



MOX-Report No. 79/2021

**Some first inferential tools for spatial regression with  
differential regularization**

Ferraccioli, F.; Sangalli, L.M.; Finos, L.

MOX, Dipartimento di Matematica  
Politecnico di Milano, Via Bonardi 9 - 20133 Milano (Italy)

[mox-dmat@polimi.it](mailto:mox-dmat@polimi.it)

<http://mox.polimi.it>

# Some first inferential tools for spatial regression with differential regularization

Federico Ferraccioli<sup>a</sup>, Laura M. Sangalli<sup>b,\*</sup>, Livio Finos<sup>c</sup>

<sup>a</sup>*Dipartimento di Scienze Statistiche, Università di Padova, Italia*

<sup>b</sup>*MOX - Dipartimento di Matematica, Politecnico di Milano, Italia*

<sup>c</sup>*Dipartimento di psicologia dello sviluppo e della socializzazione, Università of Padova, Italia*

---

## Abstract

Spatial regression with differential regularization is an innovative approach at the crossroad between functional data analysis and spatial data analysis. These models have been shown to be numerically efficient and capable to handle complex applied problems. On the other hand, their theoretical properties are still largely unexplored. Here we consider the discrete estimators in spatial regression models with differential regularization, obtained after numerical discretization, using an expansion on a finite element basis. We study the consistency and the asymptotic normality of these discrete estimators. We also propose a nonparametric test procedure for the linear part of the models, based on random sign-flipping of the score components. The test exploits an appropriate decomposition of the smoothing matrix, in order to reduce the effect of the spatial dependence, without any parametric assumption on the form of the correlation structure. The proposed test is shown to be superior to parametric alternatives.

*Keywords:* functional data analysis, roughness penalty, semiparametric regression, smoothing.

*2020 MSC:* Primary 62G09, Secondary 62G05

---

## 1. Introduction

This work presents efficient tests for the linear component of semiparametric penalized spatial regression models. The proposed procedure focuses on the semiparametric models introduced in [38], named Spatial Regression with Partial Differential Equation regularization (SR-PDE). These models are at the crossroad of spatial data analysis [8–10] and functional data analysis [13, 26, 34]. This area, often referred to as spatial functional data analysis, is receiving an increasing interest, as witnessed by the special issues [1, 16, 30] and the recent collective volume [31]. Moreover, semiparametric penalized regression and partially linear models [18, 35] are extensively used in functional data analysis, and inference for such models has been addressed by a vast literature; see, e.g., the review in [36].

SR-PDE embraces a set of numerically efficient regression models, capable to handle complex applied problems [see, e.g., 5, 12, 27, 37]. The basic formulation of SR-PDE involves a semiparametric (or partially linear) model of the form  $z_i = \mathbf{w}_i^\top \boldsymbol{\beta} + f(\mathbf{p}_i) + \epsilon_i$ ,  $i \in \{1, \dots, n\}$ , where the response variable  $z_i$  observed at location  $\mathbf{p}_i$  in the spatial domain  $\Omega$  is explained by a regression on the covariates  $\mathbf{w}_i$ , with coefficient vector  $\boldsymbol{\beta}$ , and on a non-linear smooth function  $f$ , defined on the domain  $\Omega$ . The coefficient vector  $\boldsymbol{\beta}$  (in the linear or parametric part of the model) and the function  $f$  (in the non-linear or nonparametric part of the model) are estimated minimizing a penalized least square functional, similar to other semiparametric regression models, such as those based on penalized splines regression [see, e.g., 19, 48], on thin plate splines [see, e.g., 46] and on soap film smoothing [47]. SR-PDE can handle data scattered over complicated two-dimensional domains  $\Omega$ , and can comply with different conditions at the domain boundary [5, 38]. Moreover, the regularizing term in SR-PDE can involve general forms of Partial Differential Equations (PDE), modeling various types of anisotropy and non-stationarity. These peculiarities pose new challenges for the study of the theoretical properties of these models, that are still largely unexplored. References [5, 38] demonstrate the well posedness of the estimation problem and describe the characterization of its solution. In particular, SR-PDE estimation functional cannot be solved analytically; the solution is approximated via an expansion in finite element bases. The

---

\*Corresponding author. Email address: [laura.sangalli\(at\)polimi.it](mailto:laura.sangalli(at)polimi.it)

work [4] shows that the estimator of  $f$  is asymptotically unbiased. Moreover, [3] investigates its consistency, under some simplifying hypothesis. However, these first attempts to study the asymptotic properties of the methods are restricted to a model where no covariates are present, but only the nonparametric term  $f$  is considered.

In this work, we instead consider the general semiparametric (or partially linear) model in [38], that also includes the covariates, focusing directly on the discrete estimators obtained after numerical discretization of the estimation problem. In particular, we derive the asymptotic distribution of the estimators of  $\boldsymbol{\beta}$  and  $f$ , obtained after expansion on a finite element basis. Moreover, we demonstrate the consistency of these estimators. We hence focus on hypothesis testing on  $\boldsymbol{\beta}$ .

A standard way to approach the problem is to consider the asymptotic distribution of the estimator. This naturally leads to Wald-type test statistics, either using the asymptotic variance, or with more robust approaches such as sandwich estimators [17, 48]. The sandwich estimate corrects for the potentially misspecified variance and induces an asymptotically exact test under mild assumptions. These approaches might nonetheless have poor performances in the finite sample scenario, due to the dependence on the regularizing term [see, e.g. 14, 29]. A possible solution to overcome the problem is to resort to nonparametric tests, such as permutation tests [see, e.g., 20, 33, 45]. These methods require fewer parametric assumptions and are often more powerful since they avoid imposing restrictive assumptions on the data [22, 43]. Rather than permutations, sometimes other transformations are used, such as rotations [41] and sign-flipping of residuals [45].

Here in particular we propose a sign-flip score test for  $\boldsymbol{\beta}$ . The test is influenced by a sign-flip score strategy recently proposed in [21] for classical generalized linear models. The procedure in [21], although robust against various types of misspecification, relies on the asymptotic independence of the residuals. Unfortunately, in a spatial regression setting, the dependence among the residuals can instead be very strong, due to the presence of the nonparametric part in the model and of the regularizing term in the estimation functional. For this reason, in addition to proposing a sign-flip score test for the considered SR-PDE models, we also devise a variant of this test that avoids the issues generated by spatial correlation, exploiting a spectral decomposition of the smoothing matrix. This idea is in turn inspired by approaches used for instance in [24, 25] in the context of linear regression models. This leads to the definition of an asymptotically exact test that preserves the finite sample invariance of the covariance structure of the test under random sign-flips.

The proposed resampling strategy offers a new inference approach that could be extended to other spatial semiparametric methods [e.g., 46, 47], as well as to semiparametric regression models extensively used in functional data analysis, such as those based on splines [e.g., 19, 42, 48].

The paper is organized as follows. In Section 2 we briefly review the SR-PDE framework and outline how the discrete estimators are obtained. In Section 3 we study the asymptotic properties of these estimators. In Section 4 we present the proposed nonparametric hypothesis testing procedures. In Section 5 we compare such proposals to more classical parametric approaches through simulation studies. In Section 6 we present an application to the study of rainfall measurements in Switzerland. Some final discussions and directions for future research are outlined in Section 7.

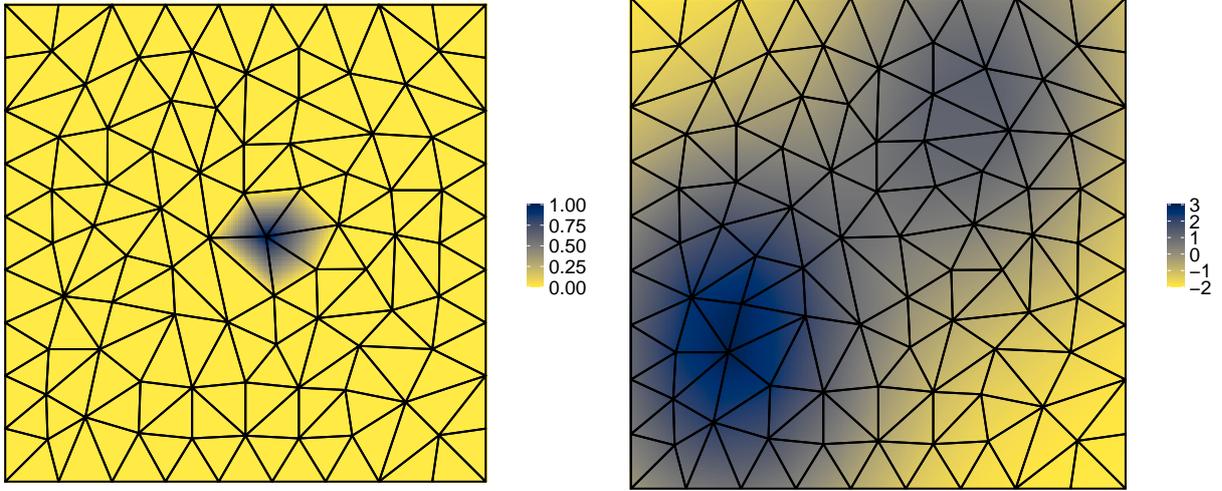
## 2. Model

In this section we briefly review the basic setting of SR-PDE. Let  $\mathbf{p}_i = (p_{i1}, p_{i2}) \in \Omega$ ,  $i \in \{1, \dots, n\}$ , be  $n$  data locations over a bounded domain  $\Omega \subset \mathbb{R}^2$  with a boundary  $\partial\Omega$  that is continuous and has continuous first and second derivatives. Let  $z_i \in \mathbb{R}$  be the value of the variable of interest observed at point  $\mathbf{p}_i$  and let  $\mathbf{w}_i \in \mathbb{R}^q$  be a vector of covariates associated with the observation  $z_i$ . SR-PDE assumes a semi-parametric model of the form

$$z_i = \mathbf{w}_i^\top \boldsymbol{\beta} + f(\mathbf{p}_i) + \epsilon_i, \quad \text{for } i \in \{1, \dots, n\}, \quad (1)$$

where  $\boldsymbol{\beta} \in \mathbb{R}^q$  is the vector of true parameters,  $f$  is an unknown deterministic mean spatial field that represents the spatial dependence structure of the problem under study, and  $\epsilon_1, \dots, \epsilon_n$  are i.i.d. random errors with zero mean and variance  $\sigma^2$ . The estimation problem can be solved minimizing the regularized least squares

$$\frac{1}{n} \sum_{i=1}^n \{z_i - \mathbf{w}_i^\top \boldsymbol{\beta} - f(\mathbf{p}_i)\}^2 + \lambda \int_{\Omega} (\Delta f)^2 \, d\mathbf{p}, \quad (2)$$



**Fig. 1:** Left: linear finite element basis function on a triangulation of a squared domain. Right: an example of linear finite element function on the considered triangulation.

where  $\lambda > 0$  is a smoothing parameter and  $\Delta$  denotes the Laplace operator, defined as

$$\Delta f(\mathbf{p}) = \frac{\partial^2}{\partial p_1^2} f(\mathbf{p}) + \frac{\partial^2}{\partial p_2^2} f(\mathbf{p}), \quad (3)$$

where  $\mathbf{p} = (p_1, p_2) \in \Omega$ . The higher the value of  $\lambda$ , the more regular is the estimate of  $f$ , and viceversa. References [4, 38] show that the estimation functional is well posed for  $\boldsymbol{\beta} \in \mathbb{R}^q$  and  $f$  in an appropriate space of functions. Moreover, the estimation problem has a unique solution once imposing appropriate boundary conditions on  $f$ ; in particular, in the following we shall assume that the normal derivative of the spatial field  $f$  at the domain boundary  $\partial\Omega$  is zero. The regularizing term in (2), instead of the simple Laplacian, may involve a PDE that formalizes the available problem-specific information, enabling a very rich modeling of space variation; see [4, 5] for details.

### 2.1. Finite elements

The solution to the minimization problem (2) cannot be found analytically. It is therefore necessary to use numerical discretization procedures. One possibility in this respect is to use the finite element method, as shown, e.g., in [37, 38].

To this end, we discretize the domain  $\Omega$  using a constrained Delaunay triangulation; this is a generalization of the Delaunay triangulation [e.g., 23] for bounded domains. Fig. 1 shows for instance a triangulation of a squared domain. We denote by  $\mathcal{T}$  the triangulation, by  $\Omega_{\mathcal{T}}$  the resulting discretized domain, union of all the triangles in  $\mathcal{T}$ , and by  $\boldsymbol{\xi}_k \in \Omega_{\mathcal{T}}, k \in \{1, \dots, N_{\mathcal{T}}\}$ , the vertices (or nodes) of the triangulation. We hence define a set of  $N_{\mathcal{T}}$  basis functions, that span the space of continuous functions over  $\Omega_{\mathcal{T}}$  that are linear once restricted to any triangle of  $\mathcal{T}$ ; this is the linear finite element space associated with the triangulation  $\mathcal{T}$  (higher polynomial orders can as well be considered [see, e.g., 37, 38]). In particular, we define a basis  $\psi_k$  for each node  $\boldsymbol{\xi}_k$ : this finite element basis is defined as the piecewise linear function on  $\Omega_{\mathcal{T}}$  that has value 1 at node  $\boldsymbol{\xi}_k$  and value 0 at any other node  $\boldsymbol{\xi}_\ell$  with  $\ell \neq k$ . As highlighted in the left panel of Fig. 1, such bases have a local support. Through expansions on these bases, it is possible to represent any globally continuous and piecewise linear function  $f$  on  $\Omega_{\mathcal{T}}$ , as  $f(\cdot) = \mathbf{f}^T \boldsymbol{\psi}(\cdot)$ , where  $\boldsymbol{\psi} := (\psi_1, \dots, \psi_{N_{\mathcal{T}}})^T$  is the vector of  $N_{\mathcal{T}}$  finite element bases and  $\mathbf{f} = (f_1, \dots, f_{N_{\mathcal{T}}})^T$  is the vector of coefficients of the basis expansion; see the example in the right panel of Fig. 1. Thanks to the definition of the finite element bases, the vector  $\mathbf{f}$  coincides with the vector of evaluations of the function  $f$  at the  $N_{\mathcal{T}}$  mesh nodes:  $\mathbf{f} = (f(\boldsymbol{\xi}_1), \dots, f(\boldsymbol{\xi}_{N_{\mathcal{T}}}))^T$ . The finite element space, spanned by  $\psi_1, \dots, \psi_{N_{\mathcal{T}}}$ , is hence constructed so that any function in this space is defined by its values at the  $N_{\mathcal{T}}$  mesh nodes.

## 2.2. Discrete estimator

Let  $\Psi$  be the  $n \times N_{\mathcal{T}}$  matrix evaluating the  $N_{\mathcal{T}}$  basis functions  $\psi_1, \dots, \psi_{N_{\mathcal{T}}}$  at the  $n$  data locations

$$\Psi = \begin{bmatrix} \psi_1(\mathbf{p}_1) & \dots & \psi_{N_{\mathcal{T}}}(\mathbf{p}_1) \\ \vdots & \ddots & \vdots \\ \psi_1(\mathbf{p}_n) & \dots & \psi_{N_{\mathcal{T}}}(\mathbf{p}_n) \end{bmatrix}$$

and  $\mathbf{R}_0, \mathbf{R}_1$  be the  $N_{\mathcal{T}} \times N_{\mathcal{T}}$  matrices

$$\mathbf{R}_0 = \int_{\Omega_{\mathcal{T}}} \psi \psi^{\top}, \quad \mathbf{R}_1 = \int_{\Omega_{\mathcal{T}}} \nabla \psi \nabla \psi^{\top}.$$

Let  $\mathbf{z} = (z_1, \dots, z_n)^{\top}$  be the vector of observed data values. Let also  $\mathbf{W}$  be the  $n \times q$  matrix whose  $i$ -th row is given by  $\mathbf{w}_i^{\top}$ , the vector of  $q$  covariates associated with observation  $z_i$  at  $\mathbf{p}_i$ , and assume that  $\mathbf{W}$  has full rank. Moreover, set  $\mathbf{Q} = \mathbf{I} - \mathbf{W}(\mathbf{W}^{\top} \mathbf{W})^{-1} \mathbf{W}^{\top}$ , the matrix that projects into the orthogonal complement of  $\mathbb{R}^n$  with respect to the subspace of  $\mathbb{R}^n$  spanned by the columns of  $\mathbf{W}$ . Finally, for any function  $f$  in the finite element space, we denote by  $\mathbf{f}$  the vector of evaluations of the function  $f$  at the  $N_{\mathcal{T}}$  nodes of the mesh,  $\mathbf{f} = (f(\xi_1), \dots, f(\xi_{N_{\mathcal{T}}}))^{\top}$ . Having defined these quantities, the problem (2) is recast in the finite element space. To this end, as detailed in [38], we introduce an auxiliary function  $g = \mathbf{g}^{\top} \psi$  that represents  $\Delta f$  (or more generally the misfit of the penalized PDE), leading to the following result.

**Proposition 1** ([38]). *There exists a unique pair of estimators  $(\hat{\beta}, \hat{f})$  which solve the discrete counterpart of the estimation problem. Moreover*

$$\hat{\beta} = (\mathbf{W}^{\top} \mathbf{W})^{-1} \mathbf{W}^{\top} (\mathbf{z} - \Psi \hat{\mathbf{f}}),$$

and  $\hat{f} = \hat{\mathbf{f}}^{\top} \psi$  with  $\hat{\mathbf{f}}$  satisfying

$$\begin{bmatrix} -\Psi^{\top} \mathbf{Q} \Psi / n & \lambda \mathbf{R}_1^{\top} \\ \lambda \mathbf{R}_1 & \lambda \mathbf{R}_0 \end{bmatrix} \begin{bmatrix} \hat{\mathbf{f}} \\ \mathbf{g} \end{bmatrix} = \begin{bmatrix} -\Psi^{\top} \mathbf{Q} \mathbf{z} / n \\ \mathbf{0} \end{bmatrix}. \quad (4)$$

Denote by  $\mathbf{P} = \mathbf{R}_1^{\top} \mathbf{R}_0^{-1} \mathbf{R}_1$  the  $N_{\mathcal{T}} \times N_{\mathcal{T}}$  symmetric positive definite matrix that discretises the penalizing term in (2). Moreover, let  $\mathbf{S}$  be the  $N_{\mathcal{T}} \times n$  matrix

$$\mathbf{S} = (\Psi^{\top} \mathbf{Q} \Psi / n + \lambda \mathbf{P})^{-1} \Psi^{\top} \mathbf{Q} / n.$$

With this notation, we can derive the analytic expressions for the estimators  $\hat{\mathbf{f}}$  and  $\hat{\beta}$  as

$$\hat{\mathbf{f}} = n^{-1} (\Psi^{\top} \mathbf{Q} \Psi / n + \lambda \mathbf{P})^{-1} \Psi^{\top} \mathbf{Q} \mathbf{z} = \mathbf{S} \mathbf{z}, \quad (5)$$

$$\hat{\beta} = (\mathbf{W}^{\top} \mathbf{W})^{-1} \mathbf{W}^{\top} (\mathbf{I} - \mathbf{S}) \mathbf{z}. \quad (6)$$

## 3. Asymptotic properties

In this section we study the asymptotic properties of the estimators in (5) and (6), proving their asymptotic normality and consistency. The following results are derived implicitly conditioning on the locations  $\{\mathbf{p}_i, i = 1, \dots, n\}$  and on the covariates  $\mathbf{W}$ . This provides the most natural setting for the introduction of the proposed tests in Section 4.2, which are conditioned to the same quantities. Analogous results could be obtained under an unconditioned framework, by imposing appropriate distributional assumptions on the covariates and locations. In particular, the mesh and locations should satisfy Assumption 1 below.

We introduce the  $q \times q$  matrix

$$\Sigma_n = \mathbf{W}^{\top} \mathbf{W} / n,$$

and the  $N_{\mathcal{T}} \times N_{\mathcal{T}}$  symmetric matrix

$$\mathbf{A}_n = (\Psi^{\top} \mathbf{Q} \Psi / n)^{-1},$$

where the subscript  $n$  underlines the dependency on the sample size. We assume that the number of basis  $N_{\mathcal{T}}$  and the triangulation  $\mathcal{T}$  are fixed, and that the triangulation is fine enough to capture all the aspects of the spatial structure of the problem. We moreover make the following assumption:

**Assumption 1.** Either the nodes of the triangulation are a subset of the data locations, i.e.,  $\{\xi_1, \dots, \xi_{N_T}\} \subset \{\mathbf{p}_1, \dots, \mathbf{p}_n\}$ , or, for  $n$  large enough, there are at least  $q + 1$  observations in the support of each basis function  $\psi_1, \dots, \psi_{N_T}$ .

Assumption 1 gives a sufficient condition for the non-singularity of the matrix  $\mathbf{A}_n$ . In fact, a sufficient condition for  $\Psi$  to be full rank is that  $n \geq N_T$ , with at least one observation in the support of each basis function. The matrix  $\mathbf{Q}$  has rank  $n - q$ . This means that, in the product  $\Psi^T \mathbf{Q} \Psi$ , the matrix  $\mathbf{Q}$  has the effect of annihilating  $q$  rows of  $\Psi$ . If the mesh nodes correspond to a subset of the data locations, it is enough to have  $n \geq q + N_T$  for  $\Psi^T \mathbf{Q} \Psi$  to be full rank. If the mesh nodes do not coincide with a subset of the data locations, a sufficient condition for  $\Psi^T \mathbf{Q} \Psi$  to be full rank is to have at least  $n \geq (q + 1)N_T$  observations, with at least  $q + 1$  observations in the support of each basis.

Under Assumption 1, we can now study the asymptotic behavior of the nonparametric component of SR-PDE model.

**Theorem 1.** Let  $(\hat{\mathbf{f}}_n)$  be a sequence of SR-PDE estimators (5). Assume that a nonsingular limit  $\mathbf{A} = \lim_n \mathbf{A}_n$  exists. If  $\lambda = \lambda_n = o(n^{-1/2})$ , then  $\hat{\mathbf{f}}_n$  has asymptotic distribution

$$\sqrt{n}(\hat{\mathbf{f}}_n - \mathbf{f}) \stackrel{d}{\sim} N_{N_T}(\mathbf{0}, \sigma^2 \mathbf{A}).$$

Moreover,  $\hat{\mathbf{f}}_n$  is a consistent estimator for  $\mathbf{f}$ , with  $\hat{\mathbf{f}}_n \xrightarrow{p} \mathbf{f}$ , where  $\xrightarrow{p}$  denotes convergence in probability.

**Proof.** To obtain the asymptotic distribution of  $\sqrt{n}(\hat{\mathbf{f}}_n - \mathbf{f})$  recall from (4) that  $\hat{\mathbf{f}}_n$  is the solution of the linear system

$$\begin{cases} -\Psi^T \mathbf{Q} \Psi \hat{\mathbf{f}}_n / n + \lambda \mathbf{R}_1^T \mathbf{g} = -\Psi^T \mathbf{Q} \mathbf{z} / n, \\ \lambda \mathbf{R}_1 \hat{\mathbf{f}}_n + \lambda \mathbf{R}_0 \mathbf{g} = \mathbf{0}. \end{cases} \quad (7)$$

Under the model assumption (1), the first equation in (7) can be rewritten as

$$-\Psi^T \mathbf{Q} \Psi \hat{\mathbf{f}}_n / n + \lambda \mathbf{R}_1^T \mathbf{g} = -\Psi^T \mathbf{Q} (\mathbf{W} \beta + \Psi \mathbf{f} + \boldsymbol{\epsilon}) / n, \quad (8)$$

where  $\boldsymbol{\epsilon} = (\epsilon_1, \dots, \epsilon_n)^T$  is the vector of i.i.d. errors. Substituting in (8) the expression for  $\mathbf{g}$  obtained from the second equation in (7), and noting that  $\mathbf{Q} \mathbf{W}$  has all entries equal to 0, we obtain

$$(\Psi^T \mathbf{Q} \Psi / n + \lambda \mathbf{P})(\hat{\mathbf{f}}_n - \mathbf{f}) + \lambda \mathbf{P} \mathbf{f} = \Psi^T \mathbf{Q} \boldsymbol{\epsilon} / n. \quad (9)$$

From (9) we can explicitly obtain  $(\hat{\mathbf{f}}_n - \mathbf{f})$ ,

$$\hat{\mathbf{f}}_n - \mathbf{f} = -\lambda (\Psi^T \mathbf{Q} \Psi / n + \lambda \mathbf{P})^{-1} \mathbf{P} \mathbf{f} + \frac{1}{n} (\Psi^T \mathbf{Q} \Psi / n + \lambda \mathbf{P})^{-1} \Psi^T \mathbf{Q} \boldsymbol{\epsilon}. \quad (10)$$

Note that the first term in (10) is not random while the second term is a weighted average of i.i.d. errors. The asymptotic normality of the estimator thus follow from the central limit theorem.

We now study separately the bias and the variance of  $\hat{\mathbf{f}}_n$ . Recall from (5) that  $\hat{\mathbf{f}}_n$  takes the form

$$\hat{\mathbf{f}}_n = n^{-1} (\Psi^T \mathbf{Q} \Psi / n + \lambda \mathbf{P})^{-1} \Psi^T \mathbf{Q} \mathbf{z}.$$

For the bias of  $\hat{\mathbf{f}}_n$  we have:

$$\begin{aligned} b_n(\lambda) &= E(\hat{\mathbf{f}}_n) - \mathbf{f} = E\{n^{-1} (\Psi^T \mathbf{Q} \Psi / n + \lambda \mathbf{P})^{-1} \Psi^T \mathbf{Q} \mathbf{z}\} - \mathbf{f} = E\{n^{-1} (\Psi^T \mathbf{Q} \Psi / n + \lambda \mathbf{P})^{-1} \Psi^T \mathbf{Q} (\Psi \mathbf{f} + \mathbf{W} \beta + \boldsymbol{\epsilon})\} - \mathbf{f} \\ &= (\Psi^T \mathbf{Q} \Psi / n + \lambda \mathbf{P})^{-1} (\Psi^T \mathbf{Q} \Psi / n) \mathbf{f} - \mathbf{f} = \{(\mathbf{A}_n^{-1} + \lambda \mathbf{P})^{-1} \mathbf{A}_n^{-1} - \mathbf{I}\} \mathbf{f} \end{aligned}$$

where we use the fact that by construction  $\mathbf{Q} \mathbf{W}$  has all entries equal to 0. Using Taylor expansion of  $b_n(\lambda)$  as a function of  $\lambda$  we obtain the bias expression

$$b_n(\lambda) = -\lambda \mathbf{A}_n \mathbf{P} \mathbf{f} + \lambda^2 \mathbf{A}_n \mathbf{P} \mathbf{A}_n \mathbf{P} \mathbf{f} + o(\lambda^2).$$

For the variance, we have

$$\begin{aligned}\text{Var}_n(\lambda) &= \text{Var}(\hat{\mathbf{f}}_n) = \text{Var}\{(\mathbf{\Psi}^\top \mathbf{Q} \mathbf{\Psi}/n + \lambda \mathbf{P})^{-1} \mathbf{\Psi}^\top \mathbf{Q}/n (\mathbf{\Psi} \mathbf{f} + \mathbf{W} \boldsymbol{\beta} + \boldsymbol{\epsilon})\} \\ &= \frac{\sigma^2}{n} (\mathbf{\Psi}^\top \mathbf{Q} \mathbf{\Psi}/n + \lambda \mathbf{P})^{-1} (\mathbf{\Psi}^\top \mathbf{Q} \mathbf{\Psi}/n) (\mathbf{\Psi}^\top \mathbf{Q} \mathbf{\Psi}/n + \lambda \mathbf{P})^{-1} = \frac{\sigma^2}{n} (\mathbf{A}_n^{-1} + \lambda \mathbf{P})^{-1} \mathbf{A}_n^{-1} (\mathbf{A}_n^{-1} + \lambda \mathbf{P})^{-1},\end{aligned}$$

where we used the fact that the matrix  $\mathbf{Q}$  is idempotent and that the variance of the error vector is  $\text{Var}(\boldsymbol{\epsilon}) = \sigma^2 \mathbf{I}_n$ , with  $\mathbf{I}_n$  the identity matrix. We can then obtain the expansion for the variance of the estimator as

$$\text{Var}_n(\lambda) = \frac{\sigma^2}{n} \{\mathbf{A}_n - 2\lambda \mathbf{A}_n \mathbf{P} \mathbf{A}_n + 3\lambda \mathbf{A}_n \mathbf{P} \mathbf{A}_n \mathbf{P} \mathbf{A}_n + O(\lambda^3)\}.$$

The mean square error for the SR-PDE estimator can hence be computed as

$$\begin{aligned}MS E_n(\lambda) &= \text{Var}_n(\lambda) + b_n(\lambda) b_n(\lambda)^\top \\ &= \frac{\sigma^2}{n} \{\mathbf{A}_n - 2\lambda \mathbf{A}_n \mathbf{P} \mathbf{A}_n + 3\lambda \mathbf{A}_n \mathbf{P} \mathbf{A}_n \mathbf{P} \mathbf{A}_n + O(\lambda^3)\} + \lambda^2 \mathbf{A}_n \mathbf{P} \mathbf{f} \mathbf{f}^\top \mathbf{P} \mathbf{A}_n + o(\lambda^2).\end{aligned}\quad (11)$$

The bias of the estimator for the nonparametric component is of the order  $O(\lambda)$ . If  $\lambda = o(n^{-1/2})$ , the asymptotic bias of  $\sqrt{n}(\hat{\mathbf{f}}_n - \mathbf{f})$  vanishes. Otherwise, if  $\lambda \rightarrow 0$  slower than  $n^{-1/2}$ , the bias term dominates. From (11) it follows that  $E(\hat{\mathbf{f}}_n - \mathbf{f})^2 \rightarrow 0$  for  $\lambda = o(n^{-1/2})$ . Thus  $\hat{\mathbf{f}}_n$  converges to  $\mathbf{f}$  in probability and the estimator is consistent.  $\square$

**Remark 1.** In the special case where  $\lambda = \lambda_n = cn^{-1/2}$ , with  $c > 0$ , the first term of  $b_n(\lambda)$  does not disappear. Therefore, we obtain

$$\sqrt{n}(\hat{\mathbf{f}}_n - \mathbf{f}) \stackrel{n}{\sim} \mathcal{N}_{N_T}(-c \mathbf{A} \mathbf{P} \mathbf{f}, \sigma^2 \mathbf{A}).$$

The field estimator thus achieves the  $\sqrt{n}$  consistency also in this case, but with a more involved asymptotic mean.

Given (1), we now consider a similar result for the linear component of the model.

**Theorem 2.** Let  $(\hat{\boldsymbol{\beta}}_n)$  be a sequence of SR-PDE estimators (6). Let  $\Theta$  be a compact parameter space, with  $\boldsymbol{\beta}$  as interior point. Assume  $\boldsymbol{\Sigma} = \lim_n \boldsymbol{\Sigma}_n$  exists and is nonsingular. Then, given a consistent estimator  $\hat{\mathbf{f}}_n$ , the estimator  $\hat{\boldsymbol{\beta}}_n$  has asymptotic distribution

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) \stackrel{n}{\sim} \mathcal{N}_q(\mathbf{0}, \sigma^2 \{\boldsymbol{\Sigma}^{-1} + (1/n^2) \boldsymbol{\Sigma}^{-1} \mathbf{W}^\top \boldsymbol{\Psi} \mathbf{A} \boldsymbol{\Psi}^\top \mathbf{W} \boldsymbol{\Sigma}^{-1}\}).$$

Moreover,  $\hat{\boldsymbol{\beta}}_n$  is a consistent estimator for  $\boldsymbol{\beta}$ , with  $\hat{\boldsymbol{\beta}}_n \xrightarrow{p} \boldsymbol{\beta}$ .

**Proof.** Given a consistent estimator  $\hat{\mathbf{f}}_n$ , the vector  $\hat{\boldsymbol{\beta}}$  is the solution of the score equation

$$\frac{1}{n} \mathbf{W}^\top (\mathbf{z} - \mathbf{W} \hat{\boldsymbol{\beta}}_n - \boldsymbol{\Psi} \hat{\mathbf{f}}_n) = 0.$$

Using (1), we get

$$\frac{1}{n} \mathbf{W}^\top \mathbf{W} (\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) + \frac{1}{n} \mathbf{W}^\top \boldsymbol{\Psi} (\hat{\mathbf{f}}_n - \mathbf{f}) = \frac{1}{n} \mathbf{W}^\top \boldsymbol{\epsilon},$$

that can be rewritten as

$$\hat{\boldsymbol{\Sigma}}_n (\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) = \frac{1}{n} \mathbf{W}^\top \boldsymbol{\epsilon} - \frac{1}{n} \mathbf{W}^\top \boldsymbol{\Psi} (\hat{\mathbf{f}}_n - \mathbf{f}).\quad (12)$$

From (12) we can explicitly obtain  $(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta})$  as

$$\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta} = \frac{1}{n} \hat{\boldsymbol{\Sigma}}_n^{-1} \mathbf{W}^\top \boldsymbol{\epsilon} - \frac{1}{n} \hat{\boldsymbol{\Sigma}}_n^{-1} \mathbf{W}^\top \boldsymbol{\Psi} (\hat{\mathbf{f}}_n - \mathbf{f}).\quad (13)$$

Note that the first term on the right side of (13) is a weighted average of i.i.d. errors. Moreover, the term  $(\hat{\mathbf{f}}_n - \mathbf{f})$  is independent of  $\boldsymbol{\epsilon}$  by construction. Using the central limit theorem and Theorem 1, we thus obtain the asymptotic distribution of the estimator.

To obtain the consistency, we need  $E(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta})^2 \rightarrow 0$ . We thus need that both  $E(\boldsymbol{\epsilon}^\top \mathbf{W} \mathbf{W}^\top \boldsymbol{\epsilon})$  and  $E(\hat{\mathbf{f}}_n - \mathbf{f})$  go to zero. Convergence to zero of the first term is obtained via central limit theorem and convergence to zero of the second term follows from Theorem 1 for  $\lambda = o(n^{-1/2})$ . The right side of (13) thus converges to 0 in probability, yielding the consistency of  $\hat{\boldsymbol{\beta}}_n$ .  $\square$

Note that the estimation of the nonparametric component affects the variance of the estimator of the parametric component. Similar results are derived in [19] in the simpler case of univariate semiparametric models, and in [48] in the case of multivariate penalized spline regression.

#### 4. Hypothesis testing

The asymptotic results in the previous section lead to the natural question on how to define appropriate inference procedures. In the case of semiparametric regression, we might be interested to test whether the linear component has an effect on the variable of interest. We are thus interested in the system of hypotheses

$$H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0 \quad \text{versus} \quad H_1 : \boldsymbol{\beta} \neq \boldsymbol{\beta}_0.$$

The results in the previous section suggest the use of a Wald-type test [39]. However, since the asymptotic variance in Theorem 2 is not available, this is estimated by the empirical variance of the estimator, which is biased due to the presence of the regularization. This may result in poor control of the Type-I error and in general in an under-conservative behavior of the test, as highlighted by the simulation studies in Section 5. Because of this, some corrections to Wald type inference have been proposed in the literature on semiparametric and nonparametric penalized regression. For instance, sandwich estimators [e.g., 48] correct for the potentially misspecified variance and induce an asymptotically exact test under mild assumptions. In the finite sample scenario, though, the sandwich variant of the Wald test may nonetheless have poor performance, due to an overestimation of the variance, that leads to low power when a strong spatial structure in the covariates is present, as indicated by the simulations carried out in Section 5.

To avoid these issues, in Section 4.1 we propose a nonparametric alternative, designing a sign-flip score test for SR-PDE models. The proposed method uses the score function as test statistic, but does not rely on the estimation of the Fisher Information matrix to define the null distribution, which is implicitly recovered by random sign-flips of the contributions of the score. The test follows a similar strategy as the one discussed in [21] for generalized regression. However, in the standard generalized regression setting considered in [21], the nonparametric component and the associated penalizing term are not present, and the test can rely on the asymptotic independence of the residuals. Instead, in the context of SR-PDE, we must face the issues caused by the presence of the nonparametric term, with its spatial structure. In particular, the spatial regularization induces a strong dependence among the residuals, causing in turn a loss in the power of the sign-flip score test for SR-PDE models, as highlighted in the simulations carried out in Section 5. To solve this problem, in Section 4.2 we propose a modification of this nonparametric test, which relies on a spectral decomposition of the smoothing matrix. This approach leads to an asymptotically exact test, named Eigen sign-flip score test, which preserves the finite sample invariance of the covariance structure of the test under random sign-flips.

##### 4.1. Sign-flip score test for SR-PDE

The sign-flip score uses the score function  $T = \mathbf{W}^\top (\mathbf{z} - \mathbf{W}\boldsymbol{\beta} - \boldsymbol{\Psi}\mathbf{f})$  as test statistic. The observed score function, evaluated under the null hypothesis above, i.e.  $\boldsymbol{\beta} = \boldsymbol{\beta}_0$ , is therefore

$$T^{obs} = \mathbf{W}^\top (\mathbf{z} - \mathbf{W}\boldsymbol{\beta}_0 - \boldsymbol{\Psi}\hat{\mathbf{f}}_{H_0}),$$

where  $\hat{\mathbf{f}}_{H_0}$  is the spatial field in (5) estimated under the null hypothesis, i.e.,  $\hat{\mathbf{f}}_{H_0} = (\boldsymbol{\Psi}^\top \boldsymbol{\Psi} / n + \lambda \mathbf{P})^{-1} \boldsymbol{\Psi}^\top (\mathbf{z} - \mathbf{W}^\top \boldsymbol{\beta}_0) / n$ . Under the standard assumption of i.i.d. errors  $\epsilon_1, \dots, \epsilon_n$ , the score test can also be viewed as a sum of  $n$  contributions to the score that are 0-centered under the null hypothesis. In the sign-flip approach, this information is used to derive

the null distribution of the test statistic, that is, the set of test statistics computed under random sign-flips of the contributions of the score statistic. See also [45] for a review on other proposals based on sign-flipping.

More formally, let  $\pi = (\pi_1, \dots, \pi_n)^\top$  be a random vector uniformly distributed in  $\{-1, 1\}^n$ . Let also  $\mathbf{\Pi}$  be a diagonal matrix with entries  $\mathbf{\Pi}_{ii} = \pi_i$ . Given a matrix  $\mathbf{\Pi}$ , we define the sign-flip score statistic  $T_{\mathbf{\Pi}}$  as

$$T_{\mathbf{\Pi}} = \mathbf{W}^\top \mathbf{\Pi}(\mathbf{z} - \mathbf{W}\boldsymbol{\beta}_0 - \boldsymbol{\Psi}\hat{\mathbf{f}}_{H_0}).$$

Note that the observed statistic  $T^{obs}$  corresponds to the case where  $\pi_i = 1$ , for  $i = 1, \dots, n$ . As standard in permutational approaches, the  $p$ -value is thus computed as the rank of  $T^{obs}$  with respect to a sample of  $M$  sign-flips divided by  $M$  [32].

We first consider the behavior of both the expected value and the variance of  $T_{\mathbf{\Pi}}$  under the null hypothesis. In what follows we drop the subscript of  $\hat{\mathbf{f}}_{H_0}$  to ease the notation. Without loss of generality, we consider the case where  $\boldsymbol{\beta}_0 = 0$ .

**Proposition 2.** *For a given  $\mathbf{\Pi}$ , the expected value and the variance of the test statistic  $T$ , under the null hypothesis, are*

$$E(T_{\mathbf{\Pi}}) = \lambda \mathbf{W}^\top \mathbf{\Pi} \boldsymbol{\Psi} \mathbf{B}_n \mathbf{P} \mathbf{f}, \quad \text{Var}(T_{\mathbf{\Pi}}) = \sigma^2 \mathbf{W}^\top \left\{ \mathbf{I} + \frac{1}{n^2} \mathbf{\Pi} (\boldsymbol{\Psi} \mathbf{B}_n \boldsymbol{\Psi}^\top)^2 \mathbf{\Pi} \right\} \mathbf{W},$$

where  $\mathbf{B}_n = (\boldsymbol{\Psi}^\top \boldsymbol{\Psi} / n + \lambda \mathbf{P})^{-1}$ .

**Proof.** For the expected value, we have

$$E(T_{\mathbf{\Pi}}) = E\{\mathbf{W}^\top \mathbf{\Pi}(\mathbf{z} - \boldsymbol{\Psi}\hat{\mathbf{f}})\} = \mathbf{W}^\top \mathbf{\Pi} E(\boldsymbol{\Psi}\mathbf{f} - \boldsymbol{\Psi}\hat{\mathbf{f}} + \boldsymbol{\epsilon}).$$

Using the expression in (9), we obtain

$$E(T_{\mathbf{\Pi}}) = -\mathbf{W}^\top \mathbf{\Pi} \boldsymbol{\Psi} E(\lambda \mathbf{B}_n \mathbf{P} \mathbf{f} + \mathbf{B}_n \boldsymbol{\Psi}^\top \boldsymbol{\epsilon} / n) + E(\boldsymbol{\epsilon}) = -\lambda \mathbf{W}^\top \mathbf{\Pi} \boldsymbol{\Psi} \mathbf{B}_n \mathbf{P} \mathbf{f}.$$

A similar procedure can be used to compute the variance:

$$\text{Var}(T_{\mathbf{\Pi}}) = \text{Var}\{\mathbf{W}^\top \mathbf{\Pi}(\mathbf{z} - \boldsymbol{\Psi}\hat{\mathbf{f}})\} = \mathbf{W}^\top \mathbf{\Pi} \text{Var}(\boldsymbol{\Psi}\mathbf{f} - \boldsymbol{\Psi}\hat{\mathbf{f}} + \boldsymbol{\epsilon}) \mathbf{\Pi} \mathbf{W}.$$

Using again the expression in (9) and the independence of the residuals, we obtain

$$\text{Var}(T_{\mathbf{\Pi}}) = \mathbf{W}^\top \mathbf{\Pi} \boldsymbol{\Psi} \text{Var}(\lambda \mathbf{B}_n \mathbf{P} \mathbf{f} + \mathbf{B}_n \boldsymbol{\Psi}^\top \boldsymbol{\epsilon} / n) \boldsymbol{\Psi}^\top \mathbf{\Pi} \mathbf{W} + \mathbf{W} \text{Var}(\boldsymbol{\epsilon}) \mathbf{W} = \frac{\sigma^2}{n^2} \mathbf{W}^\top \mathbf{\Pi} (\boldsymbol{\Psi} \mathbf{B}_n \boldsymbol{\Psi}^\top)^2 \mathbf{\Pi} \mathbf{W} + \sigma^2 \mathbf{W}^\top \mathbf{W}.$$

□

The following proposition demonstrates that the test  $T_{\mathbf{\Pi}}$  is asymptotically exact.

**Proposition 3.** *If  $\lambda = o(n^{-1/2})$ , the test  $T_{\mathbf{\Pi}}$  is asymptotically exact. In particular, the test  $T_{\mathbf{\Pi}}$  is asymptotically unbiased and second order exchangeable.*

**Proof.** We consider the expected value and the variance under the null hypothesis. For the expected value, for any given  $\mathbf{\Pi}$ , we have

$$E(T_{\mathbf{\Pi}}) = -\mathbf{W}^\top \mathbf{\Pi} \boldsymbol{\Psi} E(\hat{\mathbf{f}} - \mathbf{f}) + \mathbf{W}^\top \mathbf{\Pi} E(\boldsymbol{\epsilon}).$$

Using Theorem 1, we derive

$$E(T_{\mathbf{\Pi}}) \xrightarrow{n} 0, \quad n \rightarrow \infty,$$

thus proving the asymptotic unbiasedness of the test. For the variance, for any given  $\mathbf{\Pi}$ , we have

$$\begin{aligned} \text{Var}(T_{\mathbf{\Pi}}) &= \text{Var}\{\mathbf{W}^\top \mathbf{\Pi}(\mathbf{z} - \boldsymbol{\Psi}\hat{\mathbf{f}})\} = \mathbf{W}^\top \mathbf{\Pi} \text{Var}(\boldsymbol{\Psi}\mathbf{f} - \boldsymbol{\Psi}\hat{\mathbf{f}} + \boldsymbol{\epsilon}) \mathbf{\Pi} \mathbf{W} \\ &= \mathbf{W}^\top \mathbf{\Pi} \boldsymbol{\Psi} \text{Var}(\hat{\mathbf{f}} - \mathbf{f}) \boldsymbol{\Psi}^\top \mathbf{\Pi} \mathbf{W} + \sigma^2 \mathbf{W}^\top \mathbf{W}. \end{aligned}$$

Using again Proposition 1, we obtain

$$\text{Var}(T_{\Pi}) \xrightarrow{n} \sigma^2 \mathbf{W}^T \mathbf{W}, \quad n \rightarrow \infty,$$

concluding the proof.  $\square$

Note that the finite-sample variance  $\text{Var}(T_{\Pi})$  derived in Proposition 2 depends on the sign-flip and it is only asymptotically second order exchangeable. This may lead to a decrease in power in the finite sample scenario. This problem is discussed in detail in the next section, where we propose a possible solution that preserves the finite sample invariance of the covariance structure of the test under random sign-flips.

#### 4.2. Eigen sign-flip score test for SR-PDE

We now propose a modification of the sign-flip test introduced in Section 4.1, which is aimed to avoid the issues related to the lack of invariance under sign-flips of the covariance of the statistics  $T_{\Pi}$ . We name this new test Eigen sing-flip test.

Let us define the eigen decomposition  $\Psi \mathbf{B}_n \Psi^T = \mathbf{U} \mathbf{D} \mathbf{U}^T$ , where  $\mathbf{U}$  is the square  $N_{\mathcal{T}} \times N_{\mathcal{T}}$  matrix having as columns the eigenvectors of  $\Psi \mathbf{B}_n \Psi^T$ , and  $\mathbf{D}$  is the diagonal matrix whose diagonal elements are the corresponding eigenvalues. We define a new test statistic as

$$\tilde{T}_{\Pi} = \mathbf{W}^T \mathbf{U} \mathbf{D} \mathbf{U}^T (\mathbf{z} - \hat{\Psi} \mathbf{f}). \quad (14)$$

The post-multiplication by  $\mathbf{U}$  of  $\mathbf{W}$  and the pre-multiplication by  $\mathbf{U}^T$  of the vector of residuals act as distance-preserving transformations, orthogonal to each other. We now show the effect of this transformation on the variance of the test statistic.

**Proposition 4.** *The variance of the test  $\tilde{T}_{\Pi}$ , under the null hypothesis, is invariant with respect to  $\Pi$ , that is*

$$\text{Var}(\tilde{T}_{\Pi}) = \sigma^2 \mathbf{W}^T \left\{ \mathbf{I} + \frac{1}{n^2} (\Psi \mathbf{B}_n \Psi^T)^2 \right\} \mathbf{W}.$$

**Proof.** For any given  $\Pi$ , we have

$$\begin{aligned} \text{Var}(\tilde{T}_{\Pi}) &= \text{Var}\{\mathbf{W}^T \mathbf{U} \mathbf{D} \mathbf{U}^T (\mathbf{z} - \hat{\Psi} \mathbf{f})\} = \mathbf{W}^T \mathbf{U} \mathbf{D} \mathbf{U}^T \Psi \text{Var}(\lambda \mathbf{B}_n \mathbf{P} \mathbf{f} + \mathbf{B}_n \Psi^T \boldsymbol{\epsilon} / n) \mathbf{U} \mathbf{D} \mathbf{U}^T \mathbf{W} + \mathbf{W} \text{Var}(\boldsymbol{\epsilon}) \mathbf{W} \\ &= \frac{\sigma^2}{n^2} \mathbf{W}^T \mathbf{U} \mathbf{D} \mathbf{U}^T (\Psi \mathbf{B}_n \Psi^T) (\Psi \mathbf{B}_n \Psi^T) \mathbf{U} \mathbf{D} \mathbf{U}^T \mathbf{W} + \sigma^2 \mathbf{W}^T \mathbf{W}. \end{aligned}$$

We can now use the spectral decomposition  $\Psi \mathbf{B}_n \Psi^T = \mathbf{U} \mathbf{D} \mathbf{U}^T$  and the fact that  $\mathbf{U}$  is an orthogonal matrix to obtain

$$\text{Var}(\tilde{T}_{\Pi}) = \frac{\sigma^2}{n^2} \mathbf{W}^T \mathbf{U} \mathbf{D} \mathbf{U}^T (\mathbf{U} \mathbf{D} \mathbf{U}^T) (\mathbf{U} \mathbf{D} \mathbf{U}^T) \mathbf{U} \mathbf{D} \mathbf{U}^T \mathbf{W} + \sigma^2 \mathbf{W}^T \mathbf{W} = \frac{\sigma^2}{n^2} \mathbf{W}^T \mathbf{U} \mathbf{D}^2 \mathbf{U}^T \mathbf{W} + \sigma^2 \mathbf{W}^T \mathbf{W},$$

leading to the desired result.  $\square$

The idea behind the eigen decomposition is inspired by similar approaches proposed by [24, 25] for the classical linear model. In the context of classical linear models, the covariance matrix of the residuals is not diagonal, thus making the observations not exchangeables. To ensure a safe application of the permutation principle, the authors thus introduce the premultiplication of residuals and predictors by the eigenvectors of the residual projection matrix. Since the residual projection matrix is positive semi-definite and idempotent, its eigenvalues are only ones and zeros. This makes the resulting transformed residuals exchangeable (and reduced in number, equal to the rank of the residual projection matrix). In our case, the smoothing matrix is not a projection matrix; therefore, the eigenvalues do not take values in  $\{0, 1\}$ . Because of this, the rescaled residuals, on which the test statistic in (14) is based, are not homoscedastic, and the permutation approach is hence inapplicable. For this reason our proposal, unlike the method of [25], leverages instead on the sign-flip procedure. The proposed approach, together with the decomposition of the smoothing matrix, leads to a preservation of the covariance structure with respect to random sign-flips, as shown in Proposition 4.

As highlighted by the simulations in Section 5, the Eigen sign-flip score test here presented, with respect to the sign-flip score test in Section 4.1, ensures an higher power, while keeping a good control of the Type-I error.

## 5. Simulation studies

In this section we present some simulation studies that investigate the finite sample performances of the proposed tests. We compare our tests to a classical Wald type test, based on the asymptotic distribution of  $\hat{\beta}$ , and to the possibly more robust sandwich version of the same test, that can be derived along the lines detailed in [48].

We consider the true spatial field  $f(x, y) = \cos\{(2x+y)/4\} + \{(x+y)/15\}^2$  on the square with vertices  $[0, 10] \times [0, 10]$ . We generate the covariates as random fields with zero mean and different covariance structures, using the function `RFsimulate` of the R package `RandomFields` [40]. In particular, we consider four different test cases, with one covariate each, generated as follows:

- a. Gaussian random field with scale  $s = 0.3$ ;
- b. Matérn random field with scale  $s = 0.5$ , smoothness  $\eta = 1$  and variance  $v = 8$ ;
- c. Exponential random field with scale  $s = 2$  and variance  $v = 1$ ;
- d. Matérn random field with scale  $s = 1$ , smoothness  $\eta = 5$  and variance  $v = 2$ .

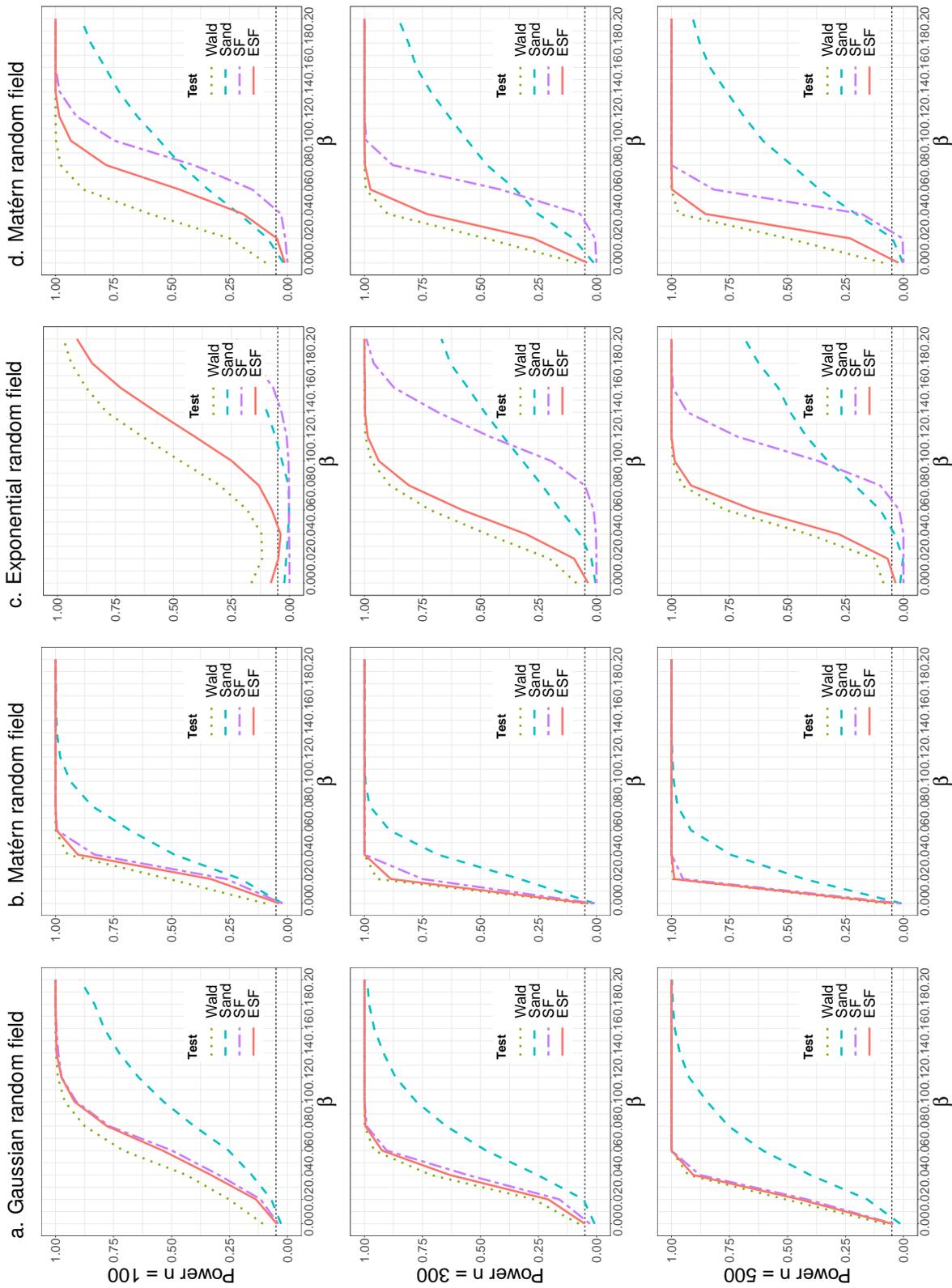
The covariates considered in test cases a-d have increasing spatial dependence. We sample the spatial field and covariates at  $n$  random locations on the square, considering increasing values of  $n$ , equal to 100, 300 and 500. We hence obtain the samples of the response variable from model (1), adding i.i.d. random errors with standard deviation 0.25, and considering increasing values of the coefficient  $\beta$ . The value  $\beta = 0$  corresponds to the null hypothesis and is used to check the control of Type-I error. The other values of  $\beta$  are considered to verify the power of the test with increasing values of the true regression coefficient. Data generation is repeated 500 times, keeping fixed the data locations and covariates, and regenerating the i.i.d. noise.

The SR-PDE model is estimated using the R package `fdaPDE` [28], with a regular mesh with 278 nodes, and selecting the smoothing parameter via generalized cross-validation. The critical value of the sign-flip test statistics are computed using 1000 random sign-flips.

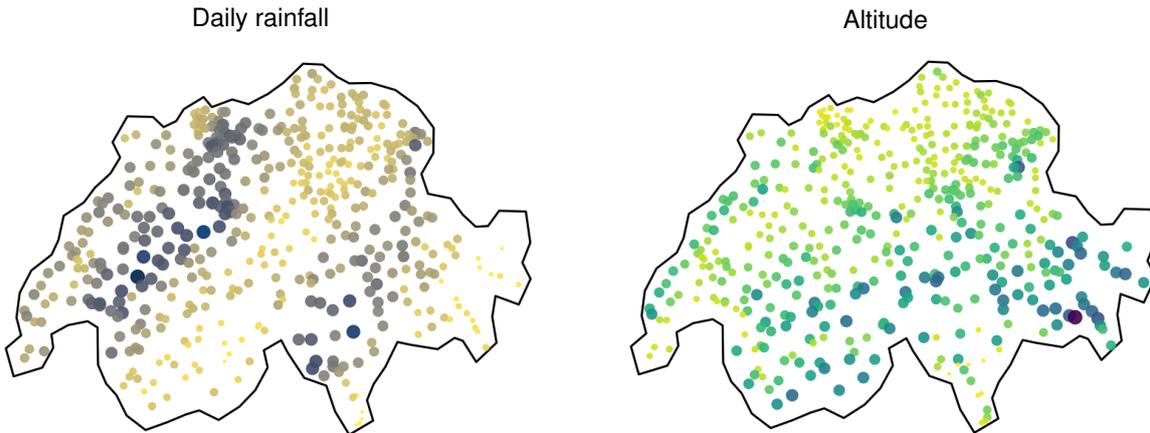
The results are presented in Fig. 2, which shows the power curves for the considered tests: the dotted olive lines correspond to the classical parametric Wald test (Par), the dashed cyan lines to the sandwich variant of the Wald test (Sand), the dashed-dotted purple lines to the sign-flip test in Section 4.1 (S-F), and the solid red lines to the Eigen sign-flip test in Section 4.2 (Eig S-F). The classical Wald test (dotted olive lines) shows a poor control of the Type-I error in all test cases, especially for small samples. This behavior is possibly due to the poor estimation of the variance induced by the regularized estimates. The sandwich variant (dashed cyan lines) is more robust and has a better control of Type-I error; on the other hand, this test is over-conservative, especially when a strong spatial dependence is present (see in particular cases c and d), showing a lower power with respect to the alternatives. The sign-flip procedure (dashed-dotted purple lines) displays a similar behaviour: although it has a decent control of Type-I error, it loses power in the case of strong dependence between the score contributions (see in particular cases c and d). This is due to the non exchangeability of the score contributions, that leads to an over-estimation of the variance of the estimator and therefore to a lower power. The eigen sign-flip (solid red lines), on the contrary, it is not affected by such problem and it is able to preserve the covariance structure of the test statistic also in the finite sample scenario, even when the covariate exhibits a strong spatial dependence. This leads, even in the more complicated cases, to a very good control of Type-I error, accompanied by a high power.

## 6. Application to Switzerland rainfall

We finally show an application of the proposed Eigen sign-flip test to real data, concerning rainfall measurements. In particular, we consider a dataset of daily rainfall values recorded in Switzerland on May 8, 1986, in 467 meteorological stations; this dataset was used for the Spatial Interpolation Comparison 97 [11]. The data are displayed in the left panel of Fig. 3: the size and color of the point marker represent the rainfall measurement. These data display a strong spatial correlation, with a clearly anisotropic structure, with dominating North-East/South-West direction. The data also include the elevation at the same 467 locations, shown in the right panel of the same figure, that we here consider as a covariate. Intuition in fact suggests that the orography of the region may play an important role in the rainfall phenomenon; this is also hinted by a visual inspection of the map of altitudes of the meteorological



**Fig. 2:** Comparison of power curves for the different test statistics. The dotted olive lines correspond to the Wald test; the dashed cyan lines to sandwich variant of the Wald test; the dashed-dotted purple lines to the standard sign-flip; the solid red lines to the eigen sign-flip. Sample sizes are  $n=100$  (top panels),  $n=300$  (central panels) and  $n=500$  (bottom panels).



**Fig. 3:** Switzerland rainfall data. Left: values of daily rainfall recorded at 467 meteorological stations in Switzerland on May 8, 1986; darker color and bigger size of the point marker represent higher rainfall; a strongly anisotropic pattern is clearly appreciable, with dominating North-East / South-West direction. Right: altitude of the meteorological stations; darker color and bigger size of the point marker represent higher altitude).

stations, which highlights elongated mountain chains and valleys with North-East/South-West orientation. We here explore such intuition, formalizing the analysis as a testing problem on the linear regression term of a semiparametric SR-PDE model.

To appropriately account for the strong anisotropy in the rainfall data, we consider the anisotropic SR-PDE method described in [6]. In particular, we consider a regularizing term involving the diffusion term  $\nabla \cdot (\mathbf{K}\nabla f)$ , where  $\nabla = (\partial/\partial p_1, \partial/\partial p_2)^\top$  and  $\mathbf{K}$  is a  $2 \times 2$  symmetric and positive definite matrix that controls the anisotropy; this operator coincide with the Laplacian in (3) when the matrix  $\mathbf{K}$  is the identity. The matrix  $\mathbf{K}$  is considered unknown and estimated from data, thus identifying the main direction and the intensity of the anisotropy, as detailed in [6]. The proposed Eigen sign-flip test naturally extends to this setting, by appropriate changes of the matrix  $\mathbf{P}$  that discretises the penalty term. More generally, as discussed in the final Discussion, the Eigen sign-flip test could be extended to SR-PDE models with complex regularizations involving general forms of PDEs, as considered in [5], thus broadening the range of applicability of the test to complex spatial variation settings.

The anisotropic SR-PDE model is constructed under the same specifications in [6], where these data are also analysed. In [6] elevation was considered not-significant, on the basis of parametric inference. We may though wonder about the validity of this result, in the light of the poor performances of parametric inference on  $\beta$ , in the context of semiparametric spatial regression, evidenced by the simulations in Section 5. We hence apply the Eigen sign-flip test, with 1000 resamples, testing the null hypothesis  $\beta_0 = 0$ . The test returns a  $p$ -value of 0.326; this concludes quite clearly that the elevation does indeed not have a significant impact on the rainfall. This is probably due to the fact that the effect of elevation on rainfall is not linear, so that this effect is not captured by a model where elevation is included as a linear regressor. Rather, the distribution of rainfall is the result of complex interaction between the geomorphology of the region and the atmospheric circulation.

## 7. Discussion

The definition of the proposed test is fairly flexible and allows for various types of generalization. In particular, the proposed approach could be extended to deal with the semiparametric regression models based on penalized regression splines [19, 42, 48], thin plate splines [46] and soap film smoothing [47]. The main difference would reside in the construction of the smoothing matrices. A thorough analysis of the estimators and the properties of the derived test statistics in these cases is necessary, and will be the object of future studies. Moreover, extension to SR-PDE models over nonplanar two-dimensional domains proposed in [12] appears natural. An interesting direction for future research also concerns the extension to the generalized linear SR-PDE model framework presented in [44]. The definition of eigen sign-flip tests for the spatio-temporal regression models in [2, 7] is also highly desirable, even

more since the study of the asymptotic properties of these models is far from trivial. Finally, while this paper focuses on hypothesis testing, similar ideas could be used for the definition of appropriate confidence intervals, as in [15].

**Acknowledgments.** We are grateful to the Guest Editor and external Reviewers for constructive comments.

## References

- [1] G. Aneiros, R. Cao, R. Fraiman, C. Genest, P. Vieu, Recent advances in functional data analysis and high-dimensional statistics, *J. Multivariate Anal.* 170 (2019) 3–9.
- [2] E. Arnone, L. Azzimonti, F. Nobile, L. M. Sangalli, Modeling spatially dependent functional data via regression with differential regularization, *J. Multivariate Anal.* 170 (2019) 275–295.
- [3] E. Arnone, A. Kneip, F. Nobile, L. M. Sangalli, Some first results on the consistency of spatial regression with partial differential equation regularization, *Statistica Sinica* (2021) DOI: 10.5705/ss.202019.0346.
- [4] L. Azzimonti, F. Nobile, L. M. Sangalli, P. Secchi, Mixed finite elements for spatial regression with pde penalization, *SIAM/ASA Journal on Uncertainty Quantification* 2 (2014) 305–335.
- [5] L. Azzimonti, L. M. Sangalli, P. Secchi, M. Domanin, F. Nobile, Blood flow velocity field estimation via spatial regression with pde penalization, *Journal of the American Statistical Association* 110 (2015) 1057–1071.
- [6] M. S. Bernardi, M. Carey, J. O. Ramsay, L. M. Sangalli, Modeling spatial anisotropy via regression with partial differential regularization, *J. Multivariate Anal.* 167 (2018) 15–30.
- [7] 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, *Stochastic Environmental Research and Risk Assessment* 31 (2017) 23–38.
- [8] N. Cressie, C. K. Wikle, *Statistics for Spatio-Temporal Data*, Wiley Series in Probability and Statistics, John Wiley & Sons, Inc., Hoboken, NJ, 2011.
- [9] N. A. C. Cressie, *Statistics for Spatial Data*, Wiley Classics Library, John Wiley & Sons, Inc., New York, revised edition, 2015. Paperback edition of the 1993 edition [MR1239641].
- [10] P. J. Diggle, P. J. Ribeiro, Jr., *Model-based geostatistics*, Springer Series in Statistics, Springer, New York, 2007.
- [11] G. Dubois, J. Malczewski, M. De Cort, Mapping radioactivity in the environment: spatial interpolation comparison 97, Office for Official Publications of the European Communities, 2003.
- [12] B. Ettinger, S. Perotto, L. M. Sangalli, Spatial regression models over two-dimensional manifolds, *Biometrika* 103 (2016) 71–88.
- [13] F. Ferraty, P. Vieu, *Nonparametric Functional Data Analysis*, Springer Series in Statistics, Springer, New York, 2006.
- [14] D. A. Freedman, On the so-called “Huber sandwich estimator” and “robust standard errors”, *The American Statistician* 60 (2006) 299–302.
- [15] P. H. Garthwaite, Confidence intervals from randomization tests, *Biometrics* (1996) 1387–1393.
- [16] A. Goia, P. Vieu, An introduction to recent advances in high/infinite dimensional statistics, *J. Multivariate Anal.* 146 (2016) 1–6.
- [17] R. J. Gray, Spline-based tests in survival analysis, *Biometrics* (1994) 640–652.
- [18] W. Härdle, H. Liang, J. Gao, *Partially Linear Models*, Physica-Verlag, Heidelberg, 2000.
- [19] N. E. Heckman, Spline smoothing in a partly linear model, *Journal of the Royal Statistical Society: Series B (Methodological)* 48 (1986) 244–248.
- [20] J. Hemerik, J. J. Goeman, False discovery proportion estimation by permutations: confidence for significance analysis of microarrays, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 80 (2018) 137–155.
- [21] J. Hemerik, J. J. Goeman, L. Finos, Robust testing in generalized linear models by sign flipping score contributions, *J. R. Stat. Soc. Ser. B. Stat. Methodol.* 82 (2020) 841–864.
- [22] J. Hemerik, A. Solari, J. J. Goeman, Permutation-based simultaneous confidence bounds for the false discovery proportion, *Biometrika* 106 (2019) 635–649.
- [23] Ø. Hjelle, M. Dæhlen, *Triangulations and Applications*, Springer, Berlin, 2006.
- [24] M.-H. Huh, M. Jhun, Random permutation testing in multiple linear regression, *Communications in statistics-Theory and Methods* 30 (2001) 2023–2032.
- [25] S. Kherad-Pajouh, O. Renaud, An exact permutation method for testing any effect in balanced and unbalanced fixed effect anova, *Computational Statistics & Data Analysis* 54 (2010) 1881–1893.
- [26] P. Kokoszka, M. Reimherr, *Introduction to Functional Data Analysis*, Chapman & Hall/CRC Texts in Statistical Science, Chapman & Hall/CRC, Boca Raton, 2017.
- [27] E. Lila, J. A. Aston, L. M. Sangalli, et al., Smooth principal component analysis over two-dimensional manifolds with an application to neuroimaging, *The Annals of Applied Statistics* 10 (2016) 1854–1879.
- [28] E. Lila, L. M. Sangalli, E. Arnone, J. 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*, 2020. R package version 1.0-9.
- [29] C. J. Maas, J. J. Hox, Robustness issues in multilevel regression analysis, *Statistica Neerlandica* 58 (2004) 127–137.
- [30] J. Mateu, E. Romano, Advances in spatial functional statistics, *Stochastic Environmental Research and Risk Assessment* 31 (2017) 1–6.
- [31] J. Mateu Mahiques, R. Giraldo, *Geostatistical Functional Data Analysis*, Wiley, New York, 2021.
- [32] F. Pesarin, *Multivariate Permutation Tests: with Applications in Biostatistics*, volume 240, Wiley Chichester, 2001.
- [33] F. Pesarin, L. Salmaso, *Permutation Tests for Complex Data: Theory, Applications and Software*, John Wiley & Sons, New York, 2010.
- [34] J. O. Ramsay, B. W. Silverman, *Functional Data Analysis*, Springer Series in Statistics, Springer, New York, second edition, 2005.
- [35] D. Ruppert, M. P. Wand, R. J. Carroll, *Semiparametric Regression*, 12, Cambridge University Press, Cambridge, 2003.
- [36] D. Ruppert, M. P. Wand, R. J. Carroll, Semiparametric regression during 2003–2007, *Electronic Journal of Statistics* 3 (2009).

- [37] L. M. Sangalli, Spatial regression with partial differential equation regularization, *International Statistical Review* (2021) DOI: 10.1111/insr.12444.
- [38] L. M. Sangalli, J. O. Ramsay, T. O. Ramsay, Spatial spline regression models, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 75 (2013) 681–703.
- [39] M. J. Schervish, *Theory of Statistics*, Springer Science & Business Media, New York, 2012.
- [40] M. Schlather, A. Malinowski, P. J. Menck, M. Oesting, K. Stokorb, Analysis, simulation and prediction of multivariate random fields with package *RandomFields*, *Journal of Statistical Software* 63 (2015) 1–25.
- [41] A. Solari, L. Finos, J. J. Goeman, Rotation-based multiple testing in the multivariate linear model, *Biometrics* 70 (2014) 954–961.
- [42] M. Wand, J. Ormerod, On semiparametric regression with O’Sullivan penalized splines, *Australian & New Zealand Journal of Statistics* 50 (2008) 179–198.
- [43] P. H. Westfall, S. S. Young, *Resampling-Based Multiple Testing: Examples and Methods for p-value Adjustment*, volume 279, John Wiley & Sons, New York, 1993.
- [44] M. Wilhelm, L. M. Sangalli, Generalized spatial regression with differential regularization, *Journal of Statistical Computation and Simulation* 86 (2016) 2497–2518.
- [45] A. M. Winkler, G. R. Ridgway, M. A. Webster, S. M. Smith, T. E. Nichols, Permutation inference for the general linear model, *Neuroimage* 92 (2014) 381–397.
- [46] S. N. Wood, Thin plate regression splines, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 65 (2003) 95–114.
- [47] S. N. Wood, M. V. Bravington, S. L. Hedley, Soap film smoothing, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 70 (2008) 931–955.
- [48] Y. Yu, D. Ruppert, Penalized spline estimation for partially linear single-index models, *Journal of the American Statistical Association* 97 (2002) 1042–1054.

## MOX Technical Reports, last issues

Dipartimento di Matematica  
Politecnico di Milano, Via Bonardi 9 - 20133 Milano (Italy)

- 78/2021** Bucelli, M.; Dede', L.; Quarteroni, A.; Vergara, C.  
*Partitioned and monolithic algorithms for the numerical solution of cardiac fluid-structure interaction*
- 76/2021** Ponti, L.; Perotto, S.; Sangalli, L.M.  
*A PDE-regularized smoothing method for space-time data over manifolds with application to medical data*
- 77/2021** Guo, M.; Manzoni, A.; Amendt, M.; Conti, P.; Hesthaven, J.S.  
*Multi-fidelity regression using artificial neural networks: efficient approximation of parameter-dependent output quantities*
- 73/2021** Marcinno, F.; Zingaro, A.; Fumagalli, I.; Dede', L.; Vergara, C.  
*A computational study of blood flow dynamics in the pulmonary arteries*
- 75/2021** Cicci, L.; Fresca, S.; Pagani, S.; Manzoni, A.; Quarteroni, A.  
*Projection-based reduced order models for parameterized nonlinear time-dependent problems arising in cardiac mechanics*
- 74/2021** Orlando, G.; Barbante, P. F.; Bonaventura, L.  
*An efficient IMEX-DG solver for the compressible Navier-Stokes equations with a general equation of state*
- 71/2021** Franco, N.; Manzoni, A.; Zunino, P.  
*A Deep Learning approach to Reduced Order Modelling of parameter dependent Partial Differential Equations*
- 72/2021** Fresca, S.; Manzoni, A.  
*POD-DL-ROM: enhancing deep learning-based reduced order models for nonlinear parametrized PDEs by proper orthogonal decomposition*
- 70/2021** Beirao da Veiga, L.; Canuto, C.; Nochetto, R.H.; Vacca, G.; Verani, M.  
*Adaptive VEM: Stabilization-Free A Posteriori Error Analysis*
- 69/2021** Antonietti, P.F.; Caldana, M.; Dede', L.  
*Accelerating Algebraic Multigrid Methods via Artificial Neural Networks*