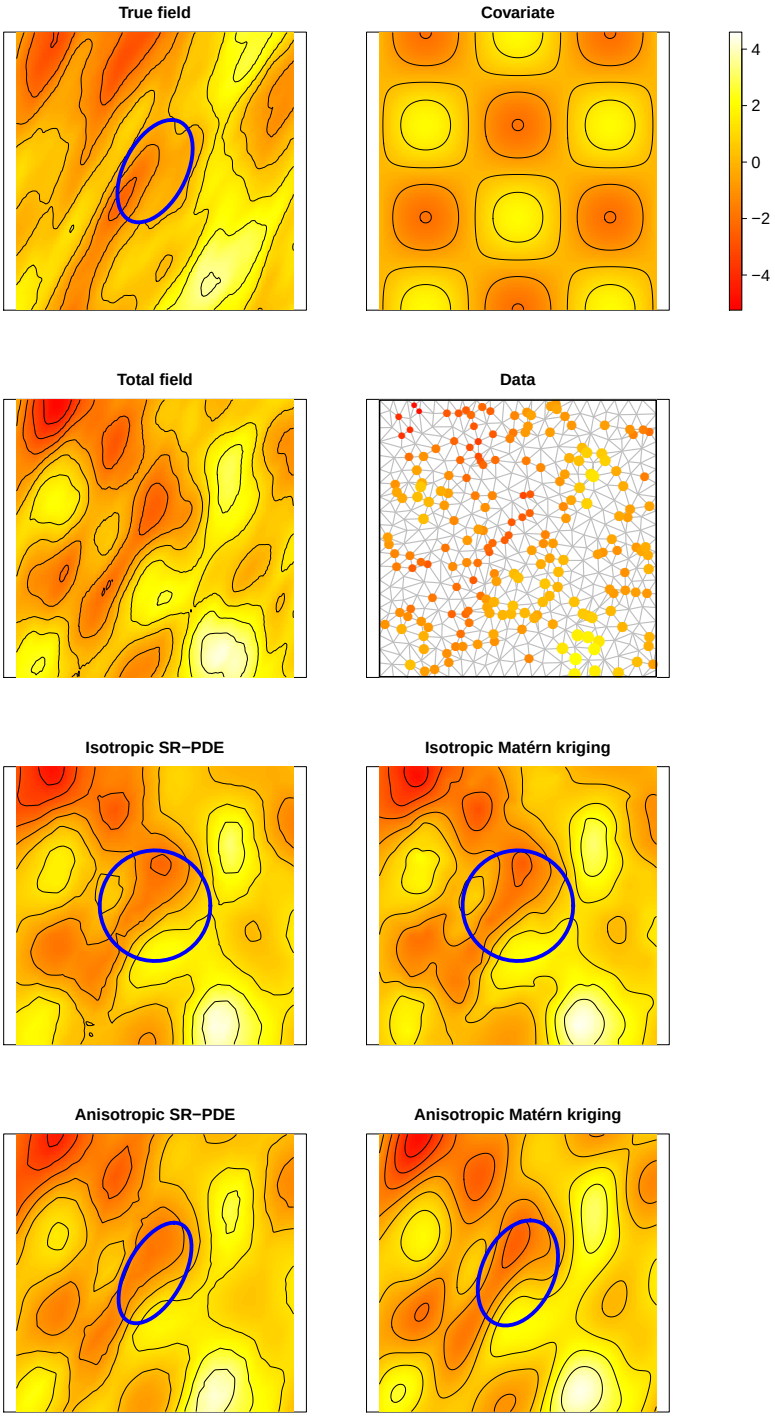

Figure 2.9: Second simulation study. First row: true Matérn field and space-varying covariate. Second row: total field and sampled data, with the triangulation used for SR-PDE estimates represented in gray. Third row: estimates provided by isotropic SR-PDE and isotropic universal Matérn kriging. Fourth row: field estimates provided by anisotropic SR-PDE and by anisotropic universal Matérn kriging.

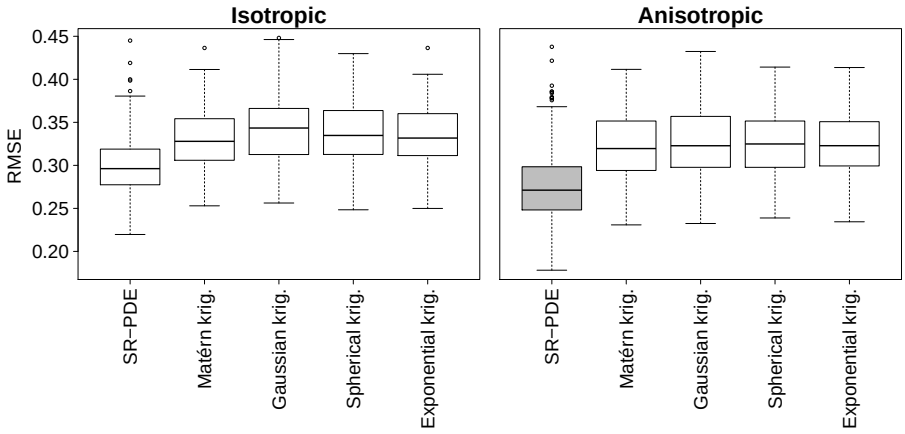

Figure 2.10: Second simulation study. Boxplots of the RMSE of the estimates obtained with isotropic and anisotropic SR-PDE and isotropic and anisotropic universal kriging with different variogram models (Matérn, Gaussian, Spherical and Exponential).

2.5.3 Third simulation study: field over irregular domain

In this simulation study, we want to test the performances of the considered methods in estimating the spatial field represented in the top left panel of Figure 2.11 from its noisy measurements. The field is defined on a rectangular domain with an elongated rectangular hole within it. The values of the field on the two sides of the hole are different; therefore the complex geometry of the domain should be taken into account to properly estimate this field.

Within the domain of side lengths 8 and 4, we uniformly sample 300 locations, imposing a minimum distance of 0.2 among them. We then sample the field at these 300 locations, adding a gaussian noise with mean 0 and standard deviation 0.1 (which corresponds to 5% of the range of the data). We repeat the simulation for 200 independent realizations of the additive gaussian noise.

The top right panel of Figure 2.11 shows the data obtained in the first simulation replicate. The central and bottom rows of the same figure show the field estimates provided by isotropic and anisotropic SR-PDE and isotropic and anisotropic Matérn kriging. Differently from kriging, SR-PDE is able to properly take into account the shape of the domain, while kriging smooths across the internal boundaries, closely connecting data points that are instead separated by the hole in the domain. Anisotropic SR-PDE improves the performances of isotropic SR-PDE, providing estimated fields with more regularity along the direction of anisotropy.

We then compare the various methods on the base of the RMSE over the 200 simulation replicates, with the RMSE computed on a fine regular square grid of step 0.05 spanning the whole domain. As apparent from the boxplots of the RMSE displayed in Figure 2.12, and confirmed by Wilcoxon pairwise tests,

the RMSE of the estimates obtained with anisotropic SR-PDE are significantly smaller than those associated with isotropic SR-PDE and to both isotropic and anisotropic kriging with any variogram model.

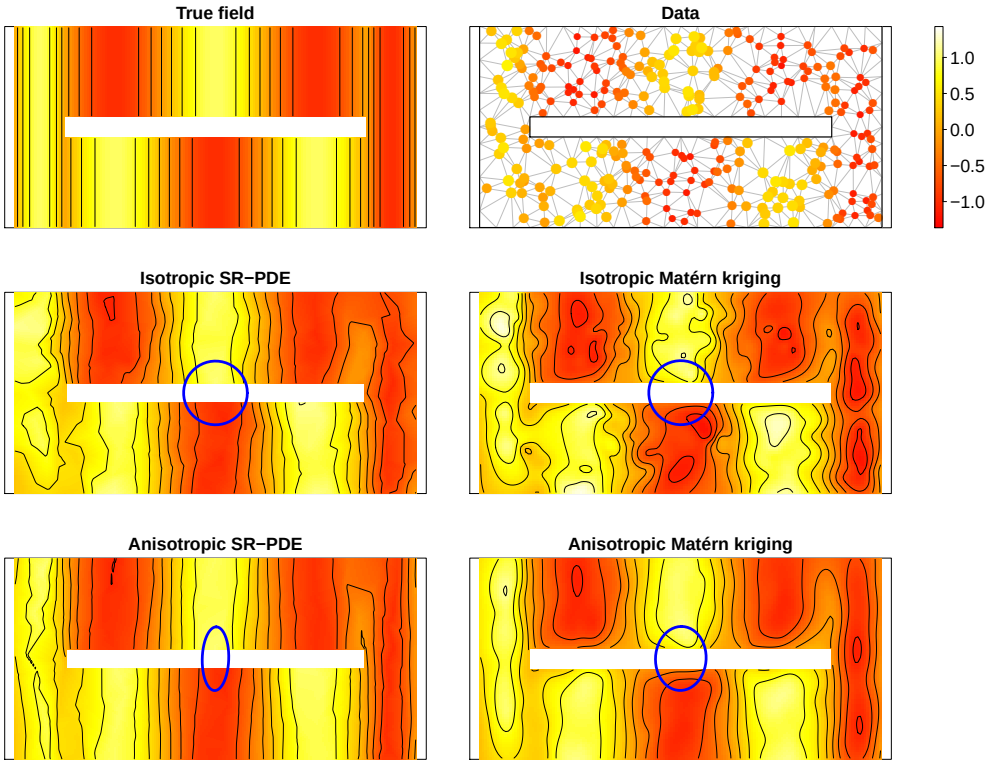

Figure 2.11: Third simulation study. First row: true field and sampled data, with the triangulation used for SR-PDE estimates in gray. Second row: field estimates provided by isotropic SR-PDE and isotropic Matérn kriging. Third row: field estimates provided by anisotropic SR-PDE and anisotropic Matérn kriging.

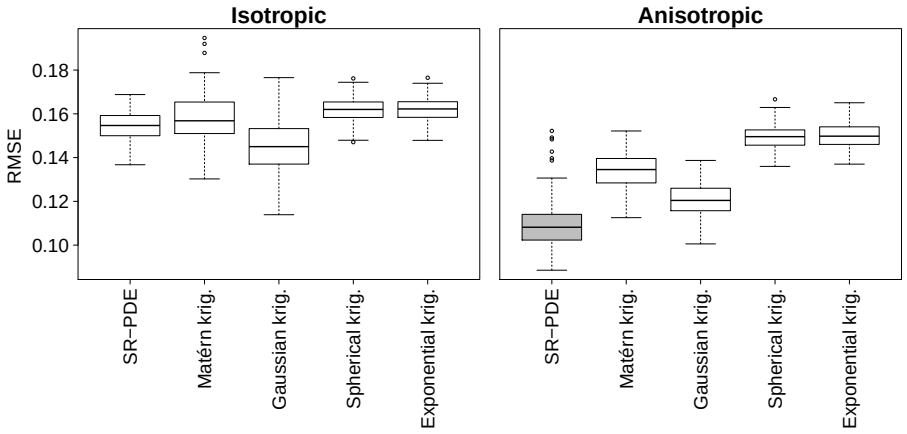

Figure 2.12: Third simulation study. Boxplots of the RMSE over the 200 simulation replicates of the estimates obtained with isotropic and anisotropic SR-PDE and isotropic and anisotropic kriging with different variogram models (Matérn, Gaussian, Spherical and Exponential).

2.6 Application to the analysis of the Switzerland rainfall data

We apply the proposed method to the analysis of the dataset of 467 daily rainfall measurements made in Switzerland, which was used for the Spatial Interpolation Comparison 97 (Dubois et al., 2003). The data are shown in Figure 2.1.

The data include the elevation at the 467 locations, that we use here as a covariate since intuition suggests that the orography of the region may play an important role in the rainfall phenomenon; see Figure 2.13 that shows the elevation over Switzerland. However, when implementing the anisotropic SR-PDE model using the elevation as a covariates, it turns out that elevation is not significant in the model. This is probably due to the fact that the effect of elevation on rainfall is not linear; the apparent anisotropy in the distribution of rainfall is the result of the interaction between the geomorphology and atmospheric circulation. Unfortunately, data about wind streams and atmospheric circulation are not included in this dataset. We thus discard the elevation from the model and compute the SR-PDE estimate without this covariate.

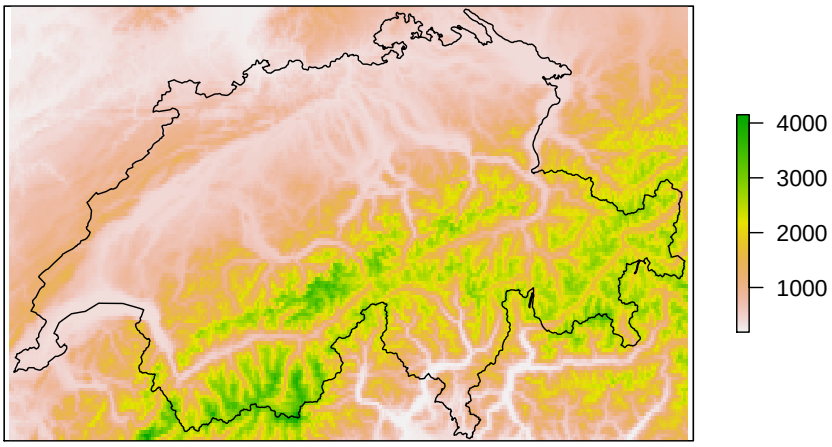

Figure 2.13: Elevation in the Switzerland region. This plot is obtained with the R package geostatsp (Brown, 2015).

Figure 2.14 displays the estimated anisotropy matrix, superimposed to the data, and Figure 2.15 shows the corresponding field estimate. These figures highlight that the proposed method correctly identified the anisotropy and that it provides a smooth field that is able to capture the important features of the data, with the elongated regions of homogeneous values well defined.

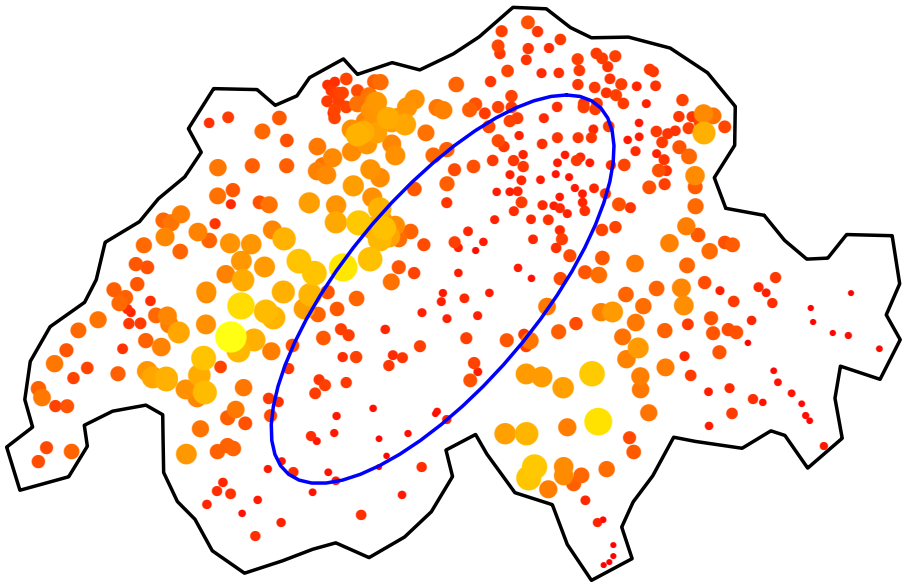

Figure 2.14: Anisotropy matrix (represented by the blue ellipse) estimated for the Switzerland rainfall data (represented by the colored dots).

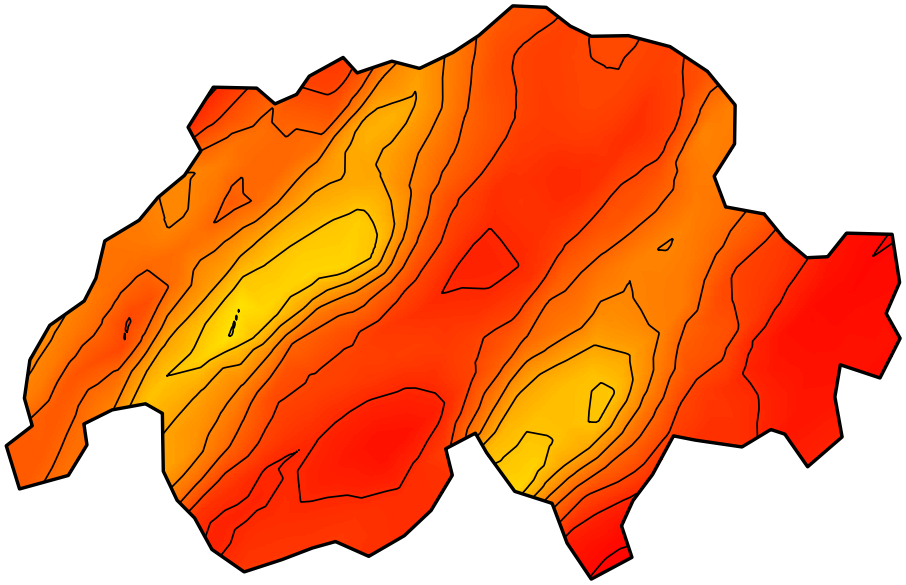

Figure 2.15: Field estimate for the Switzerland rainfall data obtained with the proposed method.

2.7 Possible model extensions and directions of future research

The SR-PDE method can be extended to handle spatio-temporal data, generalizing the model proposed by Bernardi et al. (2016), and to deal with areal data instead of geostatistical data, following a similar approach as in Azzimonti et al. (2015). Another possible extension is to data observed over curved domains, combining the method here proposed with the one described in Ettinger et al. (2016). This extension would be particularly interesting in the field of geosciences and environmental sciences, since data are often observed over regions presenting a complex orography; for instance, in the application to Switzerland rainfall data, we could include elevation in the definition of the domain of interest. Furthermore, another very interesting generalization of the proposed approach would consist in integrating it with the model framework introduced by Azzimonti et al. (2015), where the regularizing term involves a more complex PDE, $Lf = u$, suggested by prior knowledge and the phenomenon behavior, and that includes anisotropic and non-stationary diffusion, transport and reaction terms:

$$Lf = -\nabla \cdot (K(x, y, \theta)\nabla f) + \mathbf{b}(x, y, \eta) \cdot \nabla f + c(x, y, \zeta)f. \tag{2.12}$$

In Azzimonti et al. (2015) the parameters θ, η and ζ in the diffusion, transport and reaction terms in (2.12) are fixed on the base of their physical meaning,

thanks to the prior knowledge of the problem under study. On the other hand, oftentimes the prior knowledge of the phenomenon is not as detailed as to indicate the specific values for these parameters. In such contexts, these parameters could be estimated from data, generalizing the methodology presented in this chapter. This would enable us to combine prior knowledge and observed data in the definition of the anisotropic and non-stationary spatial variation of the phenomenon. In particular, the use of prior knowledge would avoid un-identifiability issues, which would otherwise be inevitably involved in a such a flexible modelling of anisotropic and non-stationary spatial variation. See also Fuglstad, Simpson, Lindgren and Rue (2015), that extend the approach based on Gaussian random fields and stochastic PDEs introduced in Lindgren et al. (2011). Other works which developed models to handle non-stationary anisotropy in spatial data within the framework of Gaussian random fields are Fuglstad, Lindgren, Simpson and Rue (2015), Fuglstad et al. (2013), Fuglstad et al. (2014). In the application to Switzerland rainfall data, if information about wind stream or air circulation were available, it could be used to define a PDE with anisotropic and non-stationary diffusion and transport terms; in particular, the transport terms can be used to induce unidirectional smoothing effects in the direction of the wind streams.

Appendix

2.A Estimation of f given the anisotropy matrix K and discretization via finite elements

As shown in Azzimonti et al. (2015), by introducing an auxiliary variable g, we can write the fourth-order problem (2.4) as a coupled system of second order problems:

$$\begin{cases} \nabla \cdot (K \nabla \hat{f}) = g & \text{in } \Omega \\ \hat{f} = 0 & \text{on } \partial\Omega \end{cases}$$

$$\begin{cases} \rho \frac{1}{|\Omega|} \nabla \cdot (K \nabla g) = -(1-\rho)\frac{1}{n}\sum_{i=1}^{n}(\hat{f} - z_i)\delta_{\mathbf{p}_i} & \text{in } \Omega \\ g = 0 & \text{on } \partial\Omega. \end{cases}$$

By integrating the differential equations against test functions h, v, and exploiting Green's theorem (or integration by parts), we then obtain the following weak formulation of the problem: find f, g such that

$$\begin{aligned} (1-\rho)\frac{1}{n}\mathbf{h}_n^T \hat{\mathbf{f}}_n - \rho \frac{1}{|\Omega|}\int_\Omega (\nabla g \cdot K \nabla h) &= (1-\rho)\frac{1}{n}\mathbf{h}_n^T \mathbf{z} \\ \int_\Omega g v + \int_\Omega (\nabla v \cdot K \nabla \hat{f}) &= 0 \end{aligned} \tag{2.13}$$

for any h, v. This reformulation of the estimation problem involves only first order derivatives and it is well suited to be solved numerically by discretization via the finite element space $V_{\mathcal{T}}^r(\Omega)$ described in Section 2.2.3 of the main text. This allow us to reduce the infinite dimensional problem to a finite dimensional one, and to reduce the estimation to the solution of a linear system. Indeed, taking the functions f, g, h, v in the finite element space $V_{\mathcal{T}}^r(\Omega)$, we obtain the following expressions for the integrals in (2.13):

$$\int_{\Omega_\mathcal{T}} (\nabla g \cdot K \nabla h) = \mathbf{g}^T R_1 \mathbf{h}, \qquad \int_{\Omega_\mathcal{T}} g v = \mathbf{g}^T R_0 \mathbf{v}, \qquad \int_{\Omega_\mathcal{T}} (\nabla v \cdot K \nabla \hat{f}) = \mathbf{v}^T R_1 \hat{\mathbf{f}}.$$

This leads to Proposition 2.2.

2.B Properties of the estimator

For a fixed value of the anisotropy matrix K, the properties of the estimator can be derived, as in Section 1.3.3 of Chapter 1.

The estimator $\hat{\mathbf{f}}$ in (2.6) is linear in the observed data values $\mathbf{z}$, and has a typical penalized least-square form. The estimate of the field f at a generic location $\mathbf{p} \in \Omega$ is given by

$$\hat{f}(\mathbf{p}) = \boldsymbol{\psi}(\mathbf{p})\hat{\mathbf{f}} = \boldsymbol{\psi}(\mathbf{p}) \left((1-\rho)\frac{1}{n}\Psi^T\Psi + \rho\frac{1}{|\Omega|}P \right)^{-1} (1-\rho)\frac{1}{n}\Psi^T\mathbf{z}$$

and its mean and variance conditional on K are given by

$$\begin{aligned} E[\hat{f}(\mathbf{p})|K] &= \boldsymbol{\psi}(\mathbf{p}) \left((1-\rho)\frac{1}{n}\Psi^T\Psi + \rho\frac{1}{|\Omega|}P \right)^{-1} (1-\rho)\frac{1}{n}\Psi^T\mathbf{f}_n \\ Var[\hat{f}(\mathbf{p})|K] &= \sigma^2\boldsymbol{\psi}(\mathbf{p}) \left((1-\rho)\frac{1}{n}\Psi^T\Psi + \rho\frac{1}{|\Omega|}P \right)^{-1} (1-\rho)\frac{1}{n}\Psi^T \\ &\quad \Psi(1-\rho)\frac{1}{n} \left((1-\rho)\frac{1}{n}\Psi^T\Psi + \rho\frac{1}{|\Omega|}P \right)^{-1} \boldsymbol{\psi}(\mathbf{p})^T. \end{aligned}$$

The covariance between the estimates of the field f at two generic locations $\mathbf{p}_1$ and $\mathbf{p}_2$ conditional on K is given by

$$\begin{aligned} &Cov[\hat{f}(\mathbf{p}_1), \hat{f}(\mathbf{p}_2)|K] = \\ &\qquad \sigma^2\boldsymbol{\psi}(\mathbf{p}_1) \left((1-\rho)\frac{1}{n}\Psi^T\Psi + \rho\frac{1}{|\Omega|}P \right)^{-1} (1-\rho)\frac{1}{n}\Psi^T \\ &\qquad \Psi(1-\rho)\frac{1}{n} \left((1-\rho)\frac{1}{n}\Psi^T\Psi + \rho\frac{1}{|\Omega|}P \right)^{-1} \boldsymbol{\psi}(\mathbf{p}_2)^T. \end{aligned}$$

.

As can be observed from the expressions of the means of the field estimators presented here and in Chapter 1, the bias of the estimators is not zero. As described in Ettinger et al. (2016) and Azzimonti et al. (2014), this bias is due to two sources: the inclusion of the regularizing term in the estimation and the discretization. The first source is common to all penalized regression techniques and disappears as n increases, if the smoothing parameter decreases with n. The second source is due to the choice of employing a basis expansion and disappears as the mesh is refined. See Lemma 1 and Theorem 3 in Azzimonti et al. (2014) that provides some preliminary results on asymptotic unbiasedness of the estimators. The study of the consistency of these estimators, which includes convergence to zero of the variance, is still an open problem. This study involves also the search for a rate of decrease of the smoothing parameter with respect to n that makes the variance vanish.

Chapter 3

Analysis of data with complex non-stationary spatial anisotropy influenced by the texture of the domain

3.1 Introduction

Data with spatial dependence arise from many fields of sciences such as meteorology, biology and geosciences. When analyzing spatial data, it is of paramount importance to understand the physical phenomenon generating the data and to adapt the estimation technique to properly take into account the characteristics of the specific phenomenon under study.

A smoothing technique able to do that is spatial regression with penalized regularization. This technique has been described in the previous chapters of this thesis and its flexibility has already been shown. The various modelling features of this technique include the ability to properly deal with complex domains featuring concavities or interior holes, the possibility of imposing known boundary conditions and the capacity of including in the model covariate information via a semi-parametric framework.

Another important feature of this technique is the flexibility given by the possibility of adapting the penalization term to the specific problem at hand. Indeed, the penalization term involves a general Partial Differential Equation (PDE) which can model various phenomena. In the simplest case, where no knowledge is available on the phenomenon under study, the PDE in the penalization term is the simple Laplacian operator (see Ramsay (2002) and Sangalli et al. (2013)). When prior knowledge can be expressed by a differential model, it can be included in the penalization in order to characterize the spatial dependence of the solution to the estimation problem (see Azzimonti et al. (2015)). If

no prior knowledge on the phenomenon is available, but the data show a spatial dependence which is not isotropic, an anisotropic diffusion operator, whose parameters are estimated from the data, can be included in the penalization, as described in Chapter 2.

In this chapter we explore the possibility of characterizing the PDE involved in the penalization considering the texture of the domain on which the data are defined. In particular, we consider data characterized by non-stationary spatial anisotropy induced by the structure of the domain.

As an illustrative example, we consider mobile phone data in the metropolitan area of Milan. We analyze the Telecom Italia database, which collects the measurements of a quantity called Erlang, consisting of the average number of mobile phones using the network for calling over time intervals of 15 minutes and square sites with dimension 232m×309m. The data are collected from Wednesday, March 18th 2009, 00:15 to Tuesday, March 31st 2009, 23:45 on a uniform lattice of 97×109 sites. These data are used to investigate population dynamics, since the Erlang measurements can be considered as an approximation of the number of active people present in that site at that time. See Manfredini et al. (2015), Secchi et al. (2015), Zanini et al. (2016) and Passamonti (2016) for previous works on these data.

In this chapter, we show some first analysis considering the Erlang data for a fixed time interval (Monday, March 23rd 2009, from 19:00 to 19:15), thus treating them as spatial data without for the moment considering the temporal evolution. Moreover, we assign the value measured for each site to the center of the 232m×309m domain on which it is measured, thus setting the analysis in the framework of geostatistical data instead of in the framework of areal data. Section 3.4 describes, among the possible extensions, how to analyze these data taking into account the temporal dependence and considering them as areal data.

The top panel of Figure 3.1 shows the distribution of the data on the spatial domain for the time instant we fixed. The data are represented over the map of the metropolitan area of Milan, which is also reported in the bottom panel.

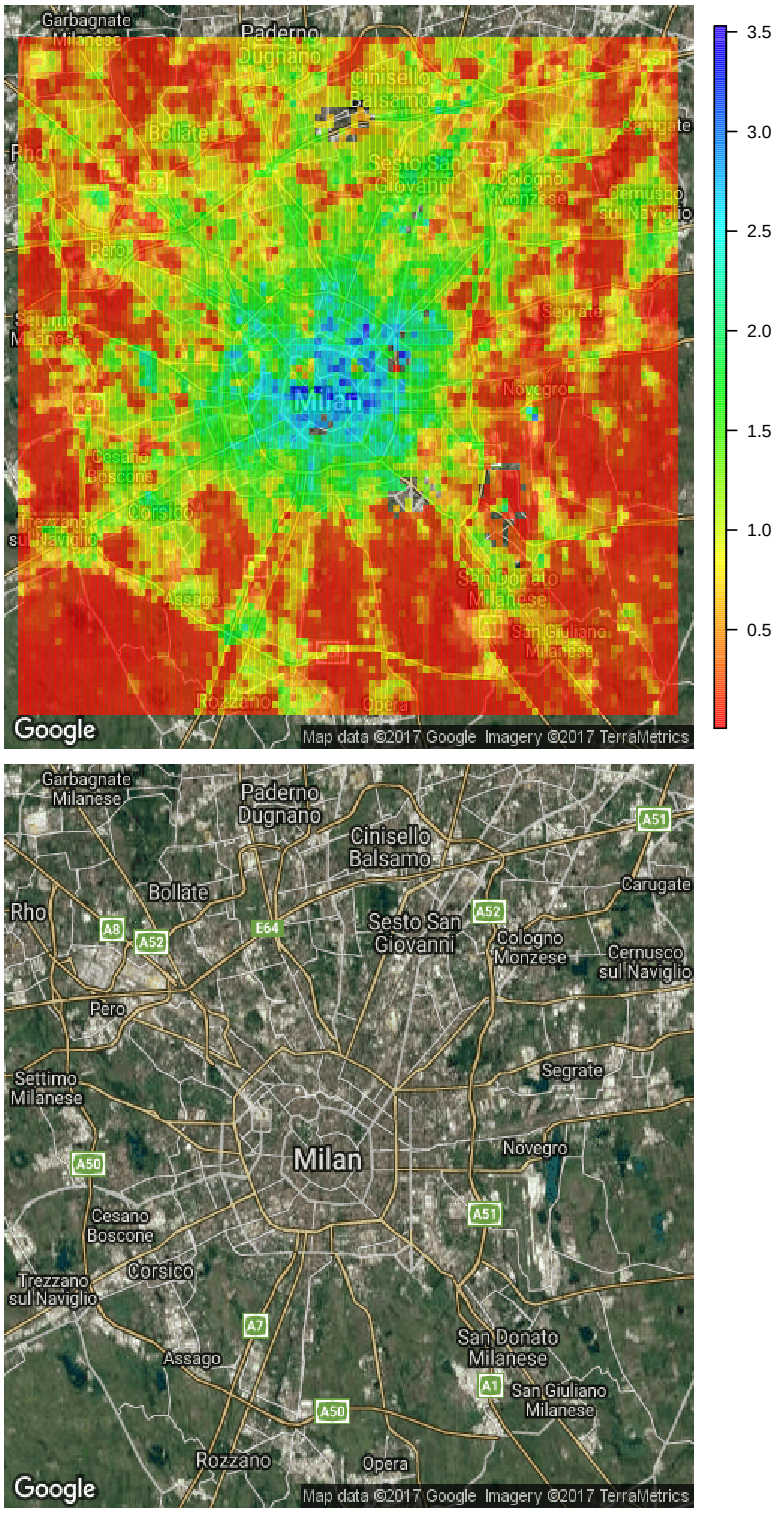

Figure 3.1: Erlang data at a fixed time instant (Monday, March 23rd 2009, 19:00) over the map of the metropolitan area of Milan, which is shown in the bottom panel.

As can be observed in the picture, the data are characterized by high values in the central part of the domain, the municipality of Milan, and lower values in the surrounding regions, the hinterland and countryside. An interesting characteristic of these data is the spatial structure the data show in the hinterland regions. In particular, large regions of low values (red areas in the plot) are separated by narrow elongated regions of higher values (represented in yellow and green in the plot). This non-homogeneity in the spatial dependence of the data is explained by the road network characterizing the domain, as shown in the right panel of the figure: the regions of the hinterland characterized by higher values follow the directions of the highways and the main roads.

This dataset is an example of a phenomenon strongly influenced by the texture of the domain. Indeed, in this case, the domain is not homogeneous, but presents regions with different characteristics which determine a non-homogeneous spatial dependence of the data. In particular, the road network induces strong local anisotropy along the main roads.

We here propose to analyze the Erlang data using a spatial regression with penalized regularization, where the PDE involved in the penalization accounts for the topology of the road network characterizing the metropolitan area of Milan. The estimation of the parameters of the PDE will be driven by the knowledge of the road network of the metropolitan area of Milan. In particular, we consider a non-stationary anisotropic diffusion operator whose anisotropy is estimated from the geometry of the road network characterizing the domain. The estimation of the anisotropy is performed with the technique presented in Della Rossa et al. (2010), where an algorithm is developed to model traffic flows in complex networks such as large urbanized areas.

The rest of the chapter is organized as follows. Section 3.2 describes the model used to analyze the data. Section 3.3 show the first results obtained for the analysis of the Erlang data. Section 3.4 gives an overview of possible extensions of the analysis and future directions of work.

3.2 Model

We consider a bounded spatial domain $\Omega \in \mathbb{R}^2$, whose boundary $\partial\Omega$ is a curve of class $\mathcal{C}^2$. Let $z_i \in \mathbb{R}$ be the value of a variable of interest observed at n point $\mathbf{p}_i = (x_i, y_i); i = 1, \ldots, n$ within Ω. We assume that $\{z_i; i = 1, \ldots, n\}$ are noisy observations of an underlying smooth function $f : \Omega \to \mathbb{R}$:

$$z_i = f(\mathbf{p}_i) + \epsilon_i \quad i = 1, \ldots, n, \tag{3.1}$$

where $\{\epsilon_i; i = 1, \ldots, n\}$ are independently distributed residuals, with mean zero and constant variance σ^2.

We want to estimate the spatial field f by minimizing the following penalized

sum-of-square-error functional:

$$J_\lambda(f) = \sum_{i=1}^{n} (z_i - f(\mathbf{p}_i))^2 + \lambda \int_\Omega (\nabla \cdot K(\mathbf{p}) \nabla f)^2 , \tag{3.2}$$

where λ is the smoothing parameter, the operator ∇ is defined as $\nabla = (\frac{\partial}{\partial x}, \frac{\partial}{\partial y})^T$ and $K(\mathbf{p})$ is a function defined on Ω taking values in the space of symmetric and positive definite 2×2-matrices. $K(\mathbf{p})$ defines the non-stationary spatial anisotropy characterizing the differential operator $\nabla \cdot K(\mathbf{p}) \nabla f$. In this setting, $K(\mathbf{p})$ has the same meaning of the matrix K considered in Chapter 2, but in this case it is not constant over the spatial domain Ω. Moreover, it is not estimated from the data, but exploiting information which comes from a different source with respect to the data.

We estimate $K(\mathbf{p})$ using the method proposed by Della Rossa et al. (2010), which was developed to model traffic flows in complex networks spanning two-dimensional regions, such as large urbanized areas. In Della Rossa et al. (2010), the traffic flow is modelled with a diffusion-advection equation characterized by two parameters which can be seen as the permeability matrix (or diffusion tensor) and the drift vector (or motion trend). These two parameters are estimated over a tessellation of the domain by means of an algorithm based on cellular automata that exploits the geometry of the network. Indeed, the method starts form a black-and-white image of the map of the roads network such as those represented in Figure 3.2 and simulates random walks along the sub-networks contained in the cells of the tessellation using a multi-start Monte Carlo algorithm. The estimators of the permeability and of the drift are obtained averaging over all the trajectories the mean and the covariance of the increments of each trajectory. For the purpose of our analysis we are interested in the estimate of the permeability matrix, which we use as estimate of $K(\mathbf{p})$.

In our implementation of the method, we estimate $K(\mathbf{p})$ on a fine regular square tessellation of the domain, obtaining an estimated function $K(\mathbf{p})$ that is piecewise constant on each subdomain of the tessellation. The estimation of $K(\mathbf{p})$ is done independently on each subdomain, considering only the part of the road network within it. For each subdomain, the algorithm simulates the motion of 20000 cars through it, considering both directions of travel for each road. Each car starts its motion from one of the intersections of the roads of the network with the border of the considered subdomain, with a starting velocity of 15km/h. The displacement of each car follows the road on which the car is, and is proportional to its velocity, which decreases if the car is approaching an intersection and otherwise increases up to a maximum velocity of 105 km/h. After a car reaches an intersection, it starts again its motion following one of the roads adjacent to that intersection, choosing randomly with a uniform probability. Each car stops its motion when it reaches the boundary of the considered subdomain; the algorithm stops when all the cars have reached the boundary. The matrix K is

then computed as the mean over the 20000 replicates of the covariance matrix of the displacements of which each trajectory is composed.

3.3 Analysis of Telecom Italia Erlang data

In order to define $K(\mathbf{p})$, we start from the map of the road network in the metropolitan area of Milan. The website of Lombardy region[1] provides the map of the main roads, secondary roads, highways and railways, which is shown in the left panel of Figure 3.2. We do not consider railways since the technique applied to estimate $K(\mathbf{p})$ is based on modeling cars flowing along roads. Moreover, we exclude from our analysis the secondary roads, since a visual inspection of the data suggests that the local spatial anisotropy is mainly due to main roads. Therefore, the road network we consider in the following takes into account main roads and highways and it is represented in the right panel of Figure 3.2.

Figure 3.2: Road network in the metropolitan area of Milan. Left panel: main roads, secondary roads, highways and railways. Right panel: main roads and highways. Data from www.geoportale.regione.lombardia.it.

The results of the estimation of $K(\mathbf{p})$ are shown in Figure 3.3 on the fine regular grid of spatial locations $\{\mathbf{p}_i; i = i, ..., N\}$ corresponding to the centers of the square subdomains of the tessellation. The symmetric and positive definite matrices $K(\mathbf{p}_i)$ are represented as ellipses whose axes are oriented according to the eigenvectors of K and have length proportional to the corresponding eigenvalues. The top panel shows the estimated $K(\mathbf{p})$ on the whole domain of the metropolitan area of Milan, while the bottom panel shows a detail of the south-east region of the domain. The resulting $K(\mathbf{p})$ models precisely the roads, assuming isotropic values in the regions without roads and strongly anisotropic values along the roads.

[1] http://www.geoportale.regione.lombardia.it

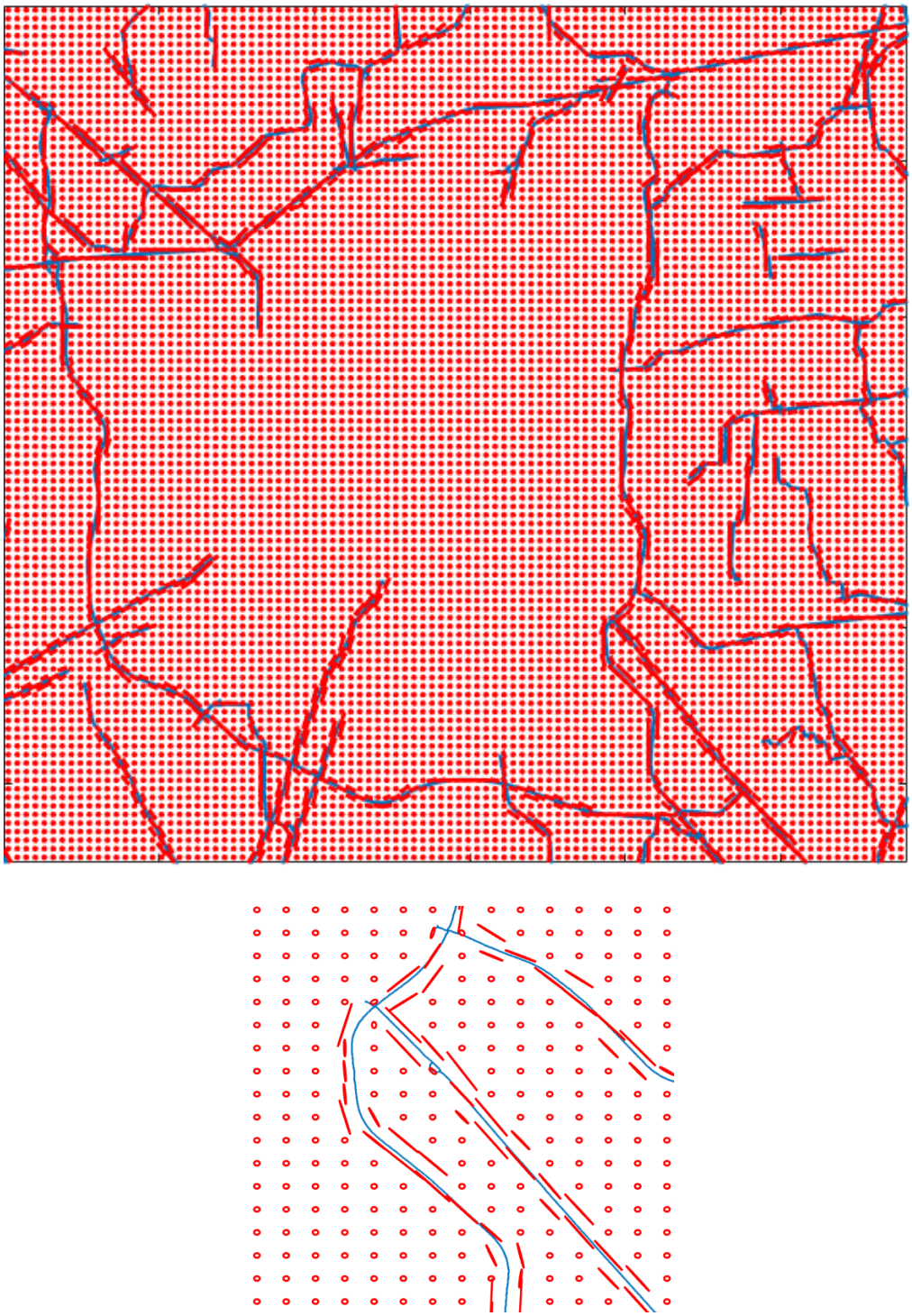

Figure 3.3: Estimated non-stationary anisotropy $K(\mathbf{p})$ represented via ellipses. The blue lines represent the road network used for the estimation of $K(\mathbf{p})$. Top panel: whole domain. Bottom panel: detail.

As a preprocessing step prior to the analysis of the Eralng data, negative values have been removed and treated as missing data since they are non-admissible and the logarithmic transformation $\text{sign}(x)\log(|x|+1)$ has been applied to he data.

The result of the estimation of the spatial field performed with the spatial regression with non-stationary anisotropic penalty is represented in the right panel of Figure 3.4, while the left panel of the same figure shows as a comparison the estimated field obtained with spatial regression with isotropic penalty. Figure 3.5 represents the difference between the two estimated fields. The difference is higher along the main roads, where the non-stationary anisotropic technique provides a more accurate estimate of the field.

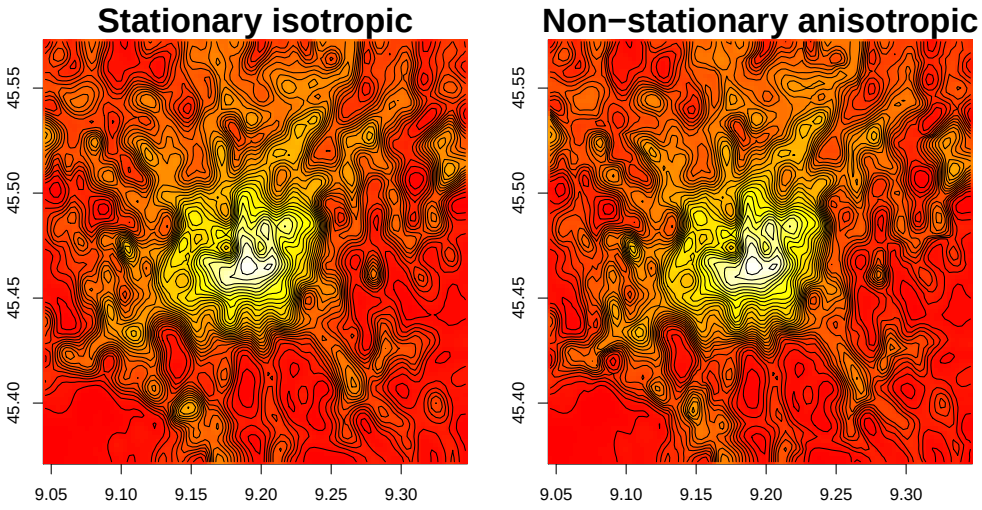

Figure 3.4: Estimated spatial fields with spatial regression with isotropic penalty (left panel) and spatial regression with non-stationary anisotropic penalty (right panel).

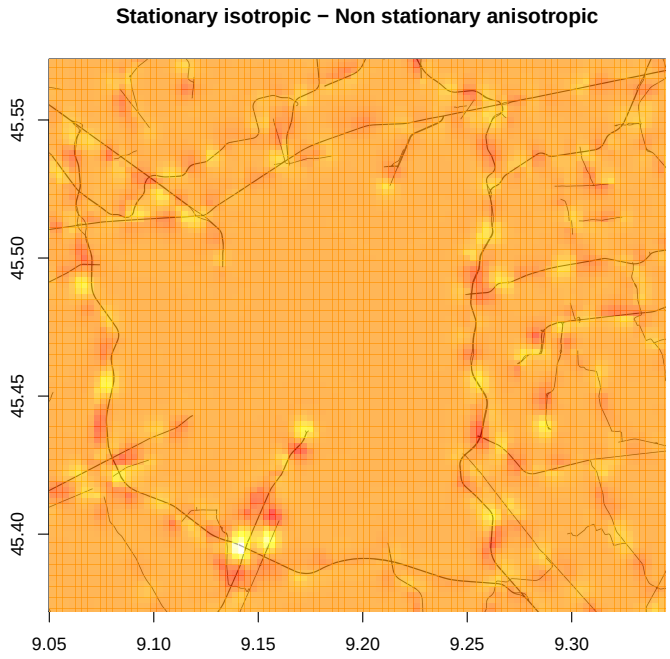

Figure 3.5: Difference between the estimated spatial fields with spatial regression with isotropic penalty and spatial regression with non-stationary anisotropic penalty. The black lines represent the road network used for the estimation of $K(\mathbf{p})$.

3.4 Model extensions

As mentioned in the Introduction, Telecom Italia Erlang data are collected over a time period of 14 days; therefore, to proper analyze these data, the temporal dependence should be taken into account. A possible approach is the one we propose in Chapter 1, where the penalty term is the sum of a spatial regularization and a temporal regularization. Moreover, the Erlang measurements are integral in time and space, since each datum refers to a temporal interval of length 15 minutes and to a square site with dimension 232m×309m. This characteristic of the data can be modeled considering the framework of areal data (see Azzimonti et al. (2015)).

In this setting, we consider $\{\Omega_i; i = 1, ..., n\}$, n sub-domains of the spatial domain Ω, and $\{[T_{j-1}, T_j]; j = 1, ..., m\}$, m sub-intervals of a temporal domain $[T_0, T_m] \subset \mathbb{R}$. In the case of Erlang data, the spatial sub-domains are the sites of the tessellation of the metropolitan area of Milan, while the temporal sub-intervals have length 15 minutes. Let z_{ij} be the measurements of the quantity of interest obtained from an underlying smooth function $f : \Omega \times [T_0, T_m] \to \mathbb{R}$ in the following way:

$$z_{ij} = \int_{T_{j-1}}^{T_j} \int_{\Omega_i} f(\boldsymbol{p}, t) d\boldsymbol{p} dt + \eta_{ij} \qquad i = 1, \dots, n \quad j = 1, \dots, m, \tag{3.3}$$

where $\{\eta_{ij}; i = 1, \ldots, n; j = 1, \ldots, m\}$ are independently distributed residuals, with mean zero and variance proportional to $|\Omega_i|(T_j - T_{j-1})$, where $|\Omega_i|$ is the area of the spatial domain Ω_i.

In this setting, the penalized sum-of-square-error functional becomes:

$$J_{\boldsymbol{\lambda}}(f) = \sum_{i=1}^{n}\sum_{j=1}^{m} \frac{1}{|\Omega_i|(T_j - T_{j-1})} \left(z_{ij} - \int_{T_{j-1}}^{T_j} \int_{\Omega_i} f(\boldsymbol{p}, t) d\boldsymbol{p} dt \right)^2 + \lambda_S \int_{T_0}^{T_m} \int_{\Omega} \Big(\nabla \cdot K(\boldsymbol{p}) \nabla f(\boldsymbol{p}, t) \Big)^2 d\boldsymbol{p} dt + \lambda_T \int_{\Omega} \int_{T_0}^{T_m} \Big(\frac{\partial^2 f(\boldsymbol{p}, t)}{\partial t^2} \Big)^2 dt d\boldsymbol{p}. \quad (3.4)$$

Another possible extension would be the inclusion of other sources of information in the definition of the regularizing differential operator. Indeed, the distribution of the data suggests that the presence of railway stations and of the Linate airport plays an important role in the phenomenon under analysis. Therefore, in the same way as the road network suggests a non-stationary anisotropic diffusion term, the railway stations and the Linate airport may suggest a non-stationary forcing term.

Another interesting line of research that could be developed in future work could be the extension to handle data with complex spatio-temporal dependence distributed over non-planar domains by integrating the approach described in this thesis with the approaches described in Ettinger et al. (2016) and Dassi et al. (2015).

Bibliography

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

Azzimonti, L., Sangalli, L. M., Secchi, P., Domanin, M. and Nobile, F. (2015), Blood flow velocity field estimation via spatial regression with PDE penalization, *Journal of the American Statistical Association* **110**(511), 1057–1071.

Azzimonti, L., Nobile, F., Sangalli, L. M. and Secchi, P. (2014), Mixed finite elements for spatial regression with pde penalization, *SIAM/ASA Journal on Uncertainty Quantification* **2**(1), 305–335.

Azzimonti, L., Sangalli, L. M., Secchi, P., Romagnoli, S. and Domanin, M. (2012), Pde penalization for spatial fields smoothing, *in* 'Proceedings of SIS Scientific Meeting 2012, 46th Scientific Meeting of the Italian Statistical Society'.

Bernardi, M. S., Carey, M., Ramsay, J. O. and Sangalli, L. M. (2017), Modelling spatial anisotropy via regression with partial differential regularization, *Submitted* .

Bernardi, M. S., Sangalli, L. M., Mazza, G. and Ramsay, J. O. (2016), A penalized regression model for spatial functional data with application to the analysis of the production of waste in venice province, *Stochastic Environmental Research and Risk Assessment, DOI: 10.1007/s00477-016-1237-3* .

Blanchet, J. and Davison, A. C. (2011), Spatial modeling of extreme snow depth, *The Annals of Applied Statistics* pp. 1699–1725.

Brown, P. E. (2015), Model-based geostatistics the easy way, *Journal of Statistical Software* **63**(12), 1–24.
URL: *http://www.jstatsoft.org/v63/i12/*

Budrikaité, L. and Dučinskas, K. (2005), Modelling of geometric anisotropic spatial variation, *Mathematical Modelling and Analysis* pp. 361–366.

Caballero, W., Giraldo, R. and Mateu, J. (2013), A universal kriging approach for spatial functional data, *Stochastic Environmental Research and Risk Assessment* **27**(7), 1553–1563.

Cao, J., Fussmann, G. F. and Ramsay, J. O. (2008), Estimating a predator-prey dynamical model with the parameter cascades method, *Biometrics* **64**(3), 959–967.

Cao, J. and Ramsay, J. (2010), Linear mixed-effects modeling by parameter cascading, *Journal of the American Statistical Association* **105**(489).

Cao, J. and Ramsay, J. O. (2007), Parameter cascades and profiling in functional data analysis, *Computational Statistics* **22**(3), 335–351.

Cao, J. and Ramsay, J. O. (2009), Generalized profiling estimation for global and adaptive penalized spline smoothing, *Computational Statistics & Data Analysis* **53**(7), 2550–2562.

Cressie, N. and Wikle, C. K. (2011), *Statistics for spatio-temporal data*, John Wiley & Sons.

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

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

Della Rossa, F., D'Angelo, C. and Quarteroni, A. (2010), A distributed model of traffic flows on extended regions., *Networks and Heterogeneous Media* **5**(3), 525–544.

Dubois, G., Malczewski, J. and De Cort, M. (2003), *Mapping radioactivity in the environment: spatial interpolation comparison 97*, Office for Official Publications of the European Communities.

Ecker, M. D. and Gelfand, A. E. (2003), Spatial modeling and prediction under stationary non-geometric range anisotropy, *Environmental and Ecological Statistics* **10**(2), 165–178.

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

Fuglstad, G.-A., Lindgren, F., Simpson, D. and Rue, H. (2015), Exploring a new class of non-stationary spatial gaussian random fields with varying local anisotropy, *Statistica Sinica* **25**, 115–133.

Fuglstad, G.-A., Simpson, D., Lindgren, F. and Rue, H. (2013), Non-stationary spatial modelling with applications to spatial prediction of precipitation, *arXiv preprint arXiv:1306.0408* .

Fuglstad, G.-A., Simpson, D., Lindgren, F. and Rue, H. (2014), Do we need non-stationarity in spatial models?, *arXiv preprint arXiv:1409.0743* .

Fuglstad, G.-A., Simpson, D., Lindgren, F. and Rue, H. (2015), Does non-stationary spatial data always require non-stationary random fields?, *Spatial Statistics* **14**, 505–531.

Giraldo, R., Delicado, P. and Mateu, J. (2011), Ordinary kriging for function-valued spatial data, *Environmental and Ecological Statistics* **18**(3), 411–426.

Goulard, M. and Voltz, M. (1993), Geostatistical interpolation of curves: A case study in soil science, *in* 'Geostatistics Tróia'92', Springer, pp. 805–816.

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

Ignaccolo, R., Mateu, J. and 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, L. M., Ramsay, J. and Formaggia, L. (2016), *fdaPDE: Functional Data Analysis and Partial Differential Equations; Statistical Analysis of Functional and Spatial Data, Based on Regression with Partial Differential Regularizations*. R package version 0.1-5.

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, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **73**(4), 423–498.

Manfredini, F., Pucci, P., Secchi, P., Tagliolato, P., Vantini, S. and Vitelli, V. (2015), Treelet decomposition of mobile phone data for deriving city usage and mobility pattern in the milan urban region, *in* 'Advances in complex data modeling and computational methods in statistics', Springer, pp. 133–147.

Marra, G., Miller, D. L. and 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. et al. (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. and 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. and Manté, C. (2010), Cokriging for spatial functional data, *Journal of Multivariate Analysis* **101**(2), 409–418.

Nychka, D., Furrer, R., Paige, J. and Sain, S. (2015), 'fields: Tools for spatial data'. R package version 8.4-1.
URL: *www.image.ucar.edu/fields*

Passamonti, F. (2016), Spatio-temporal mobile phone data in milan: Bagging-voronoi exploration and modeling through soil use and land cover data, Master's thesis, Politecnico di Milano, MOX - Dipartimento di Matematica.

Pebesma, E. J. (2004), Multivariable geostatistics in s: the gstat package, *Computers & Geosciences* **30**, 683–691.

R Core Team (2016), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.
URL: *https://www.R-project.org/*

Ramsay, J. O., Hooker, G., Campbell, D. and Cao, J. (2007), Parameter estimation for differential equations: a generalized smoothing approach, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **69**(5), 741–796.

Ramsay, J. O. and Silverman, B. W. (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.

Ribeiro Jr, P. J. and Diggle, P. J. (2016), *geoR: Analysis of Geostatistical Data*. R package version 1.7-5.2.
URL: *https://CRAN.R-project.org/package=geoR*

Sangalli, L. M., Ramsay, J. O. and Ramsay, T. O. (2013), Spatial spline regression models, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **75**(4), 681–703.

Schlather, M., Malinowski, A., Oesting, M., Boecker, D., Strokorb, K., Engelke, S., Martini, J., Ballani, F., Moreva, O., Menck, P. J., Gross, S., Ober, U., Christoph Berreth, Burmeister, K., Manitz, J., Morena, O., Ribeiro, P., Singleton, R., Pfaff, B. and R Core Team (2016), *RandomFields: Simulation and Analysis of Random Fields*. R package version 3.1.16.
URL: *http://CRAN.R-project.org/package=RandomFields*

Secchi, P., Vantini, S. and Vitelli, V. (2015), Analysis of spatio-temporal mobile phone data: a case study in the metropolitan area of milan, *Statistical Methods & Applications* **24**(2), 279–300.

Wahba, G. (1990), *Spline models for observational data*, Vol. 59 of *CBMS-NSF Regional Conference Series in Applied Mathematics*, Society for Industrial and Applied Mathematics, Philadelphia, Pennsylvania.

Wilhelm, M. and Sangalli, L. M. (2016), Generalized spatial regression with differential regularization, *Journal of Statistical Computation and Simulation* pp. 1–22.

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

Wood, S. N., Bravington, M. V. and Hedley, S. L. (2008), Soap film smoothing, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **70**(5), 931–955.

Xun, X., Cao, J., Mallick, B., Maity, A. and Carroll, R. J. (2013), Parameter estimation of partial differential equation models, *Journal of the American Statistical Association* **108**(503), 1009–1020.

Zanini, P., Shen, H. and Truong, Y. (2016), Understanding resident mobility in milan through independent component analysis of telecom italia mobile usage data, *The Annals of Applied Statistics* **10**(2), 812–833.