Modeling spatial anisotropy via regression with partial differential regularization

Mara S. Bernardi · **Michelle Carey** ·
Laura M. Sangalli · **James O. Ramsay**

Abstract We consider the problem of analyzing spatially distributed data characterized by spatial anisotropy. Following a functional data analysis approach, we propose a method based on regression with partial differential regularization, where the differential operator in the regularizing term is anisotropic and is derived from data. We show that the method correctly identifies the direction and intensity of anisotropy and returns an accurate estimate of the spatial field. The method compares favorably to both isotropic and anisotropic kriging, as tested in simulation studies under various scenarios. The method is then applied to the analysis of Switzerland rainfall data.

Keywords: Finite elements; Functional data analysis; Parameter cascading; Penalized regression

1 Introduction

Many, if not most, spatial phenomena are characterized by spatial anisotropy. 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. Figure 1 depicts a dataset of 467 daily rainfall measurements recorded in Switzerland on May 8, 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.

In this work we adopt a functional data analysis approach and propose to model the spatial anisotropy via regression with partial differential regularization. Ramsay (2002), Sangalli et al. (2013) and Wood et al. (2008) 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; 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 regression with differential regularization. Azzimonti et al. (2015) extend the method in Ramsay (2002)

M. S. Bernardi
MOX - Department of Mathematics, Politecnico di Milano, Milano, Italy

M. Carey
School of Mathematics and Statistics, University College Dublin, Ireland

L. M. Sangalli (✉)
MOX - Department of Mathematics, Politecnico di Milano, Milano, Italy
E-mail: laura.sangalli@polimi.it

J. O. Ramsay
Department of Psychology, McGill University, Montreal, Canada
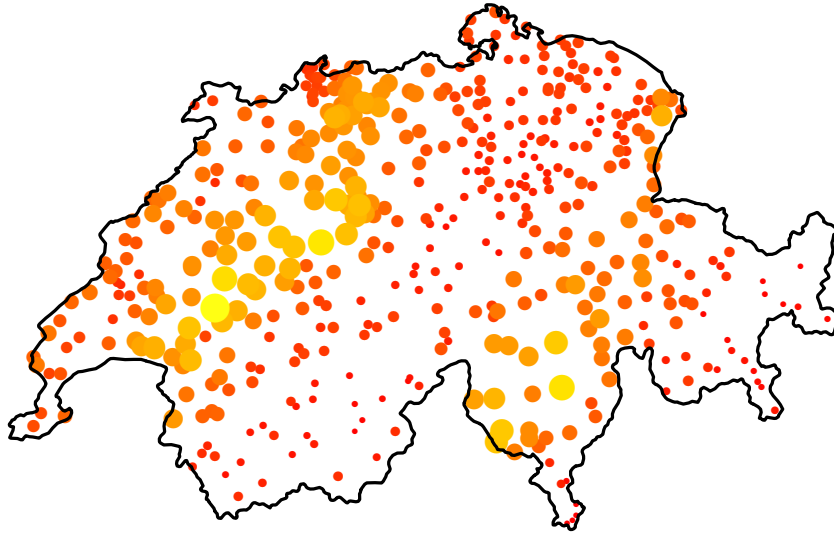
Fig. 1: Switzerland rainfall data. These include 467 daily rainfall measurements recorded in Switzerland on May 8, 1986. The size and color of point markers represent the value of the rainfall at each location, highlighting a strong spatial anisotropy.

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. In particular, Azzimonti et al. (2015) assume that the PDE in the regularizing term 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; moreover, we do not assume the existence of a physical law governing the system. We though use the PDE in the regularizing term to model the spatial variation of the phenomenon, learning the anisotropy directly from the data. Specifically, we consider PDEs that induce a stationary anisotropic smoothing effect; the parameters in the PDE determine the direction and the intensity of the anisotropy; these parameters are here considered unknown and are derived from data. In particular, we select 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. The same 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 parameter cascading to the estimation of 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. In fact, as mentioned earlier, the PDE is not here used to model the phenomenon which generated the data, but rather as a regularizing term, to characterize the spatial distribution of the data. Moreover, spatially varying covariates are included in the model.

Other spatial smoothing techniques which take into account the anisotropy of the data have been considered by Wood (2000), which describes an anisotropic thin plate spline smoother, and by Gu (2013), which presents a tensor product spline model. Unlike these techniques, the method we propose does not use global radial basis functions, nor tensor-product basis functions. In particular, likewise Azzimonti et al. (2015), Ramsay (2002) and Sangalli et al. (2013), we represent the spatial field via finite elements, which provide a non-tensor product basis for piecewise polynomial surfaces. This enable the method to efficiently handle data distributed over irregularly shaped domains, featuring concavities and interior holes, as extensively shown for instance in Bernardi et al.

(2017), Ramsay (2002) and Sangalli et al. (2013), and as here illustrated in a simulation study. The implementation of the model is based on the R package fdaPDE (Lila et al., 2016).

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 in this simulation setting 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 outperforms all other methods when generating data from a field defined over an irregularly shaped domain.

The paper is organized as follows. Section 2 describes the model. Section 3 provides implementation details. Section 4 extends the model for the inclusion of space-varying covariates. Section 5 is devoted to simulation studies. Section 6 shows the application of the proposed method to the analysis of the Switzerland rainfall data. Finally, Section 7 draws some directions for future research.

2 Model

Let $\{\mathbf{p}_1 = (x_1, y_1), \ldots, \mathbf{p}_n = (x_n, y_n)\}$ be a set of n points on a bounded domain $\Omega \in \mathbb{R}^2$, whose boundary $\partial \Omega$ is a curve of class $\mathscr{C}^2$. 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 $z_1, \ldots, z_n$ are noisy observations of an underlying smooth function $f : \Omega \to \mathbb{R}$. That is, for all $i \in \{1, \ldots, n\}$,

$$z_i = f(\mathbf{p}_i) + \varepsilon_i \tag{1}$$

where $\varepsilon_1, \ldots, \varepsilon_n$ are independently distributed residuals, with mean zero and constant variance σ^2.

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

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

where the operator ∇ is defined as $\nabla = (\partial/\partial x, \partial/\partial y)^\top$ and K is a symmetric and positive definite matrix with elements K_{ij}, for $i, j \in \{1, 2\}$. The inclusion in the regularizing term of the anisotropic diffusion operator

$$\nabla \cdot (K \nabla f) = K_{11} \frac{\partial^2 f}{\partial x^2} + (K_{12} + K_{21}) \frac{\partial^2 f}{\partial x \partial y} + K_{22} \frac{\partial^2 f}{\partial y^2}$$

provides an anisotropic smoothing effect, where the direction and intensity of the anisotropy is determined by the matrix K. In particular, the eigenvectors of K determine the directions of maximum and minimum smoothing, while the corresponding eigenvalues control the intensity of the smoothing in each direction, with the ratio between the two eigenvalues determining the intensity of the anisotropy. Figure 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, 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 Laplace operator $\Delta f = \nabla \cdot \nabla f = \partial^2 f / \partial x^2 + \partial^2 f / \partial y^2$. In the latter 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 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 the example represented in the right panel.

Beside the matrix K, which controls the direction and intensity of the regularization, there is another tuning parameter: the smoothing parameter $\rho \in (0, 1)$, which weighs the contribution of the data fitting term (the least
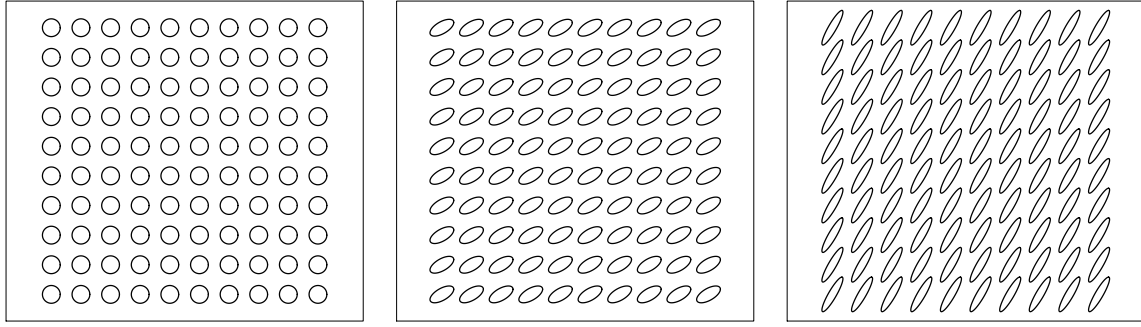
Fig. 2: Graphical representation of the smoothing effect implied by stationary isotropic (left) and anisotropic (center and right) penalizations. The anisotropy matrix K is represented by ellipses: its eigenvectors (the ellipses axes) identify the direction of the anisotropy, while the eigenvalues (proportional to the axes lengths) determine the intensity of the anisotropy. The angles of the first eigenvectors are $\pi/6$ and $\pi/3$, respectively, for the second and third case, while the ratios between the eigenvalues are 5 and 25, respectively.

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 Azzimonti et al. (2015), Ramsay (2002) and Sangalli et al. (2013). The new parametrization of the smoothing parameter in the functional (2) is introduced to more efficiently balance the least squares term and the regularizing term, using a parameter that has a bounded range; moreover, the least squares and the regularizing term are both normalized respectively by the number of observations and the area of the domain.

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 chosen to properly weigh the effect of the regularization and to accurately identify the optimal anisotropy matrix K. We here propose a two-step algorithm which alternates between the optimal selection of the anisotropy K and the estimation of spatial field f. The optimal anisotropy matrix K is selected using a parameter cascading approach, as described in Section 2.1. Hence, the spatial field f is estimated using the anisotropy matrix K selected in the previous step of the algorithm. In particular, in Section 2.2 we show that, for any given K, the estimation of the spatial field f, minimizing the functional $J_\rho(f,K)$, has a unique solution. This infinite-dimensional estimation problem is discretized via finite elements, as detailed in the Sections 2.3 and 2.4.

2.1 Selection of the anisotropy matrix K via parameter cascading

To select the anisotropy matrix K we use parameter cascading. Ramsay et al. (2007) and Xun et al. (2013) use this technique respectively to 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 differential equation, but this 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 3.2.

The parameter cascading algorithm distinguishes two classes of parameters: 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 selecting 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 $\hat{f}_{\rho,K}$ 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

$$H(K) = \sum_{i=1}^{n} \{z_i - \hat{f}_{\rho,K}(\mathbf{p}_i)\}^2, \tag{3}$$

which depends on K implicitly, through $\hat{f}_{\rho,K}$. Since $\hat{f}_{\rho,K}$ 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 Azzimonti et al. (2015) and Sangalli et al. (2013), as detailed in the following sections.

2.2 Estimation of the spatial field f for a given anisotropy matrix K

Consider the Sobolev space $H^2(\Omega)$ of functions f such that f and its first and second derivatives belong to $L^2(\Omega)$; see, e.g., Rudin (1991). The functional $J_\rho(f,K)$ in (2) is well defined for $f \in H^2(\Omega)$. Moreover, as shown in the following proposition, given a symmetric and positive definite matrix K, the minimizer of $J_\rho(f,K)$ exists and is unique for $f \in H^2(\Omega)$ with appropriate boundary conditions. In particular, the boundary conditions ensure the uniqueness of the solution. Various boundary conditions may be considered, concerning for instance the value of f and/or the value of the normal derivative of f at the boundary $\partial\Omega$, thus enabling a very flexible modeling of the behavior of the field at the boundary of the domain of interest. Specifically, Dirichlet conditions control the value of the function at the boundary, i.e., $f|_{\partial\Omega} = \gamma_D$; Neumann conditions set the value of the normal derivative of the function at the boundary, i.e., $(K\nabla f \cdot \boldsymbol{v})|_{\partial\Omega} = \gamma_N$, where $\boldsymbol{v}$ is the outward unit normal vector to $\partial\Omega$, thus controlling the flow across the boundary; Robin conditions involve a linear combination of the value of the first derivative and the value of the function at the boundary, i.e., $(K\nabla f \cdot \boldsymbol{v} + \chi f)|_{\partial\Omega} = \gamma_R$. Different boundary conditions may be imposed on different portions of the boundary, that form a partition of $\partial\Omega$. All the admissible boundary conditions can be summarized as

$$\begin{cases} f = \gamma_D & \text{on } \Gamma_D, \\ K\nabla f \cdot \boldsymbol{v} = \gamma_N & \text{on } \Gamma_N, \\ K\nabla f \cdot \boldsymbol{v} + \chi f = \gamma_R & \text{on } \Gamma_R, \end{cases}$$

where γ_D, γ_N, γ_R, Γ_D, Γ_N, and Γ_R have to satisfy some regularity conditions to obtain a well-defined functional $J_\rho(f,K)$. See Azzimonti et al. (2014) for details. For simplicity of exposition, in the following we consider homogeneous conditions, i.e., null γ_D, γ_N and γ_R. All other boundary conditions may be handled similarly to Azzimonti et al. (2014). We denote by $V(\Omega)$ the subspace of $H^2(\Omega)$ characterized by the chosen boundary conditions. We set $\mathbf{z} = (z_1, \ldots, z_n)^\top$ and, for any function $h : \Omega \to \mathbb{R}$, we set $\mathbf{h}_n = (h(\mathbf{p}_1), \ldots, h(\mathbf{p}_n))^\top$. The estimator is characterized by the following proposition.

Proposition 1 *Given a symmetric and positive definite matrix K, there exists a unique estimator $\hat{f} \in V(\Omega)$ which minimizes (2). Moreover, $\hat{f}$ satisfies, for every $h \in V(\Omega)$,*

$$(1-\rho)\frac{1}{n}\mathbf{h}_n^\top \hat{\mathbf{f}}_n + \rho \frac{1}{|\Omega|} \int_\Omega \{\nabla \cdot (K\nabla h)\}\{\nabla \cdot (K\nabla \hat{f})\} = (1-\rho)\frac{1}{n}\mathbf{h}_n^\top \mathbf{z}. \tag{4}$$

Proof The result follows from Theorem 2 in Azzimonti et al. (2014), by appropriate reparametrization of the smoothing parameter ρ, and by setting $L = \nabla \cdot K\nabla f$ and $u = 0$.

The fourth-order problem (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 allows us 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.
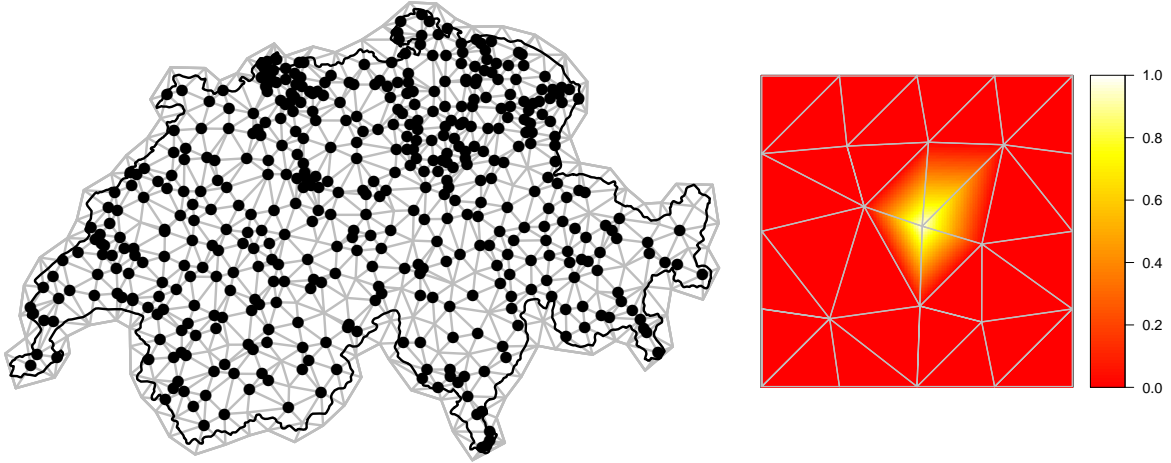
Fig. 3: Left: domain and locations of Switzerland rainfall data, with simplified domain boundary and triangulation in gray. Right: Example of linear finite element basis function.

2.3 Finite elements

To construct a finite element space, we first obtain a regular triangulation $\mathscr{T}$ 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 $\Omega_{\mathscr{T}}$, and the boundary $\partial\Omega$ is approximated by a polygon (or multiple polygons in case of a domain with holes).

As an example, the left panel of Figure 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. This simplified boundary is shown in gray in the figure, where the black line instead represents the actual border of Switzerland, and the black dots indicate the data locations. We first obtain a constrained Delaunay triangulation selecting as vertices the locations of the data and the points defining the simplified boundary; the triangulation is then refined with additional vertices imposing a constraint on the maximum value of the triangle areas. The constrained Delaunay triangulation and the refined mesh are obtained using the R package fdaPDE (Lila et al., 2016). The final triangulation is shown in gray in the figure.

On the triangulation $\mathscr{T}$, we define the finite element space $V^r_{\mathscr{T}}(\Omega)$, with $r \in \{1, 2, \dots\}$, as the space of continuous surfaces over $\Omega_{\mathscr{T}}$ that are polynomials of degree r when restricted to any triangle in $\mathscr{T}$. To define a set of N basis functions $\psi_1, \dots, \psi_N$, that span such space, it is convenient to consider the so-called nodes of the triangulation, denoted by $\boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_N$. For linear finite elements, the nodes coincide with the vertices of the triangles in $\mathscr{T}$. For higher order finite elements, the nodes are a super-set of the triangle vertices; for instance, for quadratic finite elements the nodes coincide with the triangle vertices and the middle points of the triangle edges. For each $j \in \{1, \dots, N\}$, the basis ψ_j is then associated with one node $\boldsymbol{\xi}_j$, and is a locally supported piecewise polynomial function of order r, that takes value 1 at the associated node and 0 on all other nodes, i.e., $\psi_j(\boldsymbol{\xi}_i) = \delta_{ji}$, where $\delta_{ji} = 1$ if $j = i$ and $\delta_{ji} = 0$ if $j \neq i$. Bases associated with nodes in Γ_D and Γ_R, where homogeneous Dirichlet or Robin conditions are imposed, are discarded. For the simulation studies and for the application here presented, we use linear finite elements. The right panel of Figure 3 shows an example of linear finite element basis function.

Let $\boldsymbol{\psi} = (\psi_1, \dots, \psi_N)^\top$ be the N-vector of the basis functions. Then we can represent any function $h \in V^r_{\mathscr{T}}(\Omega)$ as an expansion on this basis system, viz.

$$h(\mathbf{p}) = \mathbf{h}^\top \boldsymbol{\psi}(\mathbf{p}),$$

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

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

For an introduction to finite elements, see, e.g., Gockenbach (2006) and Quarteroni (2010).

2.4 Discretization of the spatial field f via finite elements

Let Ψ be the $n \times N$ matrix evaluating the N basis functions $\psi_1, \ldots, \psi_N$ at the n data locations $\mathbf{p}_1, \ldots, \mathbf{p}_n$:

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

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

$$R_0 = \int_{\Omega_{\mathscr{T}}} \boldsymbol{\psi}\boldsymbol{\psi}^\top, \quad R_1 = \int_{\Omega_{\mathscr{T}}} \nabla\boldsymbol{\psi}(K\nabla\boldsymbol{\psi})^\top.$$

The following proposition states that, once recast in the finite element space, finding the estimator $\hat{f}$, for a given K, reduces to solving a linear system. See the Appendix for details.

Proposition 2 *Given a symmetric and positive definite matrix K, the estimator $\hat{f} \in V_{\mathscr{T}}^r(\Omega)$ exists and is unique and is given by $\hat{f} = \hat{\mathbf{f}}\boldsymbol{\psi}$, where*

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

Proof The existence and uniqueness of the solution are ensured by the invertibility of the matrices R_0 and $\{(1-\rho)/n\}\Psi^\top\Psi + (\rho/|\Omega|)R_1 R_0^{-1} R_1$.

From Proposition 2, we obtain the following expression for $\hat{\mathbf{f}}$, the vector of coefficients of the basis expansion for the spatial field estimate $\hat{f}$:

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

where $P = R_1 R_0^{-1} R_1$ represents the discretization of the penalty term in Eq. (2). The discretization of the penalty term only involves first order derivatives, thanks to the weak formulation (8) of the estimation problem, derived in the Appendix.

3 Implementation details

3.1 Parametrization of the anisotropy matrix K

Various choices are possible for the parametrization of the anisotropy matrix K. In this work, we parametrize 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

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

The choice of this parametrization implies a periodicity of the functional H in (3) with respect to the parameter α; the numerical implementation of the optimization method appropriately takes this point into account.

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. Note that the value of ρ that enables the optimal selection of the anisotropy matrix K differs from the optimal value of ρ for the estimation of the spatial field f, for a given value of K. In fact, when selecting 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, the smoothing parameter ρ should be chosen 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 value of ρ that enabled the optimal selection of K was at least one order of magnitude larger than the optimal value of ρ for the estimation of f. Specifically, we proceed as follows.

Having no knowledge of the right amount of regularization to impose for the optimal selection of the anisotropy matrix K, we consider a grid of d possible values for the smoothing parameter $\rho_1, \ldots, \rho_d$ spanning regularly the interval $(0,1)$. The grid we use in the simulations and in the analysis of the Switzerland rainfall data is as follows: $(0.01, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9)$. For each value of the smoothing parameter in this grid, the optimal anisotropy matrix is found, thus leading to the identification of d anisotropy matrices $K_1, \ldots, K_d$. At this point, the choice of the optimal anisotropy matrix, among $K_1, \ldots, K_d$, is led by the goal of our analysis: the optimal estimation of the spatial field. In particular, we want to select the anisotropy matrix leading to the best estimation of the spatial field. To do so, we estimate an f for each of the d possible anisotropy matrices $\{K_1, \ldots, K_d\}$, obtaining the spatial field estimates $\hat{f}_1, \ldots, \hat{f}_d$; to obtain these estimates we choose in each case the smoothing parameter ρ via generalized cross validation (GCV). We then select, among the d anisotropy matrices $K_1, \ldots, K_d$ and the corresponding estimated spatial fields $\hat{f}_1, \ldots, \hat{f}_d$, the couple $K, \hat{f}$ having the smallest GCV value.

The GCV index is available in closed-form for this model, and is

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

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.,

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

The selection of the anisotropy matrix K for a grid of d 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 corresponding to its minimum. As an example, Figure 4 shows the surfaces $H\{K(\boldsymbol{\theta})\}$ for two extremes values of ρ ($\rho = 0.01$ and $\rho = 0.9$), for the Switzerland rainfall data. The optimization algorithm based on gradient descent can easily identify the global minimum for the first surface (corresponding to $\rho = 0.01$), but needs a starting point within the basin of attraction of the minimum to reach the global minimum of the second surface (corresponding to $\rho = 0.9$). In order to improve the performance of the optimization algorithm and to avoid local minima, for each value ρ_j, for $j \in \{2, \ldots, d\}$, we provide, as initial value of $\boldsymbol{\theta}$ in the optimization of $H\{K(\boldsymbol{\theta})\}$, the optimum for $\boldsymbol{\theta}$ that was obtained from the optimization for the value ρ_{j-1}. Since the complexity of the optimization surface increases as ρ increases, the initial values provided by the optimization carried out for with the previous value of ρ guarantee to remain in the basin of attraction of the global minimum. In the numerical tests we performed, we never experienced any problem of convergence of the optimization algorithm.
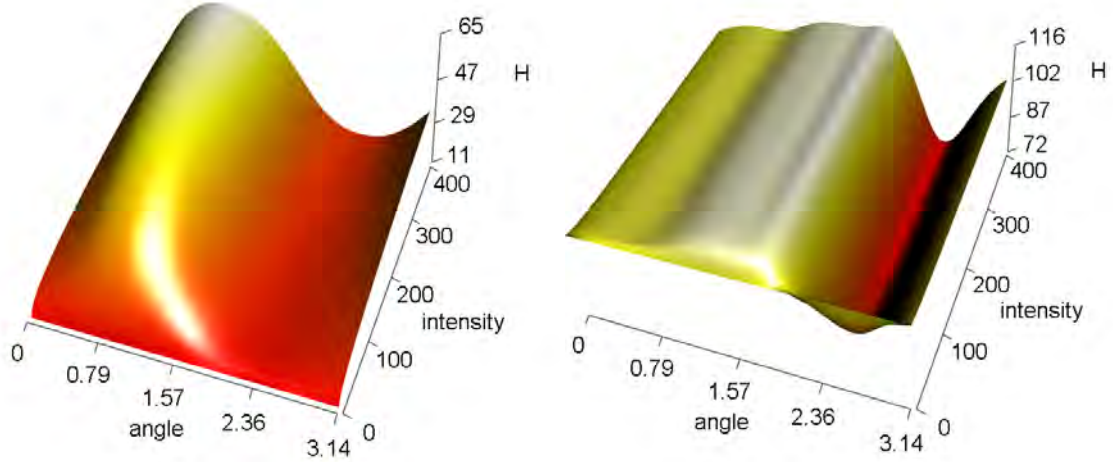
Fig. 4: Surface $H\{K(\boldsymbol{\theta})\}$ for two values of the smoothing parameter: $\rho = 0.01$ (left panel) and $\rho = 0.9$ (right panel).

4 Inclusion of space-varying covariates

The method described in the previous sections can be extended to handle space-varying covariates. Let $\mathbf{w}_i = (w_{i1}, \ldots, w_{iq})^\top$ be a q-vector of covariates observed at $\mathbf{p}_i$. The model in Eq. (1) can be modified to include an additive term which takes into account the contribution of the covariates. Specifically, we consider the semi-parametric generalized additive model defined, for all $i \in \{1, \ldots, n\}$, by

$$z_i = \mathbf{w}_i^\top \boldsymbol{\beta} + f(\mathbf{p}_i) + \varepsilon_i, \tag{5}$$

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

$$J_\rho(\boldsymbol{\beta}, f, K) = (1 - \rho) \frac{1}{n} \sum_{i=1}^n \{z_i - \mathbf{w}_i^\top \boldsymbol{\beta} - f(\mathbf{p}_i)\}^2 + \rho \frac{1}{|\Omega|} \int_\Omega \{\nabla \cdot (K\nabla f)\}^2. \tag{6}$$

The proposed two-step algorithm separately selects the matrix K and estimates the couple $\boldsymbol{\beta}, f$. The anisotropy matrix K is selected using a parameter cascading approach, minimizing the functional:

$$H(K) = \sum_{i=1}^n \{z_i - \mathbf{w}_i^\top \hat{\boldsymbol{\beta}}_{\rho,K} - \hat{f}_{\rho,K}(\mathbf{p}_i)\}^2,$$

where $\hat{\boldsymbol{\beta}}_{\rho,K}$ and $\hat{f}_{\rho,K}$ are implicit functions of the structural parameter K, and are in turn 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 chosen anisotropy matrix K, we can prove, in a similar way to the one described in Section 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 ith row is the vector $\mathbf{w}_i$ of covariates at location $\mathbf{p}_i$. We assume W has full rank. Let $P = W(W^\top W)^{-1} W^\top$ 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 3 *There exists a unique pair $\{\hat{\boldsymbol{\beta}} \in \mathbb{R}^q : \hat{f} \in V(\Omega)\}$ which minimize (6) for a fixed symmetric and positive definite matrix K. Moreover, $\hat{\boldsymbol{\beta}} = (W^\top W)^{-1} W^\top (\mathbf{z} - \hat{\mathbf{f}}_n)$ and $\hat{f}$ satisfies, for every $h \in V(\Omega)$,*

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

Proof The estimator $\hat{\boldsymbol{\beta}}$ is obtained by differentiating the functional $J_\rho(\boldsymbol{\beta}, f, K)$ in (6) with respect to $\boldsymbol{\beta}$; then, plugging the estimate $\hat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$ in (6) reduces the estimation problem to the minimization of $J_\rho(\hat{\boldsymbol{\beta}}, f, K)$ over f in $V(\Omega)$, and the result follows similarly as in the proof of Proposition 1. In particular, the imposition of boundary conditions in the space $V(\Omega)$ ensures the uniqueness of $\hat{f}$, and thus of $\hat{\boldsymbol{\beta}}$ as well. A detailed derivation of this result can be obtained combining *Proposition 1* in Sangalli et al. (2013) and Theorem 2 in Azzimonti et al. (2014).

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

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

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

Proof The existence and uniqueness of the solution are ensured by the invertibility of the matrices R_0 and $[\{(1-\rho)/n\}\Psi^\top Q\Psi + (\rho/|\Omega|)P]$.

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 $\hat{f}$:

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

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. Isotropic and anisotropic SR-PDE are implemented using the R package `fdaPDE` (Lila et al., 2016), considering homogeneous Neumann boundary conditions, i.e., null flow across the boundary. The isotropic kriging estimates, both ordinary and universal, 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, using the starting values suggested by the empirical variogram, as common practice in kriging. The anisotropic kriging estimates, both ordinary and universal, 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 in the estimation by anisotropic 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 re-run `likfit` with different initial values for the parameters, until finite estimates are attained. For some instances the instability could not be avoided; these simulation repetitions were excluded from the analysis, thus giving an implicit advantage to kriging over SR-PDE. For SR-PDE we instead never experienced any numerical instability. 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.
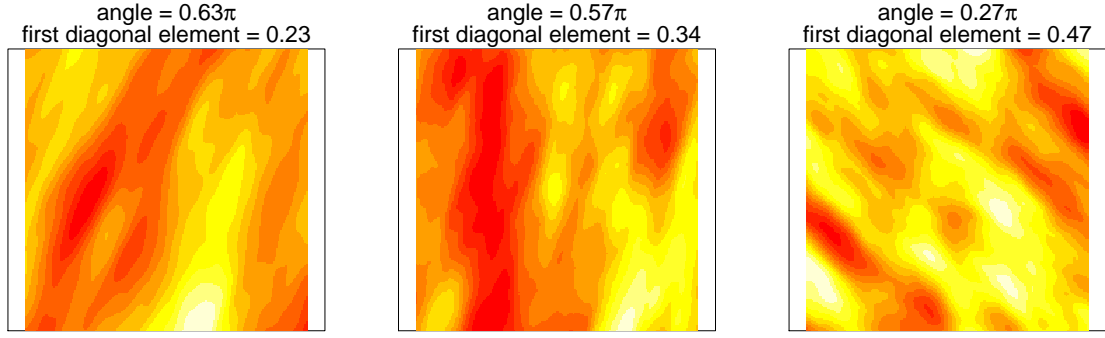
Fig. 5: First simulation study. Three generated Matérn fields with different anisotropy angles and intensities.

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 5 shows three fields generated for this first simulation study.

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).

Note that the parametrization of the anisotropy matrix used in the data generation and in the estimation by kriging is not the same as the one considered by anisotropic SR-PDE. To make these matrices comparable, one should consider the squared root of K, normalized in order to have first eigenvalue equal to 1.

The top right panel of Figure 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 respectively representing the true and the selected anisotropy matrices. In particular, to allow the visual comparison of the true anisotropy matrix, the anisotropy matrix selected by anisotropic kriging and the one selected by anisotropic SR-PDE, for the latter method we plot the squared root of K, normalized in order to have first eigenvalue equal to 1, thus considering the same parametrization for all anisotropy matrices. 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 of the estimated spatial field over the 200 simulation replicates. The RMSE is computed on a fine regular squared grid spanning the whole domain with spacing 0.05. Figure 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 any of 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 simulation setting, be the best possible model, since it assumes the same covariance structure used to generate the data.
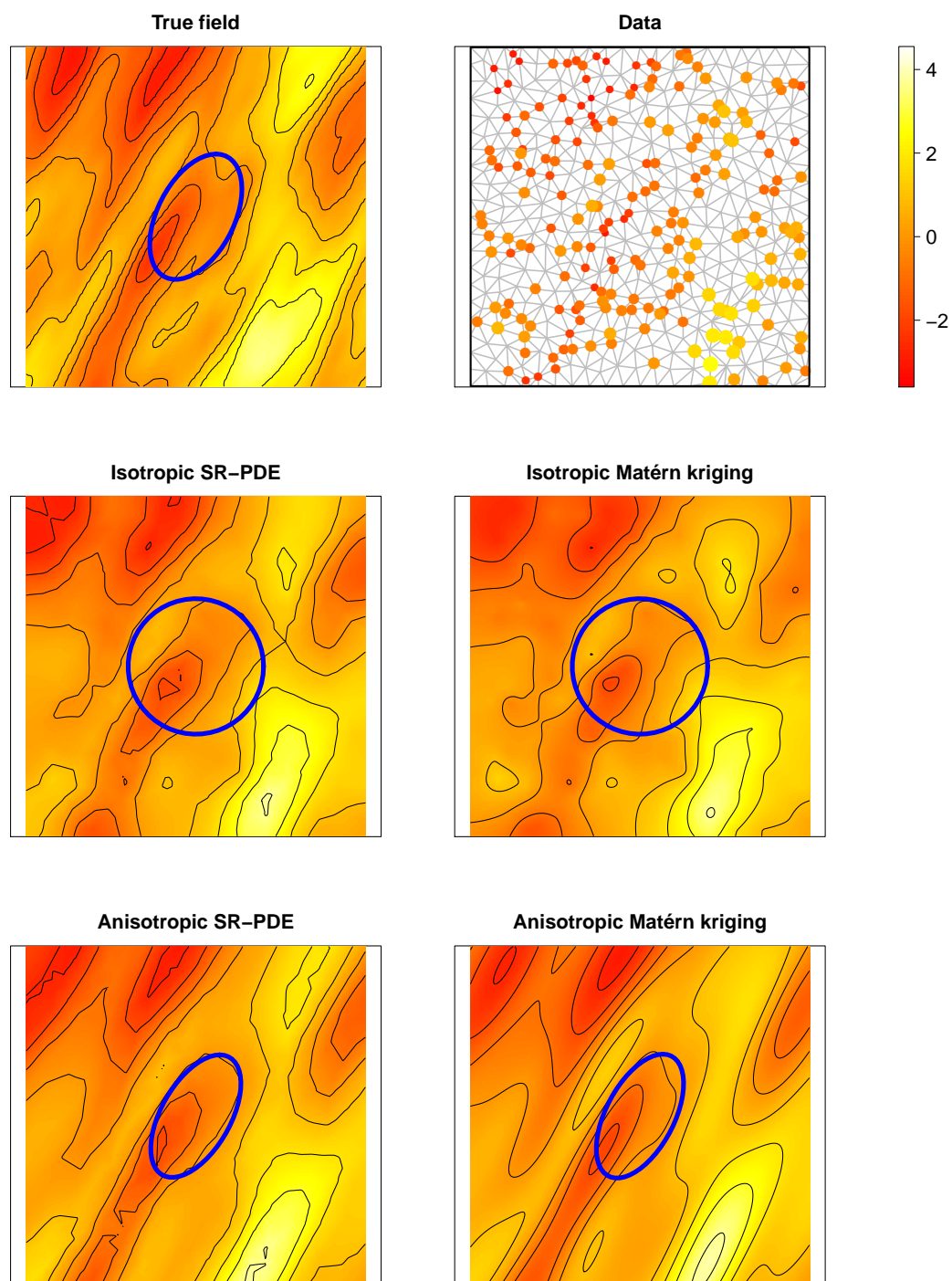
Fig. 6: First simulation study. First row: true field (left) and sampled data (right), with the triangulation used for SR-PDE estimates in gray. Second row: estimates provided by isotropic SR-PDE and isotropic Matérn kriging. Third row: estimates provided by anisotropic SR-PDE and anisotropic Matérn kriging.
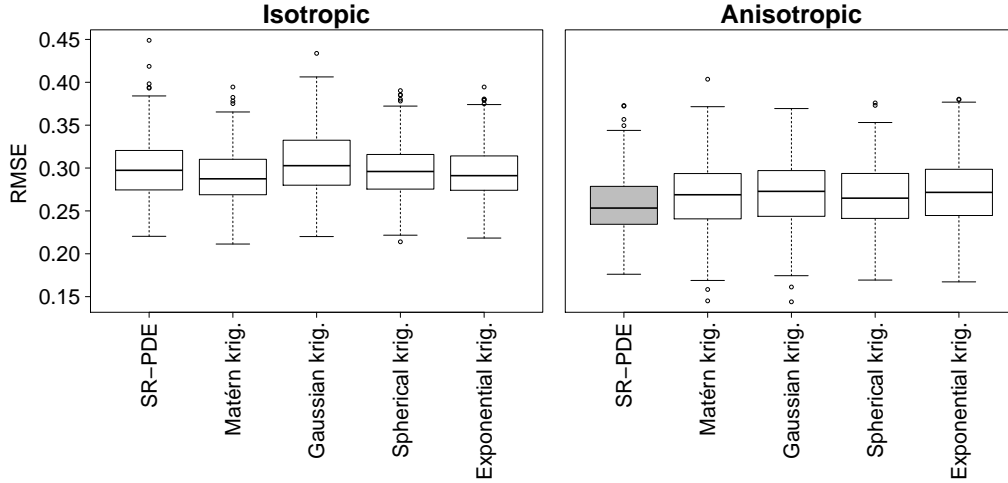
Fig. 7: First simulation study. Boxplots of the RMSE, over the 200 simulation replicates, of the spatial field estimates obtained by isotropic and anisotropic SR-PDE and by isotropic and anisotropic kriging with different variogram models (Matérn, Gaussian, Spherical and Exponential).

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 5.1, but we add a space-varying covariate. Specifically, for 200 simulation replicates, we generate from the semiparametric model (5), 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 8 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 right panel of the second row shows the data obtained in the first simulation replicate. 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. Figure 9 shows the boxplots of the RMSE of the spatial field estimates obtained with the different methods over the 200 replicates. As expected, taking into account the anisotropy significantly improves the fitting. In particular, the RMSE of the fields estimated with anisotropic SR-PDE are significantly lower than those obtained with any of 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.

True field

Covariate

Total field

Data

Isotropic SR−PDE

Isotropic Matérn kriging

Anisotropic SR−PDE
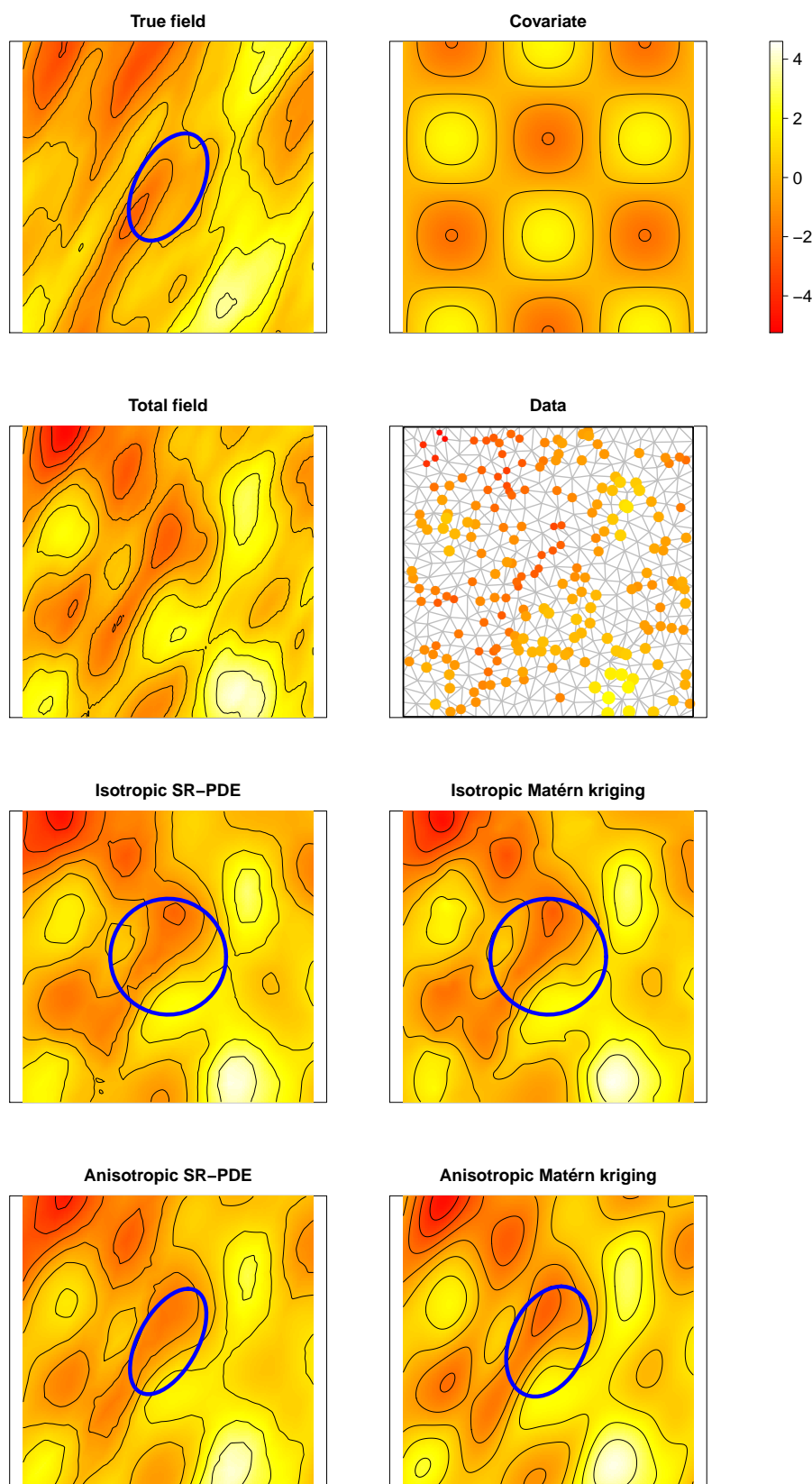
Anisotropic Matérn kriging

Fig. 8: 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.
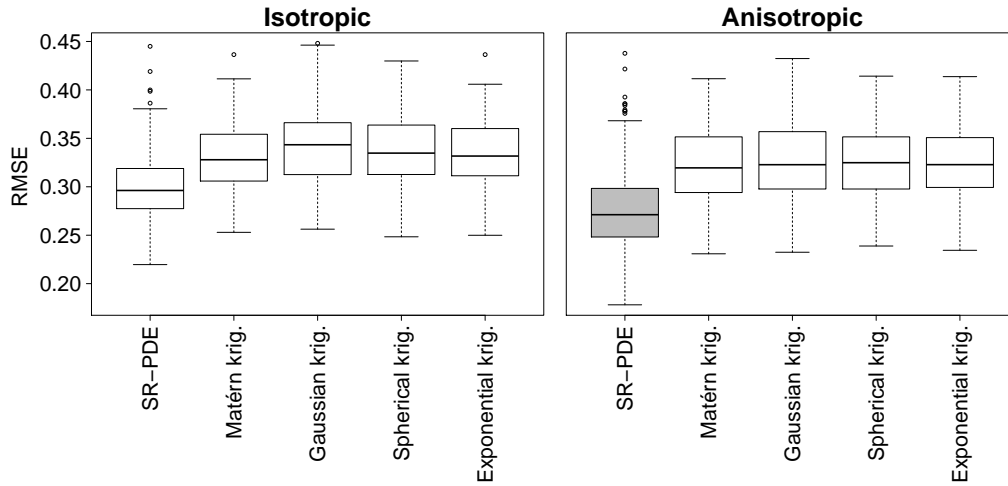
Fig. 9: Second simulation study. Boxplots of the RMSE of the spatial field estimates obtained by isotropic and anisotropic SR-PDE and by isotropic and anisotropic universal kriging with different variogram models (Matérn, Gaussian, Spherical and Exponential).

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 10 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 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 10 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. In contrast to 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 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 basis of the RMSE of the spatial field estimates, over the 200 simulation replicates. The RMSE is computed on a fine regular squared grid of step 0.05 spanning the whole domain. The boxplots of the RMSE displayed in Figure 11, and the Wilcoxon pairwise tests confirm that the RMSE of the estimates obtained with anisotropic SR-PDE are significantly smaller than those of isotropic SR-PDE, and of isotropic and anisotropic kriging with any variogram model.
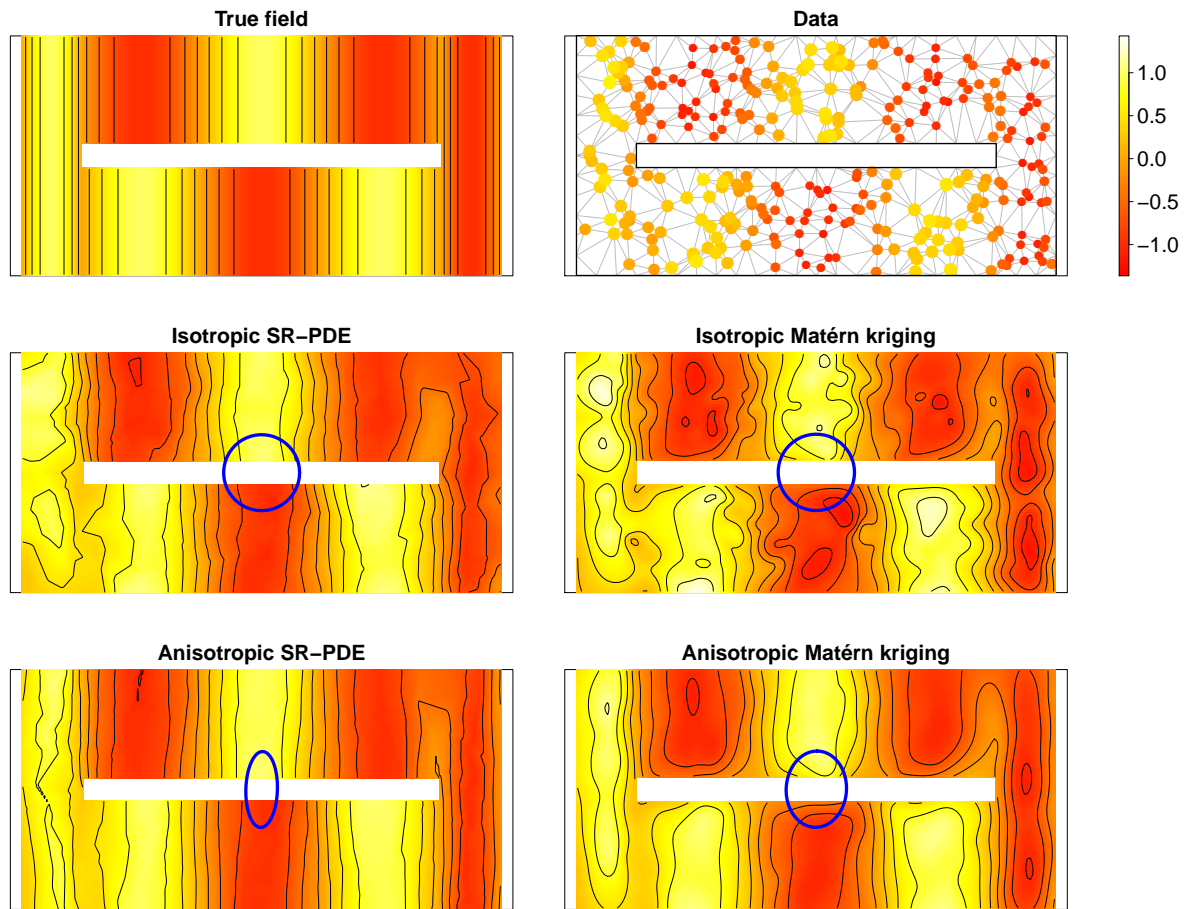
Fig. 10: Third simulation study. First row: true field and sampled data, with the triangulation used for SR-PDE estimates in gray. Second row: field estimates provided by isotropic SR-PDE and isotropic Matérn kriging. Third row: field estimates provided by anisotropic SR-PDE and anisotropic Matérn kriging.
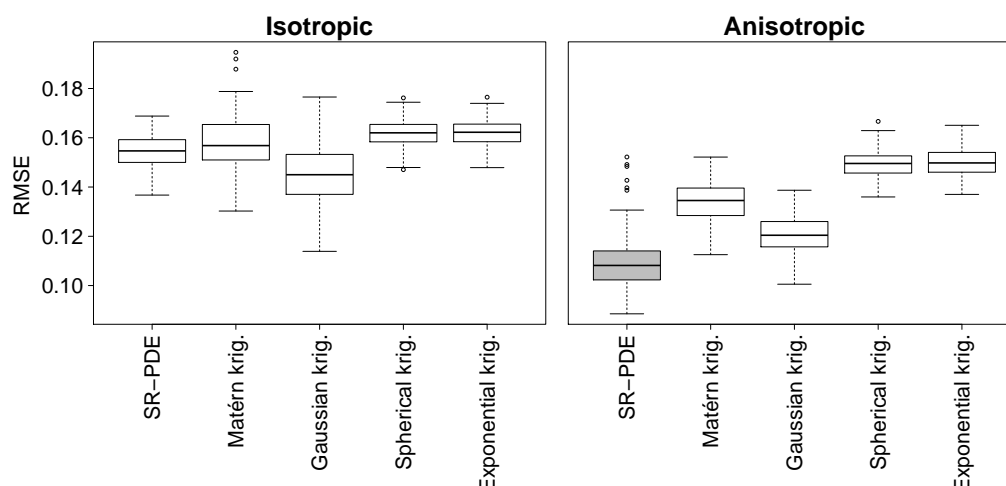
Fig. 11: Third simulation study. Boxplots of the RMSE over the 200 simulation replicates of the estimates obtained by isotropic and anisotropic SR-PDE and by isotropic and anisotropic kriging with different variogram models (Matérn, Gaussian, Spherical and Exponential).

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 recorded in Switzerland on May 8, 1986; this dataset was used for the Spatial Interpolation Comparison 97 (Dubois et al., 2003). The data are shown in Figure 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 12 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; the approximate 95% confidence interval for β (see, e.g., Sangalli et al., 2013; Azzimonti et al., 2015) includes 0. 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 complex 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.

The left panel of Figure 13 displays the selected anisotropy matrix, superimposed to the data, and the right panel of the same figure shows the corresponding field estimate. The proposed method correctly identifies the anisotropy and provides a smooth field that is able to capture the important features of the data, with the well defined elongated regions of homogeneous values.
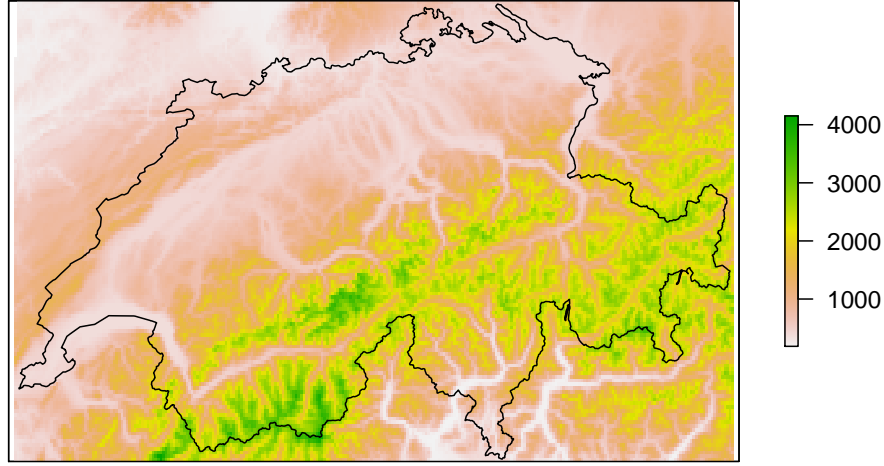
Fig. 12: Elevation in the Switzerland region. This plot is obtained with the R package `geostatsp` (Brown, 2015).
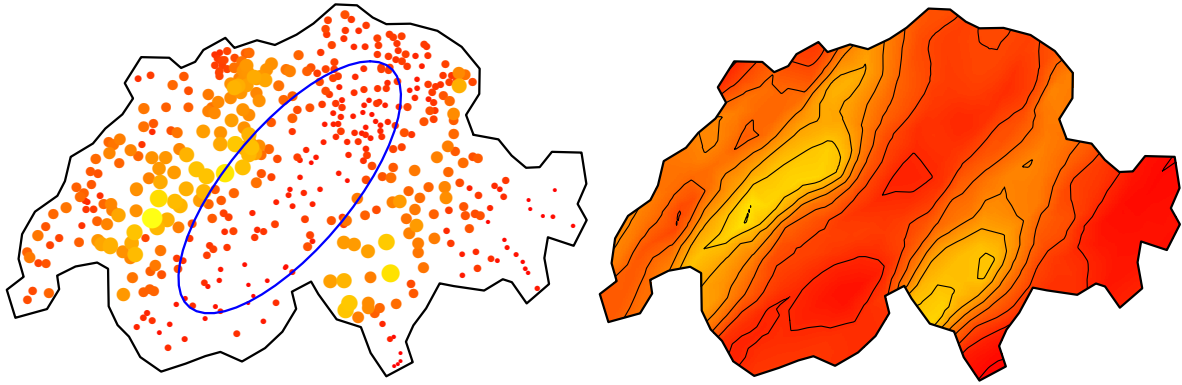

Fig. 13: Left panel: ellipse representing the anisotropy matrix selected for the Switzerland rainfall data. Right panel: field estimate obtained by the proposed method.

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. (2017), 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. 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:

$$Lf = -\nabla \cdot \{K(\mathbf{p}, \boldsymbol{\theta})\nabla f\} + \mathbf{b}(\mathbf{p}, \boldsymbol{\eta}) \cdot \nabla f + c(\mathbf{p}, \boldsymbol{\zeta})f. \tag{7}$$

In Azzimonti et al. (2015) the parameters $\boldsymbol{\theta}$, $\boldsymbol{\eta}$ and $\boldsymbol{\zeta}$ in the diffusion, transport and reaction terms in (7) are fixed on the basis of their physical meaning, thanks to the prior knowledge of the problem under study. In many instances, however, the prior knowledge of the phenomenon is not as detailed and does not indicate the specific values for these parameters. In such contexts, these parameters could be derived from data, generalizing the methodology presented in this paper. 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 such a flexible modeling of anisotropic and non-stationary spatial variation. See also Fuglstad et al. (2015), that extend the approach based on Gaussian random fields and stochastic PDEs introduced in Lindgren et al. (2011). 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.

A Estimation of f for a given anisotropy matrix K and discretization via finite elements

As shown in Azzimonti et al. (2014), by introducing an auxiliary variable g, we can write the fourth-order problem (4) as a coupled system of second order problems:

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

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

where B_c are general boundary conditions, and B_c^* are the boundary conditions associated with the adjoint problem, i.e.,

$$\begin{cases} g = 0 & \text{on } \Gamma_D, \\ K\nabla g \cdot \mathbf{v} = 0 & \text{on } \Gamma_N, \\ K\nabla g \cdot \mathbf{v} + \chi g = 0 & \text{on } \Gamma_R. \end{cases}$$

See Azzimonti et al. (2014) for details. For homogeneous boundary conditions, 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

$$(1-\rho)\frac{1}{n}\mathbf{h}_n^\top \hat{\mathbf{f}}_n - \rho \frac{1}{|\Omega|}\int_\Omega (\nabla g \cdot K\nabla h) = (1-\rho)\frac{1}{n}\mathbf{h}_n^\top \mathbf{z}, \quad \int_\Omega gv + \int_\Omega (\nabla v \cdot K\nabla \hat{f}) = 0 \tag{8}$$

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 2.3. This allow 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 (8):

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

This leads to Proposition 2.

References

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

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

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

P.E. Brown, Model-based geostatistics the easy way, J. Stat. Softw. 63 (2015) 1–24.

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

J. Cao, J.O. Ramsay, Linear mixed-effects modeling by parameter cascading, J. Amer. Statist. Assoc. 105 (2010) 365–374.

J. Cao, J.O. Ramsay, Parameter cascades and profiling in functional data analysis, Comput. Statist. 22 (2007) 335–351.

J. Cao, J.O. Ramsay, Generalized profiling estimation for global and adaptive penalized spline smoothing, Comput. Statist. Data Anal. 53 (2009) 2550–2562.

G. Dubois, J. Malczewski, M. De Cort, Mapping Radioactivity in the Environment: Spatial Interpolation Comparison 97, Office for Official Publications of the European Communities, Luxembourg, 2003.

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

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

M.S. Gockenbach, Understanding and Implementing the Finite Element Method, Society for Industrial and Applied Mathematics, Philadelphia, PA, 2006.

C. Gu, Smoothing Spline ANOVA Models, Springer, New York, 2013.

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

E. Lila, L.M. Sangalli, J.O. Ramsay, L. Formaggia, 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 0.1-4, 2016.

F. Lindgren, H. Rue, J. Lindström, An explicit link between Gaussian fields and Gaussian Markov random fields: The stochastic partial differential equation approach, J. R. Stat. Soc. Ser. B (Methodol.) 73 (2011) 423–498.

D. Nychka, R. Furrer, J. Paige, S. Sain, fields: Tools for spatial data, R Package Version 8.4-1, 2015.

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

A. Quarteroni, Numerical Models for Differential Problems, vol. 2, Springer, New York, 2010.

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

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

P.J. Ribeiro Jr, P.J. Diggle, geoR: Analysis of Geostatistical Data, R Package Version 1.7-5.2, 2016.

W. Rudin, Functional Analysis, 2nd ed., McGraw-Hill, New York, 1991.

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

M. Schlather, A. Malinowski, M. Oesting, D. Boecker, K. Strokorb, S. Engelke, J. Martini, F. Ballani, O. Moreva, P.J. Menck, S. Gross, U. Ober, C. Berreth, K. Burmeister, J. Manitz, O. Morena, P. Ribeiro, R. Singleton, B. Pfaff, R Core Team, RandomFields: Simulation and Analysis of Random Fields, R Package Version 3.1.16, 2016

G. Wahba, Spline Models for Observational Data, Society for Industrial and Applied Mathematics, Philadelphia, PA, 1990.

S.N. Wood, Modelling and smoothing parameter estimation with multiple quadratic penalties, J. R. Stat. Soc. Ser. B (Methodol.) 62 (2000) 413–428.

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

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