



MOX-Report No. 64/2024

**Sign-Flip inference for spatial regression with differential
regularization**

Cavazzutti, M.; Arnone, E.; Ferraccioli, F.; Galimberti, C.; Finos, L.; Sangalli,
L.M.

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

mox-dmat@polimi.it

<https://mox.polimi.it>

Sign-Flip inference for spatial regression with differential regularization

M. Cavazzutti¹, E. Arnone², F. Ferraccioli³, C. Galimberti¹, L. Finos³, and L. M. Sangalli^{*1}

¹MOX - Dept. of Mathematics, Politecnico di Milano, Milan, Italy

²Dept. of Management, University of Torino, Turin, Italy

³Dept. of Statistical Science, University of Padova, Padua, Italy

Abstract

We address the problem of performing inference on the linear and nonlinear terms of a semiparametric spatial regression model with differential regularization. For the linear term, we propose a new resampling procedure, based on (partial) sign-flipping of an appropriate transformation of the residuals of the model. The proposed resampling scheme can mitigate the bias effect, induced by the differential regularization. We prove that the proposed test is asymptotically exact. Moreover, we show by simulation studies that it enjoys very good control of Type-I error also in small sample scenarios, differently from parametric alternatives. Furthermore, we show that the proposed test has higher power with respect to recently proposed nonparametric tests on the linear term of semiparametric regression models with differential regularization. Concerning the nonlinear term, we develop three different inference approaches: a parametric test, and two nonparametric alternatives. The nonparametric tests are based on a sign-flip approach. One of these tests is proved to be asymptotically exact, while the other is proved to be exact also for finite samples. Simulation studies highlight the very good control of Type-I error of the nonparametric approaches, while retaining high power.

Keywords: Smoothing, roughness penalties, semiparametric regression, resampling tests.

1 Introduction

This work develops inference procedures for the linear and nonlinear terms of Spatial Regression with Partial Differential Equation regularization (SR-PDE). This is a rich class of semiparametric regression models for data distributed over space and space-time, observed over complex spatial domains. These methods, originally introduced in Sangalli et al. [2013] and Azzimonti et al. [2015], have been extended in various modeling directions by subsequent works, as reviewed, e.g., in Sangalli [2021].

Semiparametric regression models with roughness penalties have been extensively studied in more classical settings, as evidenced by the numerous textbooks and references available in the field; see, e.g., Green and Silverman 1994, Bickel et al. 1998, Eubank 1999, Härdle et al. 2000, Ruppert et al. 2003, and references therein. Due to their flexibility and versatility, these models have been the subject of a large and still ongoing research. Various approaches have been proposed to make inference for semiparametric regression models, as discussed in, e.g., Ruppert et al. [2003], Harezlak et al. [2018] and Wood [2017]. Regarding the linear component of the models, some possibilities include the generalization of undersmoothing approaches, that were originally developed for nonparametric models (a review can be found in, e.g., Hall and Horowitz 2013), as well as corrections of Wald-type test statistics, such as the Speckman's correction in Speckman [1988] and Holland [2017]. Similar results can be obtained by interpreting the model in a mixed-effects framework, as done for instance by Wood [2017]. Concerning inference on the nonlinear component, most of the research in the context of semiparametric regression focus on the asymptotic properties of the estimators; see, e.g., Xiao [2019] and Claeskens et al. [2009], for a review in the context of penalized spline estimators. For smoothing spline models, Liu and Wang [2004] conduct various simulations to compare the performances of several parametric tests, like the locally most powerful test by Cox et al. [1988], the generalized maximum likelihood ratio test, and the generalized cross validation test by Wahba [1990], estimating their null distribution via computationally intensive Monte

*Correspondence: Laura.sangalli@polimi.it

Carlo methods. Crainiceanu et al. [2005] explore likelihood ratio tests for testing polynomial regression versus an alternative model based on penalized splines, relying on a mixed-effects model representation, and assuming Gaussian errors. Within the same context, Chen et al. [2014] propose an alternative approach based on the smoothing spline estimator, in order to test the significance of the regression function and its nonparametric deviation from a polynomial model. Wood [2013] proposes a Wald type test statistic to test the significance of the smooth components of an extended generalized additive model. Yu et al. [2019] instead propose confidence bands in generalize geoadditive models. Another possibility is to reformulate the model in a Bayesian framework, as demonstrated for instance in Marra and Wood [2012], Nychka [1988, 1990], Silverman [1985], Wahba [1983], Wang [2011] and Wood [2017] in analogous contexts.

Despite this large body of literature on inference for classical semiparametric regression models with roughness penalties, the definition of correct and powerful inference procedures for SR-PDE still presents many challenges. Ferraccioli et al. [2022] and Ferraccioli et al. [2023] make some first proposals for inference on the linear term of these models, showing that parametric tests, based on Wald and Speckman approaches, have poor performances. In particular, they show that the classical parametric tests on the linear term have poor control of Type-I error, due to the bias induced by the regularizing term of SR-PDE models. They hence propose a nonparametric alternative, based on a resampling approach [Pesarin, 2001], that leverages on the innovative sign-flip procedure of the scores of the model, introduced in more classical linear regression settings by Hemerik et al. [2020]. However, this resampling test is still affected by the bias induced by the PDE regularization, resulting in a reduction in power. Moreover, inference on the nonlinear term of the models has not been explored yet.

Here we tackle inference for both the linear and nonlinear components of SR-PDE models. The work is composed of two parts. In the first one, we investigate in detail the impact of the bias on the nonparametric test for the linear component, proposed by Ferraccioli et al. [2022]. We show that the components of the Eigen-Sign-Flip statistic defined by Ferraccioli et al. [2022] are differently affected by the bias induced by the regularizing term. Moreover, we demonstrate that the different impact of the bias on the Eigen-Sign-Flip components is related to the spectral properties of the residualizing matrix associated with the nonparametric part of the model. We hence propose a Partial Eigen-Sign-Flip procedure, that limits the impact of the bias, by sign-flipping only some components of the considered statistic. We demonstrate that the proposed partial resampling test is asymptotically exact. We moreover show by means of simulation studies that the proposed test has a good control of Type-I error also in small samples, and that it enjoys higher power than the proposal by Ferraccioli et al. [2022]. In the second part of this paper, we focus on developing tests for the nonlinear component of the SR-PDE models, an issue that has never been addressed in previous studies. We derive a standard parametric approach and two nonparametric alternatives based on the sign-flip of the scores of the model. The two nonparametric tests estimate the conditional distribution of the statistic by resampling, via sign-flip of the residuals, or a transformation of the residuals, under the null hypothesis. This approach is inspired by the proposals by Huh and Jhun [2001] and Kherad-Pajouh and Renaud [2010] in the context of classical linear regression. We demonstrate that the first of these nonparametric approaches is asymptotically exact, whilst the second is exact also in final sample scenarios.

The proposed methods are implemented in the `fdapDE` library [Arnone et al., 2024], available on CRAN, with a particular attention to computational efficiency and sustainability, also in the context of massive datasets, by leveraging on state-of-the-art approaches of numerical linear algebra and stochastic approximation.

The paper is organized as follows. In Section 2 we briefly recall the SR-PDE model and the available inference approaches. In Section 3 we propose a novel approach for inference on the linear part of the model: we first study how the bias affects the Eigen-Sign-Flip test statistic; we then describe the proposed Partial Eigen-Sign-Flip test and demonstrate its asymptotic unbiasedness. In Section 4 we focus on inference for the nonlinear part of the model. We first derive a Wald-like test; we then define two nonparametric methods based on the sign-flip approach; we moreover study the large sample properties of the proposed tests. In Section 5 we report simulations that show the good performances of the proposed inference approaches. Finally, in Section 6 we discuss our results and highlight possible future research.

2 Semiparametric regression with partial differential equation regularization

In this section we briefly recall Spatial Regression with Partial Differential Equation regularization [SR-PDE; see, e.g., Sangalli, 2021; Sangalli et al., 2013]. We consider a bounded domain, $\Omega \subset \mathbb{R}^d$, with $d = 2, 3$, and a smooth boundary $\partial\Omega \subset \mathcal{C}^d$, along with a set $\mathcal{P} = \{\mathbf{p}_i\}_{i \in 1, \dots, n}$ of n spatial locations $\mathbf{p}_i \in \Omega$, for $i \in 1, \dots, n$. At each location \mathbf{p}_i , we observe a variable of interest $y_i \in \mathbb{R}$ and, possibly, a set of covariates $\mathbf{x}_i \in \mathbb{R}^q$. We consider

the following model:

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

where $\boldsymbol{\beta} \in \mathbb{R}^q$ is a vector of unknown regression coefficients, that explains the linear effects of the covariates on the mean of the variable of interest, $f : \Omega \rightarrow \mathbb{R}$ is an unknown field, that characterizes the spatial structure of the phenomenon being studied, and ϵ_i , for $i \in 1, \dots, n$, are uncorrelated homoscedastic errors, with zero mean and finite variance σ^2 . In SR-PDE, the unknown $\boldsymbol{\beta}$ and f are estimated minimizing a regularized sum-of-square-error functional:

$$\mathcal{J}(\boldsymbol{\beta}, f) = \sum_{i=1}^n (y_i - \mathbf{x}_i^\top \boldsymbol{\beta} - f(\mathbf{p}_i))^2 + \lambda \int_{\Omega} (\mathcal{L}f - u)^2, \quad (2)$$

where $\lambda > 0$ is a smoothing parameter and $\mathcal{L}f = u$ is a PDE that expresses the problem-specific knowledge about the unknown spatial field. The regularization term in (2) typically incorporates diffusion, advection, and reaction terms, which enable the modeling of different types of anisotropy and non-stationarity, allowing for the description of complex spatial phenomena; see, e.g., Azzimonti et al. [2015]; Sangalli [2021]. Additional information about the problem can be provided by specifying boundary conditions. These conditions may concern the value of f or its derivatives on the boundary $\partial\Omega$ of the spatial domain, permitting a very flexible modeling of the behaviour of the field at the boundaries of the domain of interest, as detailed, e.g., in Azzimonti et al. [2015]. For simplicity of exposition, in the following discussion we consider $u = 0$ and homogeneous Neumann conditions, that are the natural conditions used when no prior information on the behavior of f at the boundary is available. We provide some insights on the case $u \neq 0$ in Remark 1.

2.1 Discretization of the estimation problem and discrete estimator

The problem of minimization of the functional in (2) does not enjoy a closed analytical solution. A numerical solution can be obtained using the finite element method [see, e.g., Quarteroni et al., 2013]. Let $\Omega_{\mathcal{T}}$ be the discretization of Ω obtained by a triangulation of Ω , such as a Delaunay triangulation [see, e.g., Hjelle and Dæhlen, 2006]. Let $\{\boldsymbol{\xi}_j\}_{j \in 1, \dots, N}$ be the N nodes of $\Omega_{\mathcal{T}}$, that is, the vertices of the elements (triangles or tetrahedra) in $\Omega_{\mathcal{T}}$. Let $\{\psi_j(\mathbf{p}) : \Omega \rightarrow [0, 1]\}_{j \in 1, \dots, N}$ be the set of N linear finite element basis functions defined on $\Omega_{\mathcal{T}}$. These are continuous functions, that are linear once restricted to any element of $\Omega_{\mathcal{T}}$, and such that $\psi_j(\boldsymbol{\xi}_k) = 1$ if $j = k$ and $\psi_j(\boldsymbol{\xi}_k) = 0$ if $j \neq k$. Using an expansion in such bases, the spatial field $f(\cdot)$ can be expressed as:

$$f(\cdot) = \sum_{j=1}^N f_j \psi_j(\cdot) = \mathbf{f} \cdot \boldsymbol{\psi}(\cdot),$$

where $\boldsymbol{\psi} = [\psi_1, \dots, \psi_N]^\top$ is the vector of N finite element basis and $\mathbf{f} = [f_1, f_2, \dots, f_N]^\top = [f(\boldsymbol{\xi}_1), f(\boldsymbol{\xi}_2), \dots, f(\boldsymbol{\xi}_N)]^\top$ is the vector of coefficients of the basis expansion. Set $\boldsymbol{\psi}_x = [\partial\psi_1/\partial x, \dots, \partial\psi_N/\partial x]^\top$, $\boldsymbol{\psi}_y = [\partial\psi_1/\partial y, \dots, \partial\psi_N/\partial y]^\top$ and $\nabla\boldsymbol{\psi} = [\boldsymbol{\psi}_x, \boldsymbol{\psi}_y]$. Define the $n \times N$ matrix Ψ of evaluations of the basis functions at the spatial locations, $\Psi_{ij} = \psi_j(\mathbf{p}_i)$, and the following $N \times N$ matrices: $R_0 = \int_{\Omega_{\mathcal{T}}} \boldsymbol{\psi} \boldsymbol{\psi}^\top$, $R_1 = \int_{\Omega_{\mathcal{T}}} (\nabla\boldsymbol{\psi}^\top K \nabla\boldsymbol{\psi} + \nabla\boldsymbol{\psi}^\top \mathbf{b} \boldsymbol{\psi}^\top + c \boldsymbol{\psi} \boldsymbol{\psi}^\top)$, and $P = R_1^\top R_0^{-1} R_1$.

Let $\mathbf{y} = [y_1, y_2, \dots, y_n]^\top$ be the vector of the observed values at the n locations. Let also X be the $n \times q$ matrix whose i th row is given by \mathbf{x}_i^\top , the vector of covariates associated with observation y_i at \mathbf{p}_i , and assume that X has full rank. Moreover, set $Q = I - X(X^\top X)^{-1} X^\top$, where I is the identity matrix of dimension $n \times n$. It can be shown [see, e.g., Sangalli, 2021] that a unique pair of estimators $(\hat{\boldsymbol{\beta}}, \hat{\mathbf{f}})$ exists, which solves the discrete form of the estimation problem, and are given by:

$$\begin{aligned} \hat{\mathbf{f}} &= (\Psi^\top \Psi + \lambda P)^{-1} \Psi^\top (\mathbf{y} - X \hat{\boldsymbol{\beta}}) \\ \hat{\boldsymbol{\beta}} &= (X^\top X)^{-1} X^\top (\mathbf{y} - \Psi \hat{\mathbf{f}}). \end{aligned} \quad (3)$$

2.2 Current approaches for inference on the linear component $\boldsymbol{\beta}$

Ferraccioli et al. [2022] show three different approaches for testing

$$H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0 \quad \text{vs} \quad H_1 : \boldsymbol{\beta} \neq \boldsymbol{\beta}_0. \quad (4)$$

Among these, two classical parametric approaches are considered, Wald and Speckman, based respectively on the asymptotic distribution of $\hat{\boldsymbol{\beta}}$ and on Speckman's correction [Speckman, 1988]. Ferraccioli et al. [2022] show that relying on such classical parametric approaches can result in inadequate control of Type-I error, particularly when dealing with covariates that have a strong spatial structure. To address this issue, the same work also proposes a nonparametric alternative to test (4), named Eigen-Sign-Flip (ESF) test, which implicitly

recovers the null distribution via an appropriate nonparametric resampling procedure. We here briefly recall the definition of the ESF test on β . Consider the normal equation for β of model (1), i.e., $X^\top(\mathbf{y} - X\beta - \Psi\mathbf{f}) = 0$, and the corresponding score statistic

$$\mathbf{T}_\beta = X^\top(\mathbf{y} - X\beta_0 - \Psi\hat{\mathbf{f}}_{H_0}), \quad (5)$$

where $\hat{\mathbf{f}}_{H_0}$ is the estimate of \mathbf{f} under H_0 , i.e., $\hat{\mathbf{f}}_{H_0} = (\Psi^\top\Psi + \lambda P)^{-1}\Psi^\top(\mathbf{y} - X\beta_0)$. Let also B and Λ be the matrices:

$$B = \Psi(\Psi^\top\Psi + \lambda P)^{-1}\Psi^\top \quad \text{and} \quad \Lambda = I - B. \quad (6)$$

Consider the singular value decomposition of $\Lambda = VDV^\top$. Set $\Pi = \text{diag}(\pi_1, \dots, \pi_n)$, where $\boldsymbol{\pi} = (\pi_1, \dots, \pi_n)$ is a vector of random sign-flips, i.e., $\boldsymbol{\pi}$ is uniformly distributed in $\{-1, 1\}^n$. Consider the $n \times q$ matrix defined as $\tilde{X} = D^{1/2}V^\top X$ and the vector of score residuals

$$\tilde{\mathbf{r}} = D^{1/2}V^\top \mathbf{r} = D^{1/2}V^\top(\mathbf{y} - X\beta_0). \quad (7)$$

Starting from equation (5), plugging-in $\hat{\mathbf{f}}_{H_0}$ and rearranging the terms, we can define the statistic $\mathbf{T}_\beta = \tilde{X}^\top \tilde{\mathbf{r}}$. The Eigen-Sign-Flip statistic $\mathbf{T}_{\beta, \Pi}$ is then obtained by applying the sign-flip matrix Π to the score residuals $\tilde{\mathbf{r}}$:

$$\mathbf{T}_{\beta, \Pi} = n^{-\frac{1}{2}}X^\top V \Pi V^\top \Lambda(\mathbf{y} - X\beta_0) = n^{-\frac{1}{2}}\tilde{X}^\top \Pi \tilde{\mathbf{r}}, \quad (8)$$

where the term $n^{-1/2}$ ensures good asymptotic properties for $\mathbf{T}_{\beta, \Pi}$. The p-value is computed as the relative rank of the observed statistic $\mathbf{T}_{\beta, I}$ with respect to the set of M sign-flip statistics $\mathbf{T}_{\beta, \Pi}$, where Π is resampled each time. Under some regularity assumptions, discussed in Ferraccioli et al. [2022], the test statistic (8) is shown to be asymptotically unbiased and the ESF test is shown to be asymptotically exact.

The main limitation of the three discussed inferential approaches in regularized regression models is that, in finite sample scenarios, the considered statistics suffer from the effect of the bias, induced by the penalty term in the functional (2). In particular, the two parametric tests, Wald and Speckman, pay the price of such bias in term of a very poor control of Type-I error; the ESF tests instead, even though it maintains a good control of Type-I error, pays the price of the bias in terms of a reduced power with respect to the parametric alternatives. This aspect is described in depth in the next section.

Remark 1. It should be mentioned that Ferraccioli et al. [2022] restrict their attention to the special case $\mathcal{L}f = \Delta f$, where Δf is the Laplace operator, with homogeneous forcing terms and boundary conditions, thus considering a simple isotropic and stationary roughness penalty. However, thanks to the asymptotic properties recently obtained in Arnone et al. [2023], the tests proposed by Ferraccioli et al. [2022] and their properties also hold for the general regularizing terms considered in the estimation functional (2). Moreover, all the results in Ferraccioli et al. [2022], as well as the results proposed in the present work, can be extended to $u \neq 0$. Indeed, when $u \neq 0$, plugging $\hat{\mathbf{f}}_{H_0}$ in \mathbf{T}_β leads, after some rearrangements, to a statistic of the form $\mathbf{T}_{\beta, \Pi} = n^{-\frac{1}{2}}X^\top \Lambda(\mathbf{y} - X\beta_0) - \lambda X^\top (\Psi^\top\Psi + \lambda P)^{-1}R_1^\top R_0^{-1}\mathbf{u}$. However, the second term on the right-hand-side is fixed, and, therefore, it can be dropped, thus falling back to the case here detailed.

3 Bias effect on Eigen-Sign-Flip inference on the linear component

β

Consider the expression of the bias of $\mathbf{T}_{\beta, \Pi}$, for any given Π :

$$\mathbf{b}_\lambda := \mathbb{E}[\mathbf{T}_{\beta, \Pi}] = n^{-\frac{1}{2}}X^\top V \Pi D V^\top \Psi \mathbf{f}. \quad (9)$$

From (8) and (9), we observe that we can partition the statistic $\mathbf{T}_{\beta, \Pi}$ and its bias \mathbf{b}_λ into n components, each depending solely on a single sign-flip π_k and a single eigenvalue-eigenvector pair (d_k, \mathbf{v}_k) of the matrix Λ , as stated in the following proposition.

Proposition 1. The statistic (8) and its bias (9) can be decomposed as follows:

$$\mathbf{T}_{\beta, \Pi} = n^{-\frac{1}{2}} \sum_{k=1}^n \mathbf{T}_{\beta, \Pi, k} \quad \text{where} \quad \mathbf{T}_{\beta, \Pi, k} = (X^\top \mathbf{v}_k)(\pi_k d_k)(\mathbf{v}_k^\top (\mathbf{y} - X\beta_0)) \quad (10)$$

$$\mathbf{b}_\lambda = n^{-\frac{1}{2}} \sum_{k=1}^n \mathbf{b}_{\lambda, k} \quad \text{where} \quad \mathbf{b}_{\lambda, k} = (X^\top \mathbf{v}_k)(\pi_k d_k)(\mathbf{v}_k^\top \Psi \mathbf{f}). \quad (11)$$

The proof is reported in Section 1.1 of the Supporting Information. Proposition 1 provides some understanding of the impact of the bias in the ESF test. Indeed, we can rewrite $\mathbf{T}_{\beta, \Pi, k}$ as

$$\begin{aligned}\mathbf{T}_{\beta, \Pi, k} &= (X^\top \mathbf{v}_k) \pi_k d_k^{\frac{1}{2}} \tilde{\mathbf{r}}_k = \\ &= (X^\top \mathbf{v}_k) (\pi_k d_k) (\mathbf{v}_k^\top \Psi \mathbf{f}) + (X^\top \mathbf{v}_k) (\pi_k d_k) (\mathbf{v}_k^\top \boldsymbol{\epsilon}) = \\ &= \mathbf{b}_{\lambda, k} + (X^\top \mathbf{v}_k) (\pi_k d_k) (\mathbf{v}_k^\top \boldsymbol{\epsilon}),\end{aligned}\tag{12}$$

where $\tilde{\mathbf{r}}_k$ is the k th component of the score residuals $\tilde{\mathbf{r}} = D^{1/2} V^\top (\mathbf{y} - X\beta_0)$, in equation (7). Notice that the last term on the right hand side of (12) is zero centered. However, the sign-flip π_k affects not only this zero centered part of $\mathbf{T}_{\beta, \Pi, k}$, but also the bias $\mathbf{b}_{\lambda, k}$, as shown by (11). This implies that the variance of $\Pi \tilde{\mathbf{r}}$ is larger than that of $\tilde{\mathbf{r}}$. As a result, the variance of $\mathbf{T}_{\beta, \Pi}$ is larger than that of \mathbf{T}_β . Figure 1 illustrates this issue. We here consider data generated as in Section 5.1, with locations coinciding with the mesh nodes, and a single covariate generated as $x_i \sim \mathcal{N}(0, 1)$. In the top panel, we plot the distribution of $\tilde{\mathbf{r}}$, resampling the measurement error 5000 times, as described in Section 5.1. In the central panel, we plot the distribution of $\Pi \tilde{\mathbf{r}}$, sampling one sign-flip matrix Π for each generation of the measurement error $\boldsymbol{\epsilon}$. Since the distribution of $\tilde{\mathbf{r}}$, in the top panel, is not zero centered, the distribution of $\Pi \tilde{\mathbf{r}}$ displays a larger variance. The bottom panel of the same figure shows the distribution of the Partial ESF sign-flipped residuals on which our new proposed test statistics, defined in Section 3.2, rely. Before proposing such new test statistics, in Section 3.1, we investigate the link between the bias that affects the k th component of $\mathbf{T}_{\beta, \Pi}$ and the spectral properties of the residualizing matrix Λ .

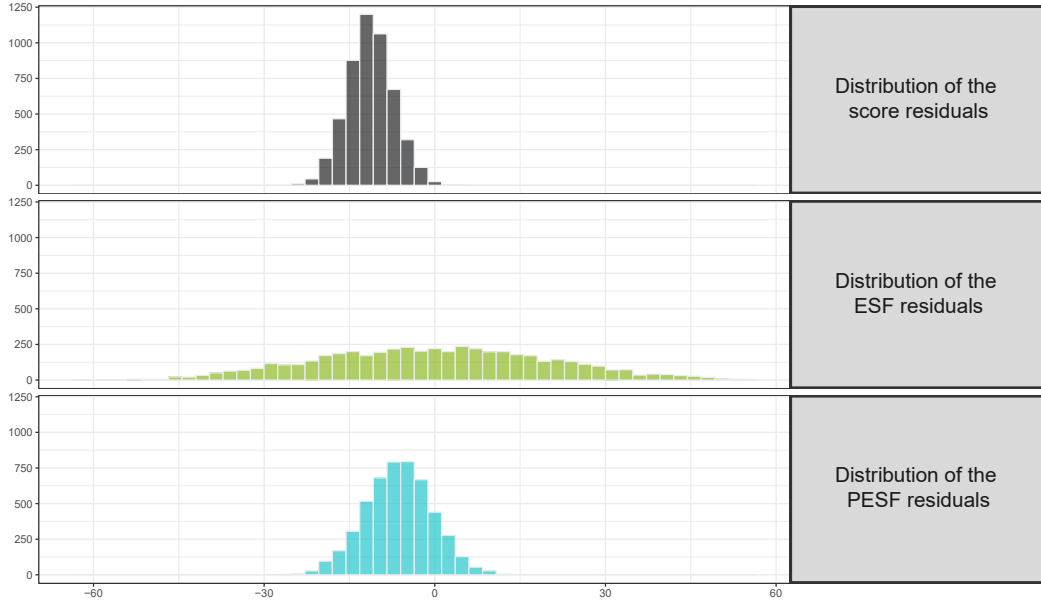


Figure 1: Top panel: distribution of the score residuals $\tilde{\mathbf{r}}$ in (7), for 5000 repetitions of the noise. Central panel: corresponding distribution of the ESF residuals $\Pi \tilde{\mathbf{r}}$, where Π is sampled once for each repetition of the noise. Bottom panel: distribution of the proposed Partial ESF (PESF) residuals $\Pi_{J_{PESF}} \tilde{\mathbf{r}}$, which corrects for the extra-variability of the ESF statistic.

Test statistics

Statistic		Equation
\mathbf{T}_β	Score statistic for the linear component β	(5)
$\mathbf{T}_{\beta, \Pi}$	Eigen-Sign-Flip (ESF) statistic for the linear component β	(8)
$\mathbf{T}_{\beta, \Pi, k}$	k th component of the Eigen-Sign-Flip test for the linear component β	(10)
$\mathbf{T}_{\beta, \Pi, J}$	Partial Eigen-Sign-Flip (PESF) statistic for the linear component β	(13)
\mathbf{T}_f	Score statistic for the nonlinear component f	(15)
$\mathbf{T}_{f, \Pi}$	Sign-Flip (SF) statistic for the nonlinear component f	(17)
$\mathbf{T}_{f, \Pi}^E$	Eigen-Sign-Flip test on the nonlinear component f	(18)

Table 1: For the convenience of the reader, this table indicates the main statistics considered.

3.1 Spectral properties of the residualizing matrix

In this section we show that the bias of the ESF statistic $\mathbf{T}_{\beta, \Pi}$ is concentrated on a few components $\mathbf{T}_{\beta, \Pi, k}$. This fact is related to the spectral properties of the residualizing matrix Λ , and of the associated smoother B , defined in equation (6). Indeed, Proposition 1 highlights that the bias $\mathbf{b}_{\lambda, k}$ of $\mathbf{T}_{\beta, \Pi, k}$ derives from the projection of $\Psi \mathbf{f}$ onto the eigenvector \mathbf{v}_k of Λ (that are the same of B since $\Lambda = I - B$). Moreover, equation (12) shows that the components more affected by bias are those in which $\mathbf{v}_k^\top \Psi \mathbf{f}$ is significantly larger than $\mathbf{v}_k^\top \epsilon$. Because of this, since \mathbf{f} is a smooth function, and ϵ is an uncorrelated noise, the bias is mostly associated with less oscillating eigenvectors \mathbf{v}_k , whilst it is instead negligible in components associated with highly oscillating eigenvectors \mathbf{v}_k . At this point, it is important to notice that the eigenvectors \mathbf{v}_k can indeed be distinguished by their oscillation frequencies.

In particular, in the simpler case of a 1-dimensional smoothing problem, with penalization of the second derivative and splines discretization, Demmler and Reinsch [1975] prove that the eigenvectors of the smoothing matrix, which plays the same role as the matrix B in equation (6), discretize the sinusoidal functions described by Grebenkov and Nguyen [2013]. Such eigenvectors thus have an increasing oscillating frequency, that is directly proportional to the order of the eigenvector k . In the case of the considered SR-PDE models, it is unfortunately not possible to give an equivalent analytical characterization of the increasing oscillatory behavior of the eigenvectors of B in full generality, namely for general forms of the domain Ω and of regularizing terms involving elliptic PDEs, and for general discretizations based on finite elements. However, such periodic behavior is always observed in practice, and can be proved analytically for specific choices of the domain, of the regularizing term, and of the finite element discretization, as shown in the following proposition.

Proposition 2. Let Ω be a squared domain, discretized by a uniform mesh. Consider data locations coinciding with the mesh nodes and an SR-PDE model with regularizing term $\int_{\Omega} (\Delta f)^2$. Then, the eigenvalues of B are monotonically non-increasing and the corresponding eigenvectors are increasingly oscillatory. Specifically, the eigenvalue-eigenvector couples of B converge to the couples $(d_{i,j} = \frac{1}{1+\lambda(i^2+j^2)}, \cos(ip_1) \cos(jp_2))$ for $i, j \in \mathbb{N}$, that are the eigenvalue-eigenfunctions couples of the Laplace eigenvalue problem.

The proof is reported in the Section 1.1 of the Supporting Information, and relies on results on the finite element discretization of the Laplace eigenvalue-eigenfunction problem [Boffi, 2010; Evans, 1998; Grebenkov and Nguyen, 2013]. The first row of Figure 2 reports some eigenvectors of B in the setting of Proposition 2, highlighting their increasing oscillatory nature. The second row shows that the same behaviour is observed in practice also for domains with different shapes. Thanks to this oscillatory behaviour, the bias is concentrated on a few components $\mathbf{T}_{\beta, \Pi, k}$, corresponding to projections of $\Psi \mathbf{f}$ on the low frequency eigenvectors \mathbf{v}_k .

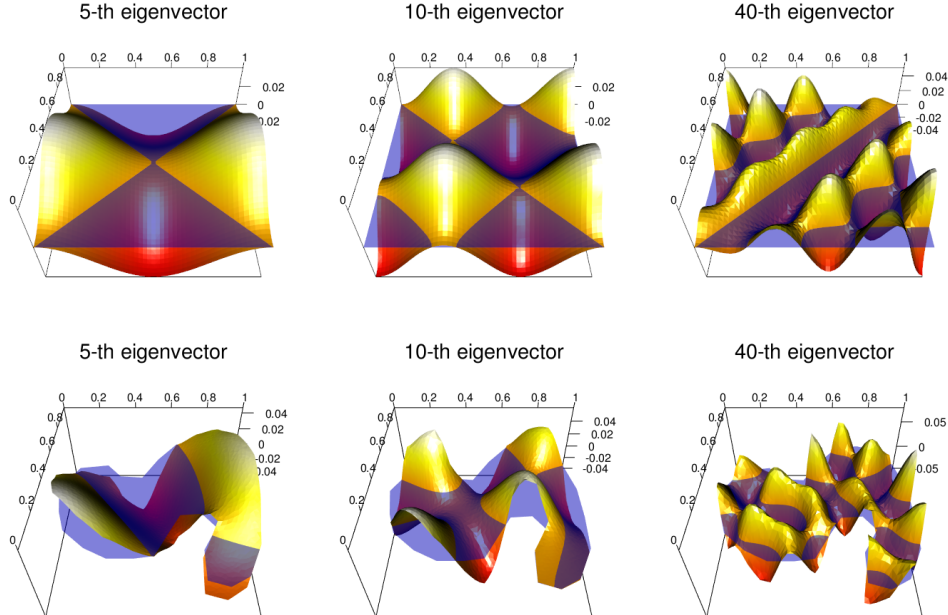


Figure 2: From left to right: 3D representations of the 5th, 10th, 40th eigenvectors of the smoothing matrix B , for Laplace regularization with regular finite element mesh and data locations at mesh nodes on a square domain (top) and on an irregular domain (bottom). In each panel, the blue transparent plane represents the zero-level. The image highlights that the eigenvectors have an increasingly oscillatory behavior along both the axes.

3.2 Partial ESF on the linear component

Our objective is to develop a more powerful inference approach on β , by defining a new sign-flip statistic, that is less affected by the impact of bias than the ESF statistic in (8). To this end, we exploit the fact that the bias \mathbf{b}_λ of $\mathbf{T}_{\beta,\Pi}$ can be decomposed as the sum of the biases $\mathbf{b}_{\lambda,k}$ specific to the components $\mathbf{T}_{\beta,\Pi,k}$ of $\mathbf{T}_{\beta,\Pi}$, as detailed in Proposition 1. The idea is then to define a new sign-flip statistic, where we do not sign-flip the components associated with the highest biases, imposing the corresponding sign-flip to $\pi_k = 1$. As described in this section, this partial sign-flip approach has the benefit of mitigating the bias effect, reducing the extra-variability induced by the bias of the observed test statistic, while retaining the test's information. Moreover, we show that the corresponding test remains asymptotically exact. In particular, Proposition 3 demonstrates that the test based on a ESF statistic is asymptotically exact, even when only a portion of the components is sign-flipped, for any given choice of the set of components to retain fixed.

Proposition 3. For a set of indices $J \subset \{1, \dots, n\}$, let Π_J be a diagonal matrix with entries $\pi_{kk} = 1$ if $k \in J$, and $\pi_{kk} \sim U\{-1, 1\}$ otherwise. Consider the Eigen-Sign-Flip statistic defined replacing Π by Π_J . Consider the test that rejects H_0 if and only if $\mathbf{T}_{\beta,I} > \mathbf{T}_{\beta,[1-\alpha]}$, i.e., if and only if the rank of $\mathbf{T}_{\beta,I}$ is at least $1 - \alpha$. Then, under the null hypothesis, the test is asymptotically exact and the rejection probability $\mathbb{P}(\mathbf{T}_{\beta,I} > \mathbf{T}_{\beta,[1-\alpha]})$ is at most α .

The proof is reported in Section 1.1 of the Supporting Information. Notice that, when the sample size n is too small, fixing many components may reduce the range of significance levels a user may be able to test. However, in practical applications, this effect is not relevant.

To understand which components of the test statistic should not be sign-flipped, we now focus on estimating the amount of bias present in each component. Consider the decomposition (11) presented in Proposition 1. An estimator $\hat{\mathbf{b}}_\lambda$ for the bias can be obtained estimating the term $\mathbf{v}_k^\top \Psi \mathbf{f}$ by $\hat{\alpha}_k = \mathbf{v}_k^\top (\mathbf{y} - X\hat{\beta})$, obtaining:

$$\hat{\mathbf{b}}_\lambda = n^{-\frac{1}{2}} \sum_{k=1}^n \hat{\mathbf{b}}_{\lambda,k} = n^{-\frac{1}{2}} \sum_{k=1}^n (X^\top \mathbf{v}_k)(d_k \pi_k) \hat{\alpha}_k.$$

We are interested in fixing the components that have a high relative bias, with respect to the test statistic $\mathbf{T}_{\beta,\Pi,k}$. We quantify this relative bias by $|\hat{\alpha}_k|$ and fix the components with $|\hat{\alpha}_k| > \gamma \cdot \hat{\sigma}^2$, where $\hat{\sigma}^2$ is the estimator of the noise variance σ^2 . The constant γ regulates the number of fixed components, and we fix $\gamma = 10$ by empirical validation. This threshold allows to correct the effect of bias, when present, at a cost of a minimal loss of sign-flips (i.e., few components fixed), whilst no sensible modification is induced when low or null bias is present (e.g., in large sample scenario). This leads to the following definition.

Definition 1 (Partial Eigen-Sign-Flip test on β). Let $J_{\text{PESF}} \subset \{1, \dots, n\}$ be the set of indices such that $|\hat{\alpha}_k| > \gamma \cdot \hat{\sigma}^2$, and let $\Pi_{J_{\text{PESF}}}$ be the associated partial sign-flip matrix. The Partial Eigen-Sign-Flip statistic \mathbf{T}_{Π_J} is defined as:

$$\mathbf{T}_{\beta,\Pi_{J_{\text{PESF}}}} = n^{-\frac{1}{2}} X^\top V \Pi_{J_{\text{PESF}}} V^\top \Lambda (\mathbf{y} - X\beta_0). \quad (13)$$

The Partial Eigen-Sign-Flip (PESF) test on β , for the set of hypotheses in (4), rejects the null hypothesis if and only if $\mathbf{T}_{\beta,I} > \mathbf{T}_{\beta,[1-\alpha]_{J_{\text{PESF}}}}$, i.e., if the rank of $\mathbf{T}_{\beta,I}$ is at least $1 - \alpha$.

For the convenience of the reader, we summarize the main statistics considered in Table 1. Proposition 3 ensures that the Partial Eigen-Sign-Flip test is asymptotically exact. In the bottom panel of Figure 1, we plot the distribution of the sign-flipped residuals $\Pi_{J_{\text{PESF}}} \tilde{\mathbf{r}}$, where $\Pi_{J_{\text{PESF}}}$ is a partial sign-flip matrix built accordingly to PESF procedure. The distribution of the partial sign-flipped residuals $\Pi_{J_{\text{PESF}}} \tilde{\mathbf{r}}$ has similar center and variance as the distribution of $\tilde{\mathbf{r}}$, contrary to the distribution of $\Pi \tilde{\mathbf{r}}$, that instead displays a larger variance.

Remark 2. The p-value for the proposed partial ESF two sided hypothesis test is computed as twice the minimum of the one-sided p-values. This approach is followed because the distribution of $\mathbf{T}_{\beta,\Pi}$ is not zero centered, and simply considering the absolute value of the test statistic $\mathbf{T}_{\beta,I}$, as done in Ferraccioli et al. [2022] would assign a different weight to the two tails of the score statistic distribution. This would marginally affect the power of the test, depending on the sign and magnitude of the bias \mathbf{b}_λ .

Remark 3. Confidence Intervals (CI) for β may be computed employing the test statistics here proposed. A $(1 - \alpha)$ -level CI can be always defined as the set of parameters for which the corresponding null hypothesis is not rejected at level α . Despite the simplicity of the theoretical definition, the computation of a CI can be demanding. Section 2 of the Supporting Information presents an iterative procedure that retrieves the set of points in the CI, avoiding the full exploration of the possible values of β .

Remark 4. It is worth to point out that, proceeding as in Section 4.4 of Ferraccioli et al. [2023], we may also test subsets of covariates. In principle, this is equivalent to treat the covariates whose coefficients are not targeted by the test as nuisance parameters, in analogy to what is done for the spatial field f . For the sake of simplicity, in this work we thus present tests on all the coefficients β simultaneously. We refer to Ferraccioli et al. [2022] for further details.

4 Inference on the nonlinear component

Inference on the nonlinear term in nonparametric or semiparametric regression has been addressed for more classical models based on splines and extensions [see, e.g., Harezlak et al., 2018; Wood, 2017, for a comprehensive overview of the existing methods]. In this section, we extend some of these approaches to the non-linear component of SR-PDE. In particular, we describe a parametric method that draws inspiration from Wald-like approach in Wood [2013], which deals with testing smooth components of generalized additive models. Furthermore, we extend to the non-linear component f the nonparametric approaches based on the sign-flip approach previously presented for the tests on β .

Our interest lies in testing the equality of the function f to a smooth function γf_0 , where f_0 is known, whilst the scaling factor γ may be unknown. The test is performed at a generic set of locations $\mathcal{Z} = \{\mathbf{z}_1, \dots, \mathbf{z}_{n_{\mathcal{Z}}}\} \subset \Omega$. We thus consider the following set of hypothesis:

$$H_0 : f(\mathbf{z}) = \gamma f_0(\mathbf{z}) \quad \forall \mathbf{z} \in \mathcal{Z} \quad \text{vs} \quad H_1 : \exists \mathbf{z} \in \mathcal{Z} \mid f(\mathbf{z}) \neq \gamma f_0(\mathbf{z}). \quad (14)$$

For simplicity, we assume that the function f_0 lives in the considered space of finite elements functions. This assumption is not restrictive, since this space is chosen rich enough to provide accurate approximations of smooth functions. Moreover, also in absence of such assumption, the results presented in the following sections hold true at the mesh nodes. Notice that, to test the equality of the function f to a function γf_0 over all the domain, it suffices to set $\mathcal{Z} = \{\boldsymbol{\xi}_j\}_{j \in 1, \dots, N}$, where $\boldsymbol{\xi}_j$ is the j th mesh node, since a finite element function is fully determined by its values of the mesh nodes. Moreover, thanks the presence of the possibly unknown scaling factor γ , the set of hypothesis (14) is rather general. For instance, we can test the significance of the nonparametric term of the SR-PDE model, by setting f_0 to a constant function.

4.1 Wald inference on f

Wald inference on f relies on the asymptotic distribution of the corresponding estimator $\hat{\mathbf{f}}$ in (3). Under some regularity assumptions, Theorem 1 in Arnone et al. [2023] shows that this estimator has asymptotic normal distribution

$$(\hat{\mathbf{f}} - \mathbf{f})|X \sim \mathcal{N}_N(\mathbf{0}, V_A),$$

where V_A is the $N \times N$ asymptotic variance given in the same theorem of Arnone et al. [2023] and $\mathbf{0}$ is the vector having N elements equal to 0. This justifies Wald type inference on f , as detailed in the following section. Notice that, similarly to what happens for the inference on β , the asymptotic variance V_A does not involve the smoothing parameter λ , which instead should be accounted for in finite sample scenarios. Hence, in practice, we employ the finite dimensional variance of $\hat{\mathbf{f}}$, given by

$$V_{\hat{\mathbf{f}}} = (\Psi^\top Q \Psi + \lambda P)^{-1} \Psi^\top Q \Psi (\Psi^\top Q \Psi + \lambda P)^{-1}$$

that asymptotically converges to V_A . We denote by $\Psi_{\mathcal{Z}}$ the matrix of basis evaluations at the testing locations \mathcal{Z} . For $\mathcal{Z} = \mathcal{P}$ the matrix $\Psi_{\mathcal{P}}$ coincides with the matrix Ψ defined in Section 2.1. We moreover denote by $\hat{\mathbf{f}}_{\mathcal{Z}}$ the estimator of f at the testing locations \mathcal{Z} , i.e., $\hat{\mathbf{f}}_{\mathcal{Z}} = \Psi_{\mathcal{Z}} \hat{\mathbf{f}}$. Finally, to lighten the notation, we let \mathbf{f}_0 be the vector of evaluations of f_0 at the testing locations \mathcal{Z} , i.e., $\mathbf{f}_0 = \mathbf{f}_{0, \mathcal{Z}}$. We now detail the Wald test, starting from the simplest case of known scaling factor γ .

Test (14) with known γ

Let $\gamma = \gamma_0$ be known and, without loss of generality, set $\gamma_0 = 1$. The Wald test statistic, under the null hypothesis H_0 in (14) with $\gamma = \gamma_0 = 1$, follows asymptotically the χ^2 distribution

$$(\hat{\mathbf{f}}_{\mathcal{Z}} - \mathbf{f}_0)^\top V_{\text{Wald}}^{-r} (\hat{\mathbf{f}}_{\mathcal{Z}} - \mathbf{f}_0) \sim \chi_r^2,$$

where V_{Wald}^{-r} indicates a rank- r pseudo-inverse of the variance matrix $V_{\text{Wald}} = \Psi_{\mathcal{Z}} V_{\hat{\mathbf{f}}} \Psi_{\mathcal{Z}}^\top$. Notice that $\text{rank}(V_{\text{Wald}}) = \min\{n_{\mathcal{Z}}, \text{rank}(V_{\hat{\mathbf{f}}})\}$, and V_{Wald} may be singular. Even when V_{Wald} is invertible, it might have eigenvalues close to zero. Thus the use of a rank- r pseudo-inverse is always recommendable, as commented in a similar context by Wood [2013].

Test (14) with unknown γ

We proceed by estimating γ . In particular, Theorem 6 in Arnone et al. [2022] proves the consistency of the estimator $\hat{\gamma}$, under mild regularity assumptions. Therefore, we can provide an estimator $\hat{\gamma}$ that converges in probability to γ . Relying on this result, we can prove the following statement.

Proposition 4. Under the null hypothesis H_0 in (14), with unknown γ , we have: $\hat{\mathbf{f}}_{\mathcal{Z}}|X \sim \mathcal{N}_N(\gamma \mathbf{f}_0, V_A)$.

The proof of Proposition 4 is reported in Section 1.2 of the Supporting Information. Leveraging on this results, we can define an asymptotically exact test, following the same procedure detailed for the case of known γ .

4.2 Score statistic for f

As we show by the simulations studies in Section 5.2, the parametric test in Section 4.1 may suffer a poor control of Type-I error, due to the presence of the bias induced by the regularizing term. For this reason, we here propose two nonparametric alternatives, based on the sign-flip approach. This approach, likewise the test for β , is based on the score statistic, and avoids the direct estimation of the Fisher information matrix for the calculation of p-values, in order to test the hypothesis (14), for both known and unknown γ . The score statistic is defined under the null hypothesis (14) and is not affected by the regularization of the model, inducing tests that enjoy good control of Type-I error as well as high power. For simplicity, we here focus on the case where the testing locations are a subset of the data locations, i.e., $\mathcal{Z} \subseteq \mathcal{P}$. Such assumption is not restrictive and permits to test equality of f to any function γf_0 over the whole domain, as long as $\Psi_{\mathcal{Z}}^{\top} \Psi_{\mathcal{Z}}$ is invertible.

Let $\mathbf{y}_{\mathcal{Z}}$, $X_{\mathcal{Z}}$ and $Q_{\mathcal{Z}}$ denote the restriction of \mathbf{y} , X , and Q to the $n_{\mathcal{Z}}$ locations in $\mathcal{Z} \subseteq \mathcal{P}$, and let $\hat{\beta}_{\mathcal{Z}, H_0} = (X_{\mathcal{Z}}^{\top} X_{\mathcal{Z}})^{-1} X_{\mathcal{Z}}^{\top} (\mathbf{y}_{\mathcal{Z}} - \Psi_{\mathcal{Z}} \mathbf{f}_0)$ be the corresponding estimator for β under H_0 in (14). Starting from the score equation of the model (1) for f , we can define the following score statistic for f :

$$\begin{aligned} \mathbf{T}_f &= n_{\mathcal{Z}}^{-1/2} \Psi_{\mathcal{Z}}^{\top} (\mathbf{y}_{\mathcal{Z}} - X_{\mathcal{Z}} \hat{\beta}_{\mathcal{Z}, H_0} - \Psi_{\mathcal{Z}} \mathbf{f}_0) = n_{\mathcal{Z}}^{-1/2} \Psi_{\mathcal{Z}}^{\top} (\mathbf{y}_{\mathcal{Z}} - X_{\mathcal{Z}} (X_{\mathcal{Z}}^{\top} X_{\mathcal{Z}})^{-1} X_{\mathcal{Z}}^{\top} (\mathbf{y}_{\mathcal{Z}} - \Psi_{\mathcal{Z}} \mathbf{f}_0) - \Psi_{\mathcal{Z}} \mathbf{f}_0) = \\ &= n_{\mathcal{Z}}^{-1/2} \Psi_{\mathcal{Z}}^{\top} ((I - X_{\mathcal{Z}} (X_{\mathcal{Z}}^{\top} X_{\mathcal{Z}})^{-1} X_{\mathcal{Z}}^{\top}) (\mathbf{y}_{\mathcal{Z}} - \Psi_{\mathcal{Z}} \mathbf{f}_0)) = n_{\mathcal{Z}}^{-1/2} \Psi_{\mathcal{Z}}^{\top} Q_{\mathcal{Z}} (\mathbf{y}_{\mathcal{Z}} - \Psi_{\mathcal{Z}} \mathbf{f}_0) = n_{\mathcal{Z}}^{-1/2} \Psi_{\mathcal{Z}}^{\top} Q_{\mathcal{Z}} \mathbf{r}_{\mathcal{Z}}, \end{aligned} \quad (15)$$

where $\mathbf{r}_{\mathcal{Z}} = (\mathbf{y}_{\mathcal{Z}} - \Psi_{\mathcal{Z}} \mathbf{f}_0)$. The statistic \mathbf{T}_f is unbiased, since $\mathbb{E}(\mathbf{T}_f) = \mathbf{0}$ due to the fact that $Q_{\mathcal{Z}} X_{\mathcal{Z}} = \mathbf{0}$. Let $V_{\mathbf{T}_f} = \sigma^2 \Psi_{\mathcal{Z}}^{\top} Q_{\mathcal{Z}} \Psi_{\mathcal{Z}}$. The following theorem states that the statistic \mathbf{T}_f has asymptotic normal distribution.

Theorem 1. Under H_0 in (14) the score statistic \mathbf{T}_f defined in (15) is asymptotically normally distributed, with: $\sqrt{n_{\mathcal{Z}}} V_{\mathbf{T}_f}^{-1/2} \mathbf{T}_f \sim \mathcal{N}(\mathbf{0}, I_N)$.

The proof is reported in Section 1.2 of the Supporting Information. It is worth noting that the score statistic is a N -dimensional vector, and each of its components is a weighted sum of the $n_{\mathcal{Z}}$ observed residuals evaluated under the null hypothesis, namely:

$$\mathbf{T}_f = n_{\mathcal{Z}}^{-1/2} \Psi_{\mathcal{Z}}^{\top} Q_{\mathcal{Z}} \mathbf{r}_{\mathcal{Z}} = \begin{bmatrix} n_{\mathcal{Z}}^{-1/2} \sum_{i=1}^{n_{\mathcal{Z}}} \psi_1(\mathbf{z}_i) (Q_{\mathcal{Z}} \mathbf{r}_{\mathcal{Z}})_i \\ \vdots \\ n_{\mathcal{Z}}^{-1/2} \sum_{i=1}^{n_{\mathcal{Z}}} \psi_N(\mathbf{z}_i) (Q_{\mathcal{Z}} \mathbf{r}_{\mathcal{Z}})_i \end{bmatrix} \quad (16)$$

where $(Q_{\mathcal{Z}} \mathbf{r}_{\mathcal{Z}})_i$ indicates the i th element of the vector $Q_{\mathcal{Z}} \mathbf{r}_{\mathcal{Z}}$. This expression highlights that the j th entry of \mathbf{T}_f can be interpreted as the estimate of the residual at the j th mesh node under the null hypothesis.

4.3 Sign-Flip test

Expression (16) highlights that each entry of the score statistic \mathbf{T}_f can be seen as a weighted sum of $n_{\mathcal{Z}}$ independent contributions, that are zero centered under the null hypothesis. We can use this information to derive the null distribution of each entry of the score statistic. Following the sign-flip approach, we can define the following Sign-Flip test on f , starting from the score statistic \mathbf{T}_f .

Definition 2 (Sign-Flip test on f). Let $\Pi = \text{diag}(\pi_1, \dots, \pi_{n_{\mathcal{Z}}})$ be a sign-flip matrix, where $\boldsymbol{\pi} = (\pi_1, \dots, \pi_{n_{\mathcal{Z}}})$ is a random vector uniformly distributed in $\{-1, 1\}^{n_{\mathcal{Z}}}$. The Sign-Flip score statistic is defined as:

$$\mathbf{T}_{f, \Pi} = n_{\mathcal{Z}}^{-1/2} \Psi_{\mathcal{Z}}^{\top} \Pi Q_{\mathcal{Z}} \mathbf{r}_{\mathcal{Z}}. \quad (17)$$

The observed test statistic $\mathbf{T}_f = \mathbf{T}_{f, I}$ corresponds to the case where $\pi_i = 1$, for all $i \in 1, \dots, n_{\mathcal{Z}}$. Let $\mathbf{T}_{f, \Pi}^{\text{comb}}$ be the sign-flip combined statistic. Then the p-value for the test on (14) is computed as the rank of $\mathbf{T}_{f, I}^{\text{comb}}$ with respect to a sample of M sign-flips $\boldsymbol{\pi}$, divided by M .

Different combination methods can be considered [Pesarin, 2001]. In this work we consider the Mahalanobis distance, that is $\mathbf{T}_{\Pi}^{\text{comb}} = n_{\mathcal{Z}}^{-1} \mathbf{T}_{f, \Pi}^{\top} V_{\mathbf{T}_f}^{-1} \mathbf{T}_{f, \Pi}$. The following propositions gives the mean and variance of the sign-flip statistic $\mathbf{T}_{f, \Pi}$ under the null hypothesis and state the asymptotic exactness of the Sign-Flip test in Definition 2.

Proposition 5. For a given sign-flip matrix Π , the expected value and variance of the score test statistic under H_0 in (14) are:

$$\begin{aligned} \mathbb{E}_{H_0}[\mathbf{T}_{f, \Pi}] &= \mathbf{0} \\ \text{Var}_{H_0}[\mathbf{T}_{f, \Pi}] &= n_{\mathcal{Z}}^{-1} \sigma^2 \Psi_{\mathcal{Z}}^{\top} \Pi Q_{\mathcal{Z}} \Pi \Psi_{\mathcal{Z}}. \end{aligned}$$

Proposition 6. The Sign-Flip test on f defined in Definition 2 is asymptotically exact.

The proofs of Propositions 5 and 6 are reported in Section 1.2 of the Supporting Information. Notice that the Sign-Flip test on f is only asymptotically exact, since the distribution of the Sign-Flip test statistic $\mathbf{T}_{f, \Pi}$ depends

on a specific matrix Π . In this test we are applying the sign-flip transformation to the score contributions, that under H_0 are symmetric around zero but not independent. In fact, we have that $\mathbb{E}_{H_0}[Q_Z \mathbf{r}_Z] = \mathbf{0}$ and $\text{Var}_{H_0}[Q_Z \mathbf{r}_Z] = \sigma^2 Q_Z$. In general, the projection matrix Q_Z is not diagonal and hence it induces correlation. However, for n_Z large enough, the projection matrix Q_Z is approximately an identity matrix and, under the null hypothesis, the distribution of $Q_Z \mathbf{r}_Z$ is approximately a multivariate normal, leading to an asymptotically exact test, as stated by Proposition 6. In the next section we instead provide a modified version of the Sign-Flip test, that is exact also in the finite sample scenarios.

Remark 5. Notice that, if the errors are assumed to be normally distributed, and no covariates are present, then the Sign-Flip tests is exact. Indeed, in this case, the projector Q_Z is an identity also in the finite sample case.

Remark 6. It is relevant to highlight the particular setting where $Z \subseteq \{\xi_j\}_{j \in 1, \dots, N}$, where ξ_j are the mesh nodes. In this case, many of the common combinations that are usually employed to produce the global statistic, among which there is the Mahalanobis distance, may not be appropriate for the sign-flip procedure here considered. This happens because Ψ_Z is a subset of the identity matrix I , and this eliminates the effect of the sign-flip transformation. However, this issue can be solved by combining the contributions of nearby nodes, as detailed in Section 3 of the Supporting Information.

4.4 Eigen-Sign-Flip for f

In this section we propose a modification of the Sign-Flip test on f described in Section 4.3, in order to correct for the lack of invariance under H_0 of the covariance of the sign-flip statistic $\mathbf{T}_{f,\Pi}$. Notice that, differently to what happens for the Eigen-Sign-Flip test on β , this test is exact also in finite-sample scenarios. Therefore, there is no need to provide a bias reduction procedure as the one discussed in Section 3.2. The new test, named Eigen-Sign-Flip test for f , is defined as follows.

Definition 3. (Eigen-Sign-Flip test on f) Let us consider the spectral decomposition $Q_Z = VV^\top$, where V is the $n_Z \times (n_Z - q)$ matrix of eigenvectors corresponding to the nonzero eigenvalues of Q_Z . Let $\Pi = \text{diag}(\pi_1, \dots, \pi_{n_Z - q})$ be a sign-flip matrix, with $\boldsymbol{\pi} = (\pi_1, \dots, \pi_{(n_Z - q)})$ a random vector uniformly distributed in $\{-1, 1\}^{(n_Z - q)}$. Then the Eigen-Sign-Flip score statistic is defined as:

$$\mathbf{T}_{f,\Pi}^E = n_Z^{-1/2} \Psi_Z^\top V \Pi V^\top Q_Z \mathbf{r}_Z = n_Z^{-1/2} \tilde{\Psi}_Z^\top \Pi \tilde{\mathbf{r}}_Z \quad (18)$$

where $\tilde{\Psi}_Z = V^\top \Psi_Z$ and $\tilde{\mathbf{r}}_Z = V^\top \mathbf{r}_Z$. The observed statistic $\mathbf{T}_f^E = \mathbf{T}_{f,I}^E \in \mathbb{R}^N$ corresponds to the case where $\pi_i = 1, i \in 1, \dots, n_\dagger$. Let $\mathbf{T}_\Pi^{\text{comb}}$ be the sign-flip combined global statistic. Then the p-value is computed as the rank of $\mathbf{T}_f^{\text{comb}}$ with respect to a sample of M sign-flips $\boldsymbol{\pi}$, divided by M .

The main idea behind the Eigen-Sign-Flip test is to exploit the fact that Q_Z is the projection matrix onto the subspace orthogonal to $\text{span}(X_Z)$. Hence it is idempotent and, assuming that X_Z has full rank q , it has exactly $n_Z - q$ eigenvalues equal to 1 and the remaining equal to 0. Then, as already proposed by Kherad-Pajouh and Renaud [2010] in the classical linear regression framework, we can premultiply the residuals by the matrix of eigenvectors corresponding to the non-zero eigenvalues. This procedure allows to reduce the cardinality of the residuals and makes the resulting $n_Z - q$ pseudo-residuals uncorrelated. The properties of the Eigen-Sign-Flip test statistic are summarized in the following propositions.

Proposition 7. For a given sign-flip matrix Π , under H_0 in (14), the expected value and variance of the score test statistic $\mathbf{T}_{f,\Pi}^E$ are:

$$\begin{aligned} \mathbb{E}_{H_0}[\mathbf{T}_{f,\Pi}^E] &= \mathbf{0} \\ \text{Var}_{H_0}[\mathbf{T}_{f,\Pi}^E] &= n_Z^{-1} \sigma^2 \Psi_Z^\top Q_Z \Psi_Z. \end{aligned}$$

Proposition 8. The Eigen-Sign-Flip test in Definition 3 is exact.

The proofs are similar to the one used in the Sign-Flip case, and are reported in Section 1.2 of the Supporting Information. As shown by the simulations in Section 5, the Sign-Flip and Eigen-Sign-Flip tests on f are more robust than the parametric Wald test, and offer a very good control of Type-I error, differently from the parametric test. We can also construct CI for $f(\mathbf{p})$ based on the Sign-Flip and Eigen-Sign-Flip tests, analogously to the procedure detailed for β in Section 2 of the Supporting Information.

5 Simulation studies

In this section we assess the validity of the proposed inference methods by means of simulations. Simulation 1 focuses on the linear component of the model β , whilst Simulation 2 focuses instead on the nonlinear component f .

The library `fdaPDE` [Arnone et al., 2024] provides an exact and an approximated implementation for all the considered approaches. In particular, all the tests involve the inversion of large sparse matrices, that, when performed exactly, yields large dense matrices, heavy to compute and difficult to store. The approximated inversion is computed by adapting the Factorized Sparse Parallel Approximate Inverse (FSPAI) algorithm, proposed by Huckle [2003] to our modeling framework. The FSPAI algorithm approximates the dense inverse of a matrix with a sparse matrix, resulting in a significant saving of time and memory.

5.1 Simulation 1: inference on the linear component

We present a simulation study that investigates the finite sample performances of the Partial Eigen-Sign-Flip test (PESF) for β proposed in Section 3.2. We consider the following competing methods:

- Wald: the classic Wald test based on the asymptotic distribution of $\hat{\beta}$;
- Speck: the variant of Wald test based on the asymptotic distribution of the Speckman version of $\hat{\beta}$ [Holland, 2017; Speckman, 1988];
- ESF: the Eigen-Sign-Flip score test introduced in Ferraccioli et al. [2022];
- PESF: the Partial Eigen-Sign-Flip test proposed in Section 3.2.

We simulate from model (1), with $\Omega = [0, 1] \times [0, 1]$ and $\mathbf{p}_1, \dots, \mathbf{p}_n$ randomly sampled from a uniform distribution on Ω , with $n = 225$. For the nonparametric component of the model, we consider the test function 2 from the function set `gamSim` in the R package `mgcv` [Wood, 2017], defined in Section 4 of the Supporting Information. We consider one covariate, and we generate $\mathbf{x}_1, \dots, \mathbf{x}_n$ according to four different stochastic processes:

- a) a Gaussian random field with zero mean and scale 0.05;
- b) a Matérn random field with $\nu = 1, \sigma = 2$ and scale 0.1;
- c) the function $\cos(5(\mathbf{p}_1 + \mathbf{p}_2)) + (2\mathbf{p}_1 - \mathbf{p}_1\mathbf{p}_2^2)^2$ with an added Gaussian random field with scale 0.05;
- d) the function $\cos(5(\mathbf{p}_1 + \mathbf{p}_2)) + (2\mathbf{p}_1 - \mathbf{p}_1\mathbf{p}_2^2)^2$ with an added Matérn random field with $\nu = 1, \sigma = 2$ and scale 0.1.

The covariate and the true f are standardized, before computation of the response variable y , so that their relative contributions to the response are comparable. We consider $H_0: \beta_0 = 0$. We generate data from model (1), both with $\beta = 0$ and with other 10 different values of β , from 0.01 to 0.1, in order to check Type-I error, as well as the power of the test. Finally, we add i.i.d. random errors $\epsilon_1, \dots, \epsilon_n$, with zero mean and standard deviation 0.1. For each test case, the generation of the covariates and noise is repeated 1000 times. For each simulation repetition, we obtain the estimate using SR-PDE with penalization of the Laplacian, employing linear finite elements on a uniform mesh with 283 nodes and smoothing parameter λ selected by generalized-cross-validation. The test are performed with nominal value 0.05. For ESF and PESF tests, we consider 1000 random sign-flips. Moreover, we show results with both exact and approximated FSPAI inference, with FSPAI tolerance $\epsilon_{FSPAI} = n \cdot 10^{-4}$. The results for the four cases (a)-(d) are shown in Table 2. Figure 3 shows the power curves for the most interesting cases (b) and (d), where the spatial correlation in the covariates induces more bias in the estimators. Cases (a) and (c) are reported in Section 4 of the Supporting Information. Both parametric alternatives, Wald and Speckman, have poor performances, with very poor control of Type-I error in all scenarios. On the converse, the nonparametric tests have an excellent control of Type-I error, even in the most challenging scenarios. The PESF test here proposed has systematically higher power with respect to the standard ESF in Ferraccioli et al. [2022]. Moreover, the proposed PESF test is robust to the FSPAI approximation, whilst ESF test may lose power, especially in the most challenging scenario (d).

5.2 Simulation 2: inference on the nonlinear component

We now show some simulation studies to assess the performances of the tests proposed for f . In particular we consider three different ways to move away from the null hypothesis in (14), generating data under the following alternative hypothesis H_1 :

- e) **Scale.** We multiply f_0 by a scale factor κ : $f(\mathbf{p}) = \kappa f_0(\mathbf{p})$, $\kappa \in \mathbb{R}$.
- f) **Shift.** We add to f_0 a shift factor c : $f(\mathbf{p}) = f_0(\mathbf{p}) + c$, $c \in \mathbb{R}$.
- g) **Linear combination.** We consider an alternative function f_1 defined in Section 5 of the Supporting Information, and we take a linear convex combination of f_0 and f_1 , with weight α : $f(\mathbf{p}) = (1 - \alpha)f_0(\mathbf{p}) + \alpha f_1(\mathbf{p})$, $0 \leq \alpha \leq 1$.

Type-I error

Exact	(a)	(b)	(c)	(d)
Wald	0.109	0.219	0.109	0.223
Speck	0.096	0.100	0.093	0.101
ESF	0.043	0.046	0.041	0.039
PESF	0.048	0.051	0.047	0.049
FSPA1	(a)	(b)	(c)	(d)
Wald	0.170	0.438	0.321	0.515
Speck	0.092	0.104	0.090	0.111
ESF	0.045	0.046	0.044	0.043
PESF	0.046	0.057	0.043	0.052

Table 2: Simulation 1, inference on β for the four simulated settings (a)-(d) and the different methods: Wald (Wald), Speckman (Speck), Eigen-Sign-Flip (ESF) from Ferraccioli et al. [2022] and Partial Eigen-Sign-Flip (PESF) proposed in Section 3. The estimates that are lower of the threshold of 0.0635 (i.e., that are inside of the confidence interval for the Type-I error for a nominal level of 0.05 with 1000 replicates) are highlighted in bold. Top: exact computations are used for all methods. Bottom: FSPA1 approximation of S (Wald) and of Λ (Speck, ESF, PESF) matrices is used for all methods.

We compare the performances of the methods proposed in Section 4, namely the classic Wald test introduced in Section 4.1; the Sign-Flip (SF) test introduced in Section 4.3; and the Eigen-Sign-Flip (ESF) test introduced in Section 4.4.

We generate data as detailed in Section 5.1, case (a), with $\beta = 1$. To assess the power of the tests we also generate data under the alternative, considering 21 values for κ from 0.80 to 1.20 with step 0.02 for case (e), and 21 values for c and α from -0.20 to 0.20 with step 0.02 for cases (f) and (g). The function f_1 used in case (g) is defined in Section 5 as the supporting information file. The generation of the noise is repeated 1000 times. For both the SF and ESF tests, we consider 1000 random sign-flips.

Figure 4 reports the results obtained considering exact computations. Section 5 of the Supporting Information reports the results obtained for the stochastic approximated inference. In all scenarios, Wald test has an extremely poor control of Type-I error, around 50 %; instead the proposed nonparametric tests have a very good control of Type-I error. The SF and ESF tests do not display any appreciable difference in the power curves.

6 Discussion

In this work we have proposed accurate and powerful nonparametric inference approaches for inference on the linear as well as nonlinear terms of SR-PDE models. The simulation studies highlight that the proposed Partial Eigen-Sign-Flip test on the linear component has higher power with respect to the already existing nonparametric tests, while retaining the same control of Type-I error. Regarding the nonlinear component, both the proposed nonparametric tests have very good control of Type-I error in small sample scenarios, differently from the parametric Wald-type inference, that has instead extremely poor control of Type-I error. All the considered tests on f have high power. Efficient algorithms for the computation of confidence intervals on the linear and nonlinear part are also presented in the supporting information file.

All the results presented in this work could be generalized to a wider class of SR-PDE models, handling for instance spatio-temporal data [Arnone et al., 2019; Bernardi et al., 2017], as well as data coming from distributions different from the Gaussian [Wilhelm et al., 2016]. Finally, in this work we assume that the λ parameter is given and fixed. A very interesting future research direction consists in exploring the effects of the choice of the smoothing parameter λ on the inference, similarly to what is done for instance in Wood et al. [2016] to correct AIC index in the context of Generalized Additive Models. These developments will be object of dedicated future studies.

Acknowledgments

L.M. Sangalli acknowledges the PRIN research project CoEnv - Complex Environmental Data and Modeling (2022E3RY23), funded by the NextGenerationEU programme of the European Union and by the Italian Min-

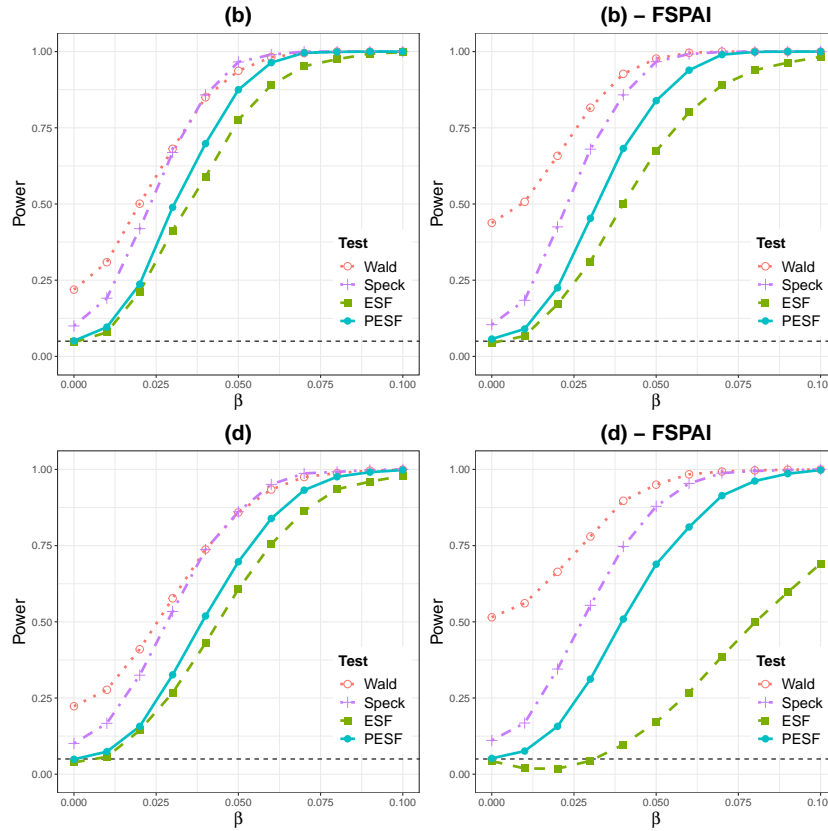


Figure 3: Simulation 1, inference on β . Rejection rates over 1000 repetitions, with covariates generated as in case (b) (top) and (d) (bottom), employing exact (left) and approximated (right) computations; Wald test (red dotted curve, with circle markers), Speckman (violet dash-and-dotted curve, with cross markers), standard ESF test (green dashed curve, with square markers) and the proposed PESF test (light blue solid curve, with circle markers).


istry for University and Research (MUR). M. Cavazzutti and L.M. Sangalli acknowledge MUR, Dipartimento d'Eccellenza 2022-2027, Dipartimento di Matematica, Politecnico di Milano.

CONFLICT OF INTEREST STATEMENT


The authors declare no potential conflicts of interest.


ORCID

Michele Cavazzutti , <https://orcid.org/0009-0007-7876-3357>

Eleonora Arnone , <https://orcid.org/0000-0002-8712-3489>

Federico Ferraccioli , <https://orcid.org/0000-0001-6832-988X>

Livio Finos , <https://orcid.org/0000-0003-3181-8078>

Laura M. Sangalli , <https://orcid.org/0000-0002-4951-9239>

Supporting information

Additional supporting information can be found online in the Supporting Information section at the end of this article.

References

E. Arnone, L. Azzimonti, F. Nobile, and L. M. Sangalli. Modeling spatially dependent functional data via regression with differential regularization. *Journal of Multivariate Analysis*, 170:275–295, 2019. doi: <https://doi.org/10.1016/j.jmva.2018.09.006>.

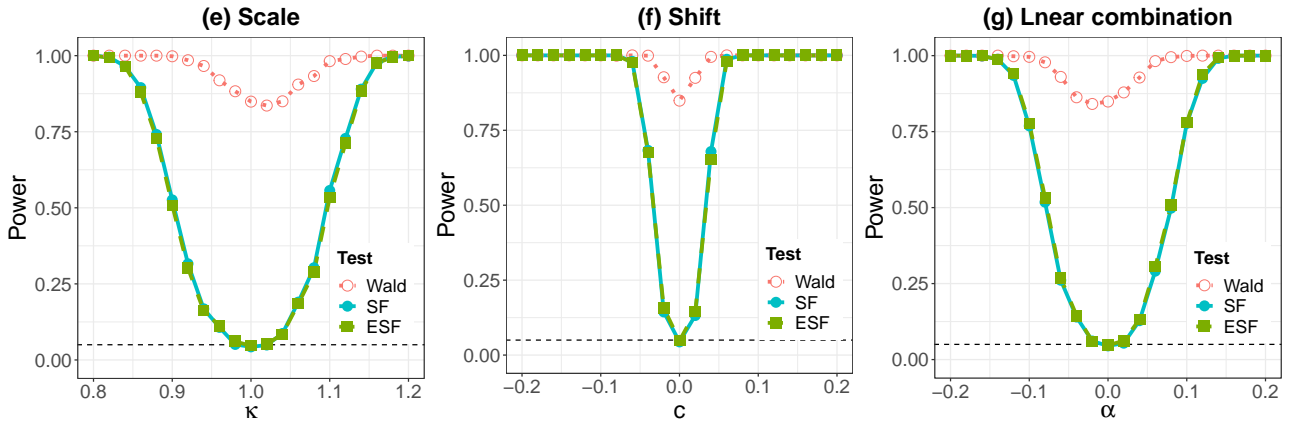


Figure 4: Simulation 2, inference on f . Rejection rates for 1000 repetitions, with exact inference computation; Wald test (red dotted curve, with circle markers), SF nonparametric test (blue solid curve, with circle markers) and ESF test (green dashed curve, with square markers).

- E. Arnone, A. Kneip, F. Nobile, and L. M. Sangalli. Some first results on the consistency of spatial regression with partial differential equation regularization. *Statistica Sinica*, 32(1):209–238, 2022. doi: <https://doi.org/10.5705/ss.202019.0346>.
- E. Arnone, L. Negri, F. Panzica, and L. M. Sangalli. Analyzing data in complicated 3d domains: Smoothing, semiparametric regression, and functional principal component analysis. *Biometrics*, 79(4):3510–3521, 2023. doi: <https://doi.org/10.1111/biom.13845>.
- E. Arnone, A. Clemente, L. M. Sangalli, E. Lila, J. Ramsay, and L. Formaggia. *fdaPDE: Physics-Informed Spatial and Functional Data Analysis*, 2024. URL <https://CRAN.R-project.org/package=fdaPDE>. R package version 1.1-18.
- L. Azzimonti, L. M. Sangalli, P. Secchi, M. Domanin, and F. Nobile. Blood flow velocity field estimation via spatial regression with pde penalization. *Journal of the American Statistical Association, Theory and Methods*, 110(511):1057–1071, 2015. doi: <https://doi.org/10.1080/01621459.2014.946036>.
- M. S. Bernardi, L. M. Sangalli, G. Mazza, and 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(1):23–38, 2017. doi: <https://doi.org/10.1007/s00477-016-1237-3>.
- P. J. Bickel, C. A. Klaassen, Y. Ritov, and J. A. Wellner. *Efficient and Adaptive Estimation for Semiparametric Models*. Springer New York, 1 edition, 1998.
- D. Boffi. Finite element approximation of eigenvalue problems. *Acta Numerica*, 19:1–120, 2010. doi: <https://doi.org/10.1017/S0962492910000012>.
- H. Chen, Y. Wang, R. Li, and K. Shear. A note on a nonparametric regression test through penalized splines. *Statistica Sinica*, 24:1143–1160, 2014. doi: <https://doi.org/10.5705/ss.2012.230>.
- G. Claeskens, T. Krivobokova, and O. J. D. Asymptotic properties of penalized spline estimators. *Biometrika*, 96(3):529–544, 2009. doi: <https://doi.org/10.1093/biomet/asp035>.
- D. Cox, E. Koh, G. Wahba, and B. S. Yandell. Testing the (parametric) null model hypothesis in (semi-parametric) partial and generalized spline models. *The Annals of Statistics*, 16(1):113–119, 1988. doi: <https://doi.org/10.1214/aos/1176350693>.
- C. Crainiceanu, D. Ruppert, G. Claeskens, and M. P. Wand. Exact likelihood ratio tests for penalised splines. *Biometrika*, 92(1):91–103, 2005. doi: <https://doi.org/10.1093/biomet/92.1.91>.
- A. Demmler and C. Reinsch. Oscillations matrices with spline smoothing. *Numerische Mathematik*, 24:375–382, 1975. doi: <https://doi.org/10.1007/BF01437406>.
- R. L. Eubank. *Nonparametric Regression and Spline Smoothing*. CRC press, 2 edition, 1999.
- L. C. Evans. *Partial Differential Equation*. American Mathematical Society, 1998.
- F. Ferraccioli, L. Finos, and L. M. Sangalli. Some first inferential tools for spatial regression with differential regularization. *Journal of Multivariate Analysis*, 189:104866, 2022. doi: <https://doi.org/10.1016/j.jmva.2021.104866>.

- F. Ferraccioli, L. M. Sangalli, and L. Finos. Nonparametric tests for semiparametric models. *Test*, 32:1106–1130, 2023. doi: <https://doi.org/10.1007/s11749-023-00868-9>.
- D. S. Grebenkov and B. T. Nguyen. Geometrical structure of laplacian eigenfunctions. *SIAM review*, 55(4): 601–667, 2013. doi: <https://doi.org/10.1137/120880173>.
- P. J. Green and B. W. Silverman. *Nonparametric Regression and Generalized Linear Models: A roughness penalty approach*. CRC Press, 1 edition, 1994.
- P. Hall and J. Horowitz. A simple bootstrap method for constructing nonparametric confidence bands for functions. *The Annals of Statistics*, 41(4):1892–1921, 2013. doi: <https://doi.org/10.1214/13-AOS1137>.
- W. Härdle, H. Liang, and J. Gao. *Partially Linear Model*. Physica, Heidelberg, 1 edition, 2000.
- D. Harezlak, D. Ruppert, and M. Wand. *Semiparametric Regression with R*. Springer, 1 edition, 2018.
- J. Hemerik, J. Goeman, and L. Finos. Robust testing in generalized linear models by sign-flipping score contributions. *Journal of the Royal Statistical Society Series B*, 82(3):841–864, 2020. doi: <https://doi.org/10.1111/rssb.12369>.
- Ø. Hjelle and M. Dæhlen. *Triangulations and Applications*. Springer, 1 edition, 2006.
- A. D. Holland. Penalized spline estimation in the partially linear model. *Journal of Multivariate Analysis*, 153: 211–235, 2017. doi: <https://doi.org/10.1016/j.jmva.2016.10.001>.
- T. Huckle. Factorized sparse approximate inverses for preconditioning and smoothing. *The Journal of Supercomputing*, 25:109–117, 2003. doi: <https://doi.org/10.1023/A:1023988426844>.
- M.-H. Huh and M. Jhun. Random permutation testing in multiple linear regression. *Communications in Statistics - Theory and Methods*, 30(10):2023, 2001. doi: <https://doi.org/10.1081/STA-100106060>.
- S. Kherad-Pajouh and O. Renaud. An exact permutation method for testing any effect in balanced and unbalanced fixed effect anova. *Computational Statistics and Data Analysis*, 54:1881—1893, 2010. doi: <https://doi.org/10.1016/j.csda.2010.02.015>.
- A. Liu and Y. Wang. Hypothesis testing in smoothing spline models. *Journal of Statistical Computation and Simulation*, 74(8):581–597, 2004. doi: <https://doi.org/10.1080/00949650310001623416>.
- G. Marra and S. N. Wood. Coverage properties of confidence intervals for generalized additive model components. *Scand. J. Stat.*, 39(1):53–74, 2012. ISSN 0303-6898. doi: <https://doi.org/10.1111/j.1467-9469.2011.00760.x>.
- D. Nychka. Bayesian confidence intervals for smoothing splines. *J. Amer. Statist. Assoc.*, 83(404):1134–1143, 1988. ISSN 0162-1459. doi: <https://doi.org/10.1080/01621459.1988.10478711>.
- D. Nychka. The average posterior variance of a smoothing spline and a consistent estimate of the average squared error. *The Annals of Statistics*, 18(1):415–428, 1990. doi: <https://doi.org/10.1214/aos/1176347508>.
- F. Pesarin. *Multivariate Permutation Tests : With Applications in Biostatistics*. John Wiley & Sons, 1 edition, 6 2001.
- A. Quarteroni, R. Sacco, F. Saleri, and P. Gervasio. *Matematica Numerica*. Springer, Berlin Heidelberg, 4 edition, 2013.
- D. Ruppert, M. Wand, and R. Carroll. *Semiparametric Regression*. Cambridge University Press, 1 edition, 2003.
- L. M. Sangalli. Spatial regression with partial differential equations regularization. *International Statistical Review*, 89(3):505–531, 2021. doi: <https://doi.org/10.1111/insr.12444>.
- L. M. Sangalli, J. O. Ramsay, and T. O. Ramsay. Spatial spline regression models. *Journal of Royal Statistical Society B*, 75:681–703, 2013. doi: <https://doi.org/10.1111/rssb.12009>.
- B. W. Silverman. Some aspects of the spline smoothing approach to non-parametric regression curve fitting. *Journal of the Royal Statistical Society Series B*, 47:1–21, 1985. URL <http://www.jstor.org/stable/2345542?origin=JSTOR-pdf>.
- P. Speckman. Kernel smoothing in partial linear models. *Journal of the Royal Statistical Society Series B*, 50 (3):413–436, 1988. doi: <https://doi.org/10.1111/j.2517-6161.1988.tb01738.x>.

- G. Wahba. Bayesian “confidence intervals” for the cross-validated smoothing spline. *J. Roy. Statist. Soc. Ser. B*, 45(1):133–150, 1983. ISSN 0035-9246. URL [http://links.jstor.org/sici?sici=0035-9246\(1983\)45:1<133:B\"IFTC>2.0.CO;2-B&origin=MSN](http://links.jstor.org/sici?sici=0035-9246(1983)45:1<133:B\).
- G. Wahba. *Spline Models for Observational Data*. Society for Industrial and Applied Mathematics, 3600 Market Street, 6th Floor Philadelphia, PA 19104 USA, 1990.
- Y. Wang. *Smoothing Splines*. Chapman and Hall/CRC, 1 edition, 2011.
- M. Wilhelm, L. Dedè, L. M. Sangalli, and P. Wilhelm. Igs: An isogeometric approach for smoothing on surfaces. *Computer Methods in Applied Mechanics and Engineering*, 302:70–89, 2016. doi: <https://doi.org/10.1016/j.cma.2015.12.028>.
- S. N. Wood. On p-values for smooth components of an extended generalized additive model. *Biometrika*, 100(1):221–228, 2013. doi: <https://doi.org/10.1093/biomet/ass048>.
- S. N. Wood. *Generalized Additive Models*. Chapman and Hall/CRC, 2 edition, 2017.
- S. N. Wood, N. Pya, and B. Säfken. Smoothing parameter and model selection for general smooth models. *Journal of the American Statistical Association*, 111(516):1548–1563, 2016. doi: <https://doi.org/10.1080/01621459.2016.1180986>.
- L. Xiao. Asymptotic theory of penalized splines. *Electronic Journal of Statistics*, 13:747–794, 2019. doi: <https://doi.org/10.1214/19-EJS1541>.
- S. Yu, G. Wang, L. Wang, C. Liu, and L. Yang. Estimation and inference for generalized geoadditive models. *Journal of the American Statistical Association*, 2019. doi: <https://doi.org/10.1080/01621459.2019.1574584>.

Supporting Information for the paper Sign-Flip inference for spatial regression with differential regularization

M. Cavazzutti¹, E. Arnone², F. Ferraccioli³, C. Galimberti¹, L. Finos³, and L. M. Sangalli^{*1}

¹MOX - Dept. of Mathematics, Politecnico di Milano, Milan, Italy

²Dept. of Management, University of Torino, Turin, Italy

³Dept. of Statistical Science, University of Padova, Padua, Italy

16 September 2024

Abstract

This document contains supporting information for the article: Sign-Flip inference for spatial regression with differential regularization. We prove the theoretical results detailed in the main article. We propose an efficient algorithm for the computation of Confidence Intervals on the linear component of the model β , based on the nonparametric tests discussed in the main article. We provide details on the nonparametric inference on f when the locations tested are a subset of the mesh nodes. Finally, we provide further details on the simulations presented in the main article.

1 Proofs

1.1 Proofs of theorems on the linear component β

Proof of Proposition 1. To lighten the notation, we omit the subscript β , using the notation $\mathbf{T} = \mathbf{T}_{\beta, \cdot}$, where \cdot is any other symbol also used as \mathbf{T} subscript. We can rewrite \mathbf{T}_{Π} as follows.

$$\begin{aligned} \mathbf{T}_{\Pi} &= n^{-\frac{1}{2}} X^{\top} V \mathbf{H} V^{\top} \Lambda (\mathbf{y} - X \beta_0) = n^{-\frac{1}{2}} X^{\top} V D \mathbf{H} V^{\top} (\mathbf{y} - X \beta_0) = n^{-\frac{1}{2}} X^{\top} \sum_{k=1}^n (\pi_k d_k \mathbf{v}_k \mathbf{v}_k^{\top}) (\mathbf{y} - X \beta_0) = \\ &= n^{-\frac{1}{2}} \sum_{k=1}^n (X^{\top} \mathbf{v}_k) (\pi_k d_k) (\mathbf{v}_k^{\top} (\mathbf{y} - X \beta_0)) = n^{-\frac{1}{2}} \sum_{k=1}^n \mathbf{T}_{\Pi, k}, \end{aligned}$$

and \mathbf{b}_{λ} , for a given Π , as:

$$\mathbf{b}_{\lambda} = \mathbb{E}[\mathbf{T}_{\Pi}] = n^{-\frac{1}{2}} \sum_{k=1}^n (X^{\top} \mathbf{v}_k) (\pi_k d_k) (\mathbf{v}_k^{\top} \mathbb{E}[\mathbf{y} - X \beta_0]) = n^{-\frac{1}{2}} \sum_{k=1}^n (X^{\top} \mathbf{v}_k) (\pi_k d_k) (\mathbf{v}_k^{\top} \Psi \mathbf{f}) = n^{-\frac{1}{2}} \sum_{k=1}^n \mathbf{b}_{\lambda, k}.$$

We obtain the desired result observing that both the expressions of $\mathbf{b}_{\lambda, k}$ and $\mathbf{T}_{\Pi, k}$ depend on (d_k, \mathbf{v}_k) . \square

Proof of Proposition 2. Let Ω be a square domain and consider the eigenvalue-eigenfunction problem $(\Delta f)^2 = \mu f$ for $f : \Omega \Rightarrow \mathbb{R}$, with homogeneous Neumann boundary conditions [see, e.g., Evans, 1998]. Grebenkov and Nguyen [2013] prove that the eigenvalue-eigenfunction couples are of the form $(\mu_{i,j} = i^2 + j^2, \cos(ip_1) \cos(jp_2))$, for $i, j \in \mathbb{N}$. Now consider a uniform mesh discretizing the square domain Ω , and the associated Finite Element discretization P of $(\Delta f)^2$. Section (3.1) of Boffi [2010] proves the convergence of the finite dimensional eigenvalue-eigenvector couples of P to the infinite-dimensional couples above, with quadratic rate. Now let the data locations $\{\mathbf{p}_i\}_{i=1, \dots, n}$ coincide with the mesh nodes $\{\mathbf{x}_j\}_{j=1, \dots, N}$. Then $\Psi = I$, and, as a consequence, the eigenvectors of B coincide with the eigenvectors of P , which converge to $\cos(ip_1) \cos(jp_2)$ for $i, j \in \mathbb{N}$, and the eigenvalues are $d_{i,j} = \frac{1}{1+\lambda\mu_{i,j}}$, which in turn converge to $d_{i,j} = \frac{1}{1+\lambda(i^2+j^2)}$ \square

Proof of Proposition 3. To lighten the notation, we omit the subscript β , using the notation $\mathbf{T} = \mathbf{T}_{\beta, \cdot}$, where \cdot is any other symbol also used as \mathbf{T} subscript. For the same reason, we replace the symbol Π_{J_r} by Π_r . Let $m = 2^{n-r}$. We want to show that, asymptotically, the distribution of $\mathbf{T} = (\mathbf{T}_I, \dots, \mathbf{T}_{\Pi_r, m})^{\top}$ is invariant under

*Correspondence: Laura.sangalli@polimi.it

partial sign-flip transformations Π_r . We denote by $\stackrel{d}{=}$ the equality in distribution. We have that $\Pi_r \circ \mathbf{T} = \Pi_r \circ (\mathbf{T}_I, \mathbf{T}_{\Pi_{r,1}}, \mathbf{T}_{\Pi_{r,2}}, \dots, \mathbf{T}_{\Pi_{r,m}})^\top = (\mathbf{T}_{I\Pi_r}, \mathbf{T}_{\Pi_{r,1}\Pi_r}, \mathbf{T}_{\Pi_{r,2}\Pi_r}, \dots, \mathbf{T}_{\Pi_{r,m}\Pi_r})$, so the desired result is equivalent to prove that $\mathbf{T} \stackrel{d}{=} \Pi_r \circ \mathbf{T}$. We proceed as in [Ferraccioli et al., 2023] to provide such proof. Consider the eigendecomposition $\Lambda = VDV^\top$, let $\tilde{\mathbf{r}} = D^{1/2}V^\top(\mathbf{y} - X\beta_0)$ and let \tilde{X} be the diagonal matrix of dimension $n \times n$ with elements $(\tilde{X}_1, \dots, \tilde{X}_n)$. The ESF test statistic can be rewritten as

$$\mathbf{T}_{\Pi_r} = n^{-\frac{1}{2}} \tilde{X}^\top \Pi_r \tilde{\mathbf{r}} = n^{-\frac{1}{2}} \mathbf{1}^\top \Pi_r \tilde{X} \tilde{\mathbf{r}},$$

where $\mathbf{1}$ is an n -dimensional vector with all entries 1. \mathbf{T}_{Π_r} can hence be interpreted as the sum of n contributions, where the elements of $\tilde{X} \tilde{\mathbf{r}}$ that are not in J_r are sign-flipped through Π_r . The variance of \mathbf{T}_{Π_r} can be written as

$$\text{Var}(\mathbf{T}_{\Pi_r}) = n^{-1} \sigma^2 \mathbf{1}^\top \tilde{X} D \tilde{X} \mathbf{1}.$$

Let Π_T be the matrix with m rows and n columns collecting all the $m = 2^{n-r}$ vectors (of length n) of sign-flips. Then, we have that $\mathbf{T} = \Pi_T \tilde{X} \tilde{\mathbf{r}}$ and $\Pi_r \circ \mathbf{T} = \Pi_T \Pi_r \tilde{X} \tilde{\mathbf{r}}$. The joint distribution of \mathbf{T} is asymptotically multivariate normal with zero mean and variance $\text{Var}(\mathbf{T}) = \sigma^2 n^{-1} \Pi_T \tilde{X} D \tilde{X} \Pi_T^\top$. It remains to show that, asymptotically, $\Pi_r \circ \mathbf{T}$ has the same distribution for all Π_r . Observing that the expected value is not affected by the transformation Π_r , we have that $\Pi_r \circ \mathbf{T}$ has asymptotically null mean. Furthermore, for the variance we have

$$\begin{aligned} \text{Var}(\Pi_r \circ \mathbf{T}) &= n^{-1} \sigma^2 \Pi_T \Pi_r \tilde{X} D \tilde{X} \Pi_r \Pi_T^\top = \\ &= n^{-1} \sigma^2 \Pi_T \Pi_r \Pi_r \tilde{X} D \tilde{X} \Pi_T^\top = \\ &= n^{-1} \sigma^2 \Pi_T \tilde{X} D \tilde{X} \Pi_T^\top. \end{aligned}$$

Applying Theorem 15.2.1 in Lehmann and Romano [2008] and Theorem 1 in Hemerik and Goeman [2018], we thus obtain the null invariance $\mathbf{T} \stackrel{d}{=} \Pi_r \circ \mathbf{T}$. Consequently, under H_0 , we have that $\mathbb{P}(\mathbf{T}_I > \mathbf{T}_{[1-\alpha]}) \leq \alpha$. \square

1.2 Proofs of theorems on the nonlinear component f

Proof of Proposition 4. Consider the expression of $\hat{\gamma} = \frac{\frac{1}{n_Z} \sum_{i=1}^{n_Z} \hat{\mathbf{f}}_{Z,i}}{\frac{1}{n_Z} \sum_{i=1}^{n_Z} f_0(\mathbf{z}_i)}$. In the following, we assume that Assumptions 3-5 in Arnone et al. [2022] are satisfied when n_Z increases. For n_Z large, the term $\frac{1}{n_Z} \sum_{i=1}^{n_Z} f_0(\mathbf{z}_i)$ converges to $\int_{\Omega} f_0(\mathbf{z}) d\mathbf{z}$. Moreover, Theorem 6 of Arnone et al. [2022] proves the consistency of the estimator \hat{f} . Thus, the numerator $\frac{1}{n_Z} \sum_{i=1}^{n_Z} \hat{\mathbf{f}}_{Z,i}$ of $\hat{\gamma}$ converges in probability to $\gamma \int_{\Omega} f_0(\mathbf{z}) d\mathbf{z}$. Applying Slutsky theorem [Gut, 2005], we have that $\hat{\gamma}$ converges in probability to γ and asymptotically, $\mathbb{E}[\hat{\mathbf{f}}_Z - \hat{\gamma} \mathbf{f}_0] = 0$. It remains to show that $\text{Var}[\hat{\mathbf{f}}_Z - \hat{\gamma} \mathbf{f}_0] \rightarrow \text{Var}[\hat{\mathbf{f}}_Z - \gamma \mathbf{f}_0]$ for $n_Z \rightarrow \infty$. This is a direct consequence of the convergence in probability of $\hat{\gamma}$ to γ and of Slutsky theorem [Gut, 2005]. The desired result is obtained proceeding as in the proof of Theorem 1 in Arnone et al. [2023]. \square

Proof of Theorem 1. To lighten the notation, we omit the subscript f , using the notation $\mathbf{T} = \mathbf{T}_f$, where \cdot is any other symbol also used as subscript of \mathbf{T} . We begin to show the proof for H_0 when $\gamma = \gamma_0 = 1$ is known, and then we extend it to the case of unknown γ . Under the null hypothesis H_0 in (14), we have $\mathbf{T} = n_Z^{-1/2} \Psi_Z^\top Q_Z (\mathbf{y}_Z - \Psi_Z \mathbf{f}_0)$. Thus:

$$\begin{aligned} \mathbb{E}_{H_0}[\mathbf{T}] &= n_Z^{-1/2} \Psi_Z^\top Q_Z \mathbb{E}[X_Z \beta + \epsilon_Z] = \mathbf{0}, \\ \text{Var}_{H_0}[\mathbf{T}] &= n_Z^{-1} \Psi_Z^\top Q_Z \text{Var}[X_Z \beta + \epsilon_Z] Q_Z^\top \Psi_Z = n_Z^{-1} \sigma^2 \Psi_Z^\top Q_Z \Psi_Z, \end{aligned}$$

where ϵ_Z is the vector containing the errors associated with the n_Z locations under test, and we have exploited the fact that under the null hypothesis $\mathbf{y}_Z = X_Z \beta + \Psi_Z \mathbf{f}_0 + \epsilon_Z$, with $\mathbb{E}[\epsilon_Z] = \mathbf{0}$ and $\text{Var}[\epsilon_Z] = \sigma^2 I$, and that $Q_Z X_Z = \mathbf{0}$. Therefore, \mathbf{T} can be rewritten as a sum of n_Z independent random vectors as follows:

$$\mathbf{T} = n_Z^{-1/2} \Psi_Z^\top Q_Z (\mathbf{y}_Z - \Psi_Z \mathbf{f}_0) = n_Z^{-1/2} \Psi_Z^\top Q_Z \epsilon_Z = n_Z^{-1/2} \sum_{i=1}^{n_Z} [\Psi_Z^\top Q_Z]_i \epsilon_i$$

where $[\Psi_Z^\top Q_Z]_i$ indicates the i th column of the $N \times n_Z$ matrix $\Psi_Z^\top Q_Z$. It then follows from the Lindeberg-Feller central limit theorem [see, e.g., Vaart, 1998] that \mathbf{T} is also asymptotically normal. Consider now the case of γ unknown. Let $\hat{\gamma}$ be defined as in the proof of Proposition 4. Under the null hypothesis H_0 in (14), we have that $\mathbf{T} = n_Z^{-1/2} \Psi_Z^\top Q_Z (\mathbf{y}_Z - \hat{\gamma} \mathbf{f}_0)$, and thus, for large n_Z :

$$\begin{aligned} \mathbb{E}_{H_0}[\mathbf{T}] &= n_Z^{-1/2} \Psi_Z^\top Q_Z \mathbb{E}[(\gamma - \hat{\gamma}) \mathbf{f}_0 + X_Z \beta + \epsilon_Z] = \mathbf{0}, \\ \text{Var}_{H_0}[\mathbf{T}] &= n_Z^{-1} \Psi_Z^\top Q_Z \text{Var}[(\gamma - \hat{\gamma}) \mathbf{f}_0 + X_Z \beta + \epsilon_Z] Q_Z^\top \Psi_Z = n_Z^{-1} \sigma^2 \Psi_Z^\top Q_Z \Psi_Z. \end{aligned}$$

Again, to derive the above expressions, we have exploited the fact that, under the null hypothesis, we have that $\mathbf{y}_Z = X_Z \beta + \gamma \mathbf{f}_0 + \epsilon_Z$, with $\mathbb{E}[\epsilon_Z] = \mathbf{0}$ and $\text{Var}[\epsilon_Z] = \sigma^2 I$, and that $Q_Z X_Z = \mathbf{0}$. Moreover, since $\hat{\gamma}$ converges to γ in probability, for Slutsky theorem [Gut, 2005], we also have that $\text{Var}[(\gamma - \hat{\gamma})\mathbf{f}_0 + X_Z \beta + \epsilon_Z] = \text{Var}[X_Z \beta + \epsilon_Z]$. Therefore, \mathbf{T} can be rewritten as a sum of n_Z independent N -dimensional random vectors

$$\mathbf{T} = n_Z^{-1/2} \Psi_Z^\top Q_Z (\mathbf{y}_Z - \hat{\gamma} \mathbf{f}_0) = n_Z^{-1/2} \Psi_Z^\top Q_Z (\epsilon_Z + (\gamma - \hat{\gamma}) \mathbf{f}_0) = n_Z^{-1/2} \sum_i^{n_Z} [\Psi_Z^\top Q_Z]_i (\epsilon_i + (\gamma - \hat{\gamma}) \mathbf{f}_{0,i}).$$

It then follows from the Lindeberg-Feller central limit theorem [Vaart, 1998] that \mathbf{T} is asymptotically normal. \square

Proof of Proposition 5. To lighten the notation we will omit the subscript f , using the notation $\mathbf{T}_\cdot = \mathbf{T}_{f,\cdot}$, where \cdot is any other symbol already used as \mathbf{T} subscript. It is trivial to prove that the expected value is zero. Concerning the variance, we have:

$$\text{Var}_{H_0}[\mathbf{T}_\Pi] = n_Z^{-1} \Psi_Z^\top \Pi Q_Z \text{Var}_{H_0}[\mathbf{y}_Z - \Psi_Z \mathbf{f}_0] Q_Z^\top \Pi \Psi_Z = n_Z^{-1} \sigma^2 \Psi_Z^\top \Pi Q_Z Q_Z^\top \Pi \Psi_Z = n_Z^{-1} \sigma^2 \Psi_Z^\top \Pi Q_Z \Pi \Psi_Z$$

exploiting the fact that $n_Z \text{Var}_{H_0}[\mathbf{y}_Z - \Psi_Z \mathbf{f}_0] = \text{Var}_{H_0}[\epsilon_Z] = \sigma^2 I$ and that $Q_Z Q_Z^\top = Q_Z$. \square

Proof of Proposition 6. To lighten the notation, we omit the subscript f , using the notation $\mathbf{T}_\cdot = \mathbf{T}_{f,\cdot}$, where \cdot is any other symbol also used as \mathbf{T} subscript. Consider the j th component of the score statistic defined in equation (17) of the main article, corresponding to the j th row of Ψ_Z^\top , for which we have $T_1^j = n_Z^{-1/2} \sum_{i=1}^{n_Z} \nu_i^j$, with $\nu_i^j = \psi_j(\mathbf{p}_i)[Q_Z \mathbf{r}_Z]_i$. Let $m = 2^{n_Z}$. Let $\Pi_{\mathbf{T}}$ be the matrix with m rows and n_Z columns, collecting all the m vectors (of length n_Z) of sign-flips, with $\Pi_{\mathbf{T},1} = \mathbf{1}$. Consider the m -dimensional vector of statistics $\mathbf{T}^j = (T_1^j, \dots, T_m^j)$, where $T_h^j = T_{\Pi_{\mathbf{T},h}}^j$, for $1 \leq h \leq m$. Under H_0 in (14), we have that $\mathbb{E}[T_h^j] = 0$ and $\text{Var}[T_h^j] = n_Z^{-1} \sum_{i=1}^{n_Z} \text{Var}[\nu_i^j] =: s_{n_Z}^2$, for $2 \leq h \leq m$, as $\Pi_{\mathbf{T},h}$ is a n_Z -dimensional vector of independent sign-flips. On the converse, the variance of the observed score statistic is:

$$\text{Var}[T_1^j] = n_Z^{-1} \sum_{i=1}^{n_Z} \text{Var}[\nu_i^j] - 2n_Z^{-1} \sum_{\substack{i \in 1, \dots, n_Z \\ l \in 1, \dots, n_Z \\ i < l}} \text{Cov}(\nu_i^j, \nu_l^j) \leq n_Z^{-1} \sum_{i=1}^{n_Z} s_{n_Z}^2,$$

with $\text{Cov}(\nu_i^j, \nu_l^j) = \sigma^2 \psi_j(\mathbf{p}_i) Q_{Z,i,l} \psi_j(\mathbf{p}_l)$, for $i, l \in 1, \dots, n_Z$, which vanishes asymptotically. Moreover, the covariance matrix of the vector \mathbf{T}^j has zeros entries off diagonal, due to the fact that $(\Pi_{\mathbf{T},h})_i$, for $2 \leq h \leq m$, are independent with zero mean. Then, by the Lindeberg-Feller central limit theorem, \mathbf{T}^j converges in distribution to a vector of i.i.d. random variables, with asymptotic multivariate normal distribution with zero mean and covariance $\lim_{n_Z \rightarrow \infty} s_{n_Z}^2 I$. Consequently, we can use Lemma 11 in Hemerik et al. [2020] to derive the asymptotic exactness of the test corresponding to the j th component. Finally, from the multivariate central limit theorem [Vaart, 1998] and Theorem 8 in Hemerik et al. [2020] follows the asymptotic exactness of the global test using the combined statistic $\mathbf{T}_\Pi^{\text{comb}}$. \square

Proof of Proposition 7. To lighten the notation, we omit the subscript f and the superscript E , using the notation $\mathbf{T}_\cdot = \mathbf{T}_{f,\cdot}^E$, where \cdot is any other symbol also used as \mathbf{T} subscript. It is trivial to prove that the expected value is zero. Concerning the variance, we have:

$$\text{Var}_{H_0}[\mathbf{T}_\Pi] = n_Z^{-1} \sigma^2 \Psi_Z^\top V \Pi V^\top Q_Z Q_Z^\top V \Pi V^\top \Psi_Z = n_Z^{-1} \sigma^2 \Psi_Z^\top V \Pi I_{n_Z-q} \Pi V^\top \Psi_Z = n_Z^{-1} \sigma^2 \Psi_Z^\top Q_Z \Psi_Z.$$

where we exploited the fact that $Q_Z Q_Z^\top = Q_Z$, $Q_Z = V V^\top$ and $V^\top V = I_{n_Z-q}$. \square

Proof of Proposition 8. To lighten the notation, we omit the subscript f and the superscript E , using the notation $\mathbf{T}_\cdot = \mathbf{T}_{f,\cdot}^E$, where \cdot is any other symbol also used as \mathbf{T} subscript. The argument proceeds along the lines of the proof of Proposition 6. Consider the j th component of the N dimensional statistic (18), corresponding to the j th row of Ψ_Z^\top , $T_1^j = n_Z^{-1/2} \sum_{i=1}^{n_Z} [\Psi_Z^\top V]_i [V^\top Q_Z \mathbf{r}_Z]_i$. Let $m = 2^{n_Z}$. Let $\Pi_{\mathbf{T}}$ be the matrix with m rows and n_Z columns collecting all the m vectors (of length n_Z) of sign-flips, with $\Pi_{\mathbf{T},1} = \mathbf{1}$. Consider the m -dimensional vector of statistics $\mathbf{T}^j = (T_1^j, \dots, T_m^j)$, where $T_h^j = T_{\Pi_{\mathbf{T},h}}^j$, for $1 \leq h \leq m$. We can rewrite T_h^j as

$$T_h^j = n_Z^{-1/2} \mathbf{1}_{n_Z}^\top \Pi_h \tilde{\Psi}_Z \tilde{\mathbf{r}}_Z$$

The above expression highlights that T_h^j is a sum of n_Z contributions, each one sign-flipped through Π . Moreover, under the null hypothesis (14):

$$\begin{aligned} \mathbb{E}[T_h^j] &= 0, \\ \text{Var}[T_h^j] &= n_Z^{-1} \sigma^2 \mathbf{1}_{n_Z}^\top \tilde{\Psi}_Z \tilde{\Psi}_Z \mathbf{1}_{n_Z}. \end{aligned}$$

We can express the vector \mathbf{T}^j as:

$$\mathbf{T}^j = n_{\mathcal{Z}}^{-1/2} \Pi_{\mathbf{T}} \tilde{\Psi}_{\mathcal{Z}} \tilde{\mathbf{r}}_{\mathcal{Z}}.$$

Thus, \mathbf{T}^j has zero mean and variance given by

$$\text{Var}[\mathbf{T}^j] = n_{\mathcal{Z}}^{-1} \sigma^2 \Pi_{\mathbf{T}} \tilde{\Psi}_{\mathcal{Z}} \tilde{\Psi}_{\mathcal{Z}} \Pi_{\mathbf{T}}^{\top}.$$

Applying a sign-flip transformation Π to \mathbf{T}^j we have that $\Pi \circ \mathbf{T}^j = n_{\mathcal{Z}}^{-1/2} \Pi_{\mathbf{T}} \Pi \tilde{\Psi}_{\mathcal{Z}} \tilde{\mathbf{r}}_{\mathcal{Z}}$. Clearly, the matrix of the sign-flips Π does not affect the expected value and also the variance is invariant to its application:

$$\text{Var}[\Pi \circ \mathbf{T}^j] = n_{\mathcal{Z}}^{-1} \sigma^2 \Pi_{\mathbf{T}} \Pi \tilde{\Psi}_{\mathcal{Z}} \tilde{\Psi}_{\mathcal{Z}} \Pi \Pi_{\mathbf{T}} = n_{\mathcal{Z}}^{-1} \sigma^2 \Pi_{\mathbf{T}} \tilde{\Psi}_{\mathcal{Z}} \tilde{\Psi}_{\mathcal{Z}} \Pi_{\mathbf{T}}^{\top}.$$

Using Lindeberg-Feller multivariate central limit theorem we have that \mathbf{T}^j and $\Pi \circ \mathbf{T}^j$ converge to the same asymptotic multivariate normal distribution, which proves the asymptotically exactness of the test corresponding to the j th component of the test statistic. We can then derive the exactness of the global test using the multivariate central limit theorem [Vaart, 1998] and Theorem 8 in Hemerik et al. [2020]. \square

2 Confidence intervals

In this section we describe how to compute Confidence Intervals (CI) for β , based on the proposed PESF statistic. Let $p(b)$ be the p-value for the nonparametric test $H_0 : \beta = b$ vs $H_1 : \beta \neq b$. For the sake of simplicity, we here focus on the case $q = 1$. Let $[L, U]$, with $L \leq U$, be the desired confidence interval of level $1 - \alpha$ for β , i.e., $\mathbb{P}(L < \beta < U) \geq 1 - \alpha$. Let L_l, L_u, U_l, U_u be four values such that $L_l \leq L \leq L_u$ and $U_l \leq U \leq U_u$. Algorithm 1 details how to determine U such that $\mathbb{P}(\beta > U) \leq \frac{\alpha}{2}$. The algorithm determines the upper bound U of the CI in two steps. In the first step, starting from a guess obtained by means of the parametric CI, either Speckman or Wald, two values U_u and U_l are determined, such that $p(U_u) \leq \frac{\alpha}{2}$ and $p(U_l) \geq \frac{\alpha}{2}$. Then U is obtained via a bisection algorithm, starting from U_u and U_l . Using Speckman in step 2 of the algorithm is computationally convenient, as it requires to compute the same quantities that are also needed for the partial ESF tests.

Algorithm 1 Computation of upper extreme of $\text{CI}(1 - \alpha)$ for β

<pre> 1: Compute $\hat{\beta}$ 2: Compute Speckman CI range R_s 3: $U_u = \hat{\beta} + \frac{R_s}{4}$, $U_l = \hat{\beta} - \frac{R_s}{4}$ 4: while ($!(p(U_u) \leq \frac{\alpha}{2}) \& (p(U_l) \geq \frac{\alpha}{2}))$) do 5: if ($p(U_u) > \frac{\alpha}{2}$) then 6: $U_u = U_u + 1.5 * \frac{U_u - U_l}{2}$ 7: end if 8: if ($p(U_l) < \frac{\alpha}{2}$) then 9: Select $U_l = \frac{U_l + \hat{\beta}}{2}$ 10: end if </pre>	<pre> 11: end while 12: Start bisection algorithm: 13: while ($U_u - U_l > \frac{R_s}{20}$) do 14: $U_n = \frac{U_u + U_l}{2}$ 15: if ($p(U_n) \leq \frac{\alpha}{2}$) then 16: $U_u = U_n$ 17: else 18: $U_l = U_n$ 19: end if 20: end while 21: $U = \frac{U_u + U_l}{2}$ </pre>
---	---

3 Sign-Flip for f : when locations coincide with mesh nodes

We detail here the proposed nonparametric inference on f for the specific case in which all tested data locations in \mathcal{Z} in eq (14) of the main manuscript are a subset of the mesh nodes. Consider, for simplicity of exposition, the Sign-Flip test. Then the score test statistic becomes:

$$\mathbf{T}_{f, \Pi} = n_{\mathcal{Z}}^{-1/2} \Psi_{\mathcal{Z}}^{\top} \Pi Q_{\mathcal{Z}} \mathbf{r}_{\mathcal{Z}}$$

where all the \mathcal{Z} -subscripts stress the fact that the test statistic is computed using the chosen $n_{\mathcal{Z}}$ data points in \mathcal{P} . In particular the $n_{\mathcal{Z}} \times N$ matrix $\Psi_{\mathcal{Z}}$ has at most one entry equal to 1 in each row, while all the other entries are zeros, due to the fact that each point coincides exactly with one node. Hence, each component of the N -dimensional score statistic is either zero, if the corresponding location has not been selected for the test,

or is equal to the residual at that location, sign-flipped according to Π . In fact, the score test statistic becomes:

$$\mathbf{T}_{f,\Pi} = n_{\mathcal{Z}}^{-1/2} \Psi_{\mathcal{Z}}^{\top} \Pi Q_{\mathcal{Z}} \mathbf{r}_{\mathcal{Z}} = n_{\mathcal{Z}}^{-1/2} \begin{bmatrix} \pi_1 r_1 \\ \vdots \\ \pi_j r_j \\ \vdots \\ \pi_N r_N \end{bmatrix},$$

where r_j is the component of $Q_{\mathcal{Z}} \mathbf{r}_{\mathcal{Z}}$ that corresponds to the j th node, if the j th node is under test, otherwise the entry is zero. Such configuration prevents the use of the global statistic \mathbf{T}_{Π}^{comb} defined in the main article. Indeed if we consider the sum of squares for combination, its value is invariant with respect to Π and the test is no longer meaningful. A possible practical solution consists in combining the values of nearby nodes. Indeed, instead of considering only the residual corresponding to the j th node, we can sum all the residuals of the closest nodes in the mesh. We define the following matrix:

$$G = \begin{bmatrix} I_{\mathcal{G}_1}(\mathbf{p}_1) & \dots & I_{\mathcal{G}_1}(\mathbf{p}_{n_{\mathcal{Z}}}) \\ \vdots & \dots & \vdots \\ I_{\mathcal{G}_N}(\mathbf{p}_1) & \dots & I_{\mathcal{G}_N}(\mathbf{p}_{n_{\mathcal{Z}}}) \end{bmatrix}$$

where $I_{\mathcal{G}_j}(\mathbf{p}_i) = 1$ if and only if point \mathbf{p}_i belongs to the neighborhood \mathcal{G}_j of node j . Then the score test statistic becomes:

$$\mathbf{T}_{f,\Pi} = n_{\mathcal{Z}}^{-1/2} G \Pi Q_{\mathcal{Z}} \mathbf{r}_{\mathcal{Z}}$$

and we can take again as global test statistic \mathbf{T}_{Π}^{comb} the sum of squares of the $\mathbf{T}_{f,\Pi}$ entries. For what concerns the Eigen-Sign-Flip test the approach is analogous, replacing $\Psi_{\mathcal{Z}}$ with G in Definition 3.

4 Simulation 1: inference on the linear component

This section reports details and additional results for the simulation studies described in Section 5.1 of the main manuscript. The true spatial function f_0 is defined as:

$$f_0(\mathbf{p}) = 0.4\pi^{0.3} \left(1.2 \exp \left(-\frac{(p_1 - 0.2)^2}{0.3^2} - \frac{(p_2 - 0.3)^2}{0.4^2} \right) + 0.8 \exp \left(-\frac{(p_1 - 0.7)^2}{0.3^2} - \frac{(p_2 - 0.8)^2}{0.4^2} \right) \right).$$

Figure 1 shows the power curves for the cases a) and c) in Simulation 5.1, in analogy to what has already displayed in the main article, concerning for the cases b) and d).

5 Simulation 2: inference on the nonlinear component

This section contains details and additional results concerning the simulations described in Section 5.2 of the main manuscript. The function f_1 is defined as:

$$f_1(\mathbf{p}) = 0.4\pi^{0.3} \left(1.6 \exp \left(-\frac{(p_1 - 0.55)^2}{0.3^2} - \frac{(p_2 - 0.15)^2}{0.4^2} \right) + 0.5 \exp \left(-\frac{(p_1 - 0.85)^2}{0.3^2} - \frac{(p_2 - 0.4)^2}{0.4^2} \right) \right).$$

Figure 2 reports the rejection rates curves when inference is computed with stochastic approximation for Simulation 5.2. The setting is the same shown in the main article, except that inference is computed efficiently using the approximation with FSPA tolerance $\epsilon_{FSPA} = n \cdot 10^{-7}$. Notice that, despite the higher precision requested in this case with respect to Simulation 5.1, the Wald inference performance greatly worsen with respect to the results observed in Simulation 5.2. On the converse, the nonparametric tests maintain perfect control of Type I error and high power.

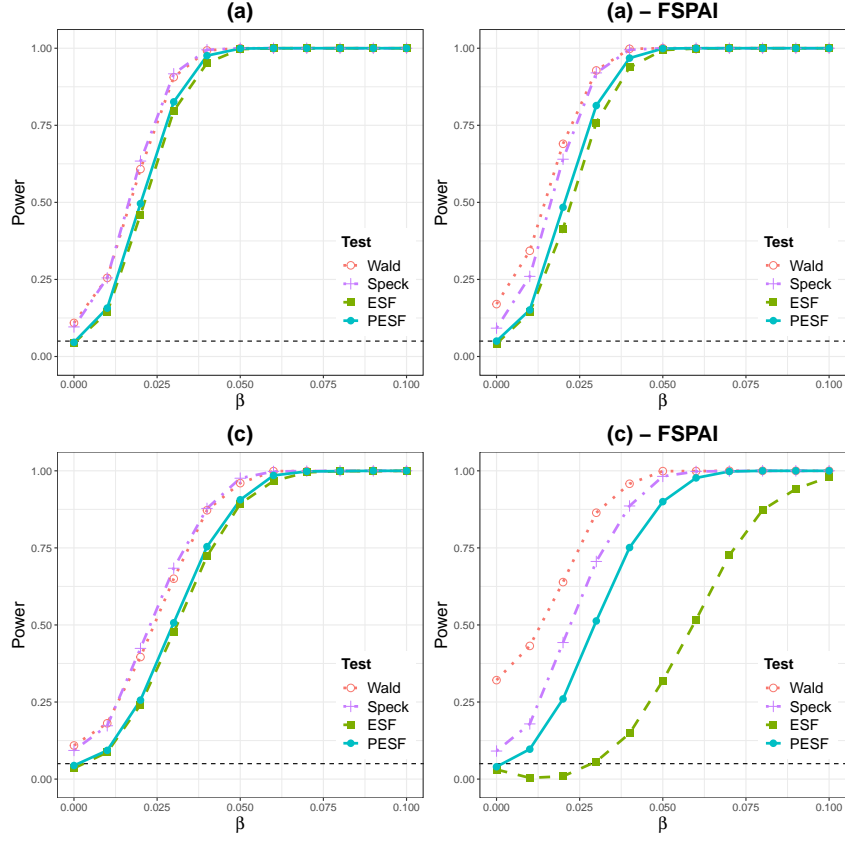


Figure 1: Simulation 1, inference on β . Rejection rates over 1000 repetitions, with covariates generated as in case (a) (top) and (c) (bottom), employing exact (left) and approximated (right) computations; Wald test (red dotted curve, with circle markers), Speckman (violet dash-and-dotted curve, with cross markers), standard ESF test (green dashed curve, with square markers) and the proposed PESF test (light blue solid curve, with circle markers).

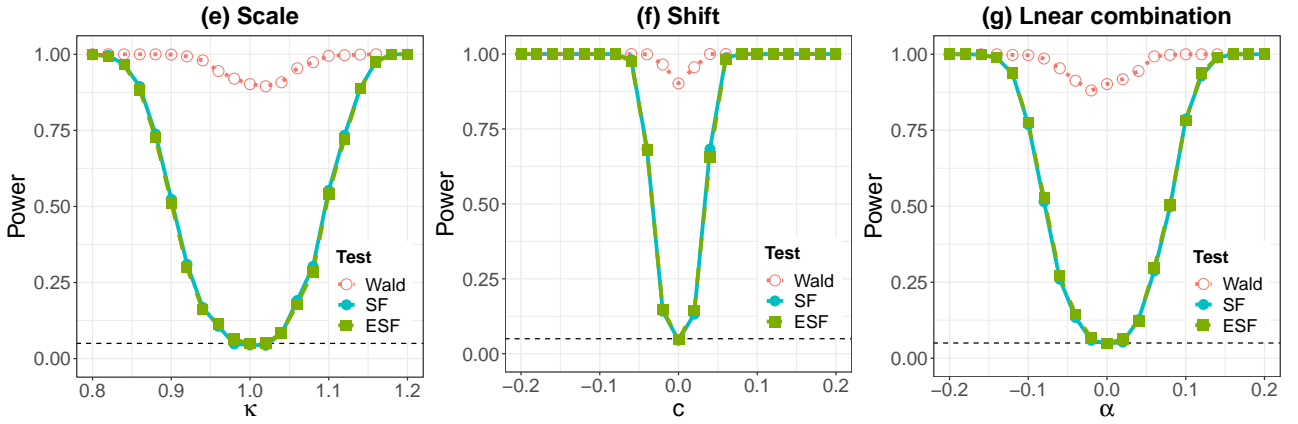


Figure 2: Simulation 2, inference on f . Rejection rates for 1000 repetitions, with approximated inference computation; Wald test (red dotted curve, with circle markers), SF nonparametric test (blue solid curve, with circle markers) and ESF test (green dashed curve, with square markers).

References

- E. Arnone, A. Kneip, F. Nobile, and L. M. Sangalli. Some first results on the consistency of spatial regression with partial differential equation regularization. *Statistica Sinica*, 32(1):209–238, 2022. doi: <https://doi.org/10.5705/ss.202019.0346>.
- E. Arnone, L. Negri, F. Panzica, and L. M. Sangalli. Analyzing data in complicated 3d domains: Smoothing, semiparametric regression, and functional principal component analysis. *Biometrics*, 79(4):3510–3521, 2023. doi: <https://doi.org/10.1111/biom.13845>.
- D. Boffi. Finite element approximation of eigenvalue problems. *Acta Numerica*, 19:1–120, 2010. doi: <https://doi.org/10.1017/S0962492910000012>.
- L. C. Evans. *Partial Differential Equation*. American Mathematical Society, 1998.
- F. Ferraccioli, L. M. Sangalli, and L. Finos. Nonparametric tests for semiparametric models. *Test*, 32:1106–1130, 2023. doi: <https://doi.org/10.1007/s11749-023-00868-9>.
- D. S. Grebenkov and B. T. Nguyen. Geometrical structure of laplacian eigenfunctions. *SIAM review*, 55(4): 601–667, 2013. doi: <https://doi.org/10.1137/120880173>.
- A. Gut. *Probability: a graduate course*. Springer-Verlag, 1 edition, 2005.
- J. Hemerik and J. Goeman. Exact testing with random permutations. *TEST*, 27(4):811–825, 2018. doi: <https://doi.org/10.1007/s11749-017-0571-1>.
- J. Hemerik, J. Goeman, and L. Finos. Robust testing in generalized linear models by sign-flipping score contributions. *Journal of the Royal Statistical Society Series B*, 82(3):841–864, 2020. doi: <https://doi.org/10.1111/rssb.12369>.
- E. L. Lehmann and J. P. Romano. *Testing Statistical Hypotheses*. Springer, 3 edition, 2008.
- A. W. Vaart. *Asymptotic Statistics*. Cambridge University Press, 1 edition, 1998.

MOX Technical Reports, last issues

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

- 63/2024** Vitullo, P.; Franco, N.R.; Zunino, P.
Deep learning enhanced cost-aware multi-fidelity uncertainty quantification of a computational model for radiotherapy
- 62/2024** Roknian, A.A.; Scotti, A.; Fumagalli, A.
Free convection in fractured porous media: a numerical study
- 59/2024** Carbonaro, D.; Ferro, N.; Mezzadri, F.; Gallo, D.; Audenino, A.; Perotto, S.; Morbiducci, U.; Chiastra, C.
Easy-to-use formulations based on the homogenization theory for vascular stent design and mechanical characterization
- 60/2024** Temellini, E.; Ferro, N.; Stabile, G.; Delgado Avila, E.; Chacon Rebollo, T.; Perotto, S.
Space - time mesh adaptation for the VMS - Smagorinsky modeling of high Reynolds number flows
- 61/2024** Speroni, G.; Ferro, N.
A novel metric - based mesh adaptation algorithm for 3D periodic domains
- Possenti, L.; Vitullo, P.; Cicchetti, A.; Zunino, P.; Rancati, T.
Modeling Hypoxia-Induced Radiation Resistance and the Impact of Radiation Sources
- Roknian, A.A.; Scotti, A.; Fumagalli, A.
Free convection in fractured porous media: a numerical study
- Speroni, G.; Ferro, N.
A novel metric-based mesh adaptation algorithm for 3D periodic domains
- Temellini, E.; Ferro, N.; Stabile, G.; Delgado Avila, E.; Chacon Rebollo, T.; Perotto, S.
Space-time mesh adaptation for the VMS-Smagorinsky modeling of high Reynolds number flows
- 58/2024** Ciaramella, G.; Vanzan, T.
Variable reduction as a nonlinear preconditioning approach for optimization problems