MOX-Report No. 98/2024

PDE-regularised spatial quantile regression

Castiglione, C.; Arnone, E.; Bernardi, M.; Farcomeni, A.; Sangalli, L.M.

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

mox-dmat@polimi.it https://mox.polimi.it

PDE-regularised spatial quantile regression

Cristian Castiglione[a,b], Eleonora Arnone[c], Mauro Bernardi[a], Alessio Farcomeni[d], Laura M. Sangalli[e,*]

[a] *Dipartimento di Scienze Statistiche, Universitá di Padova, Padova, Italia*
[b] *BIDSA, Bocconi Institute for Data Science and Analytics, Bocconi University*
[c] *Dipartimento di Management, Universitá di Torino, Torino, Italia*
[d] *Dipartimento di Economia e Finanza, Universitá di Roma Tor Vergata, Roma, Italia*
[e] *MOX, Dipartimento di Matematica, Politecnico di Milano, Milano, Italia*

Abstract

We consider the problem of estimating the conditional quantiles of an unknown distribution from data gathered on a spatial domain. We propose a spatial quantile regression model with differential regularisation. The penalisation involves a partial differential equation defined over the considered spatial domain, that can display a complex geometry. Such regularisation permits, on one hand, to model complex anisotropy and non-stationarity patterns, possibly on the basis of problem-specific knowledge, and, on the other hand, to comply with the complex conformation of the spatial domain. We define an innovative functional Expectation-Maximisation algorithm, to estimate the unknown quantile surface. We moreover describe a suitable discretisation of the estimation problem, and investigate the theoretical properties of the resulting estimator. The performance of the proposed method is assessed by simulation studies, comparing with state-of-the-art techniques for spatial quantile regression. Finally, the considered model is applied to two real data analyses, the first concerning rainfall measurements in Switzerland and the second concerning sea surface conductivity data in the Gulf of Mexico.

Keywords: Differential penalisation, Finite elements, Functional data analysis, Nonparametric quantile smoothing, Smoothing with roughness penalties
2020 MSC: Primary 62G08, Secondary 65D10

1. Introduction

In this work, we are interested in modelling the heterogeneous effects of complex spatial phenomena on the conditional quantiles of a response variable of interest, via nonparametric regression. Regression analysis is, in fact, a powerful and versatile tool for analysing the conditional distribution of noisy data given a set of covariates. However, it is generally recognised that the conditional mean is not the most suitable summary of location when samples significantly deviate from the Gaussian assumption. In particular, heteroscedasticity, skewness, lepto- or platy-kurtic tails are characteristics often present in real data, that can shadow the relationship between the response variable and the covariates, as postulated by classical conditional regression models on the data mean. Possibly restrictive parametric assumptions on the data generating mechanism could be a solution; for instance, generalised linear models [48] replace the Gaussian distribution with a more appropriate law belonging to the exponential family. Quantile regression models more generally characterise the conditional behaviour of the response variable, and are usually distribution-free and based solely on local assumptions. Conditional quantiles have the additional benefits of being invariant to monotone transformations of the data, and more robust to outlier contamination [36]. These reasons motivated the spread of conditional quantile methods, since their introduction by [38]. Quantiles play a relevant role in modern theoretical and applied statistical approaches, including, for instance, high-dimensional learning [70, 74], conformal prediction [61], risk assessment [62], and environmental statistics [69]. Moreover, they deliver optimal point estimates [15, 29] and classification [18].

*Corresponding author. Email address: laura.sangalli@polimi.it

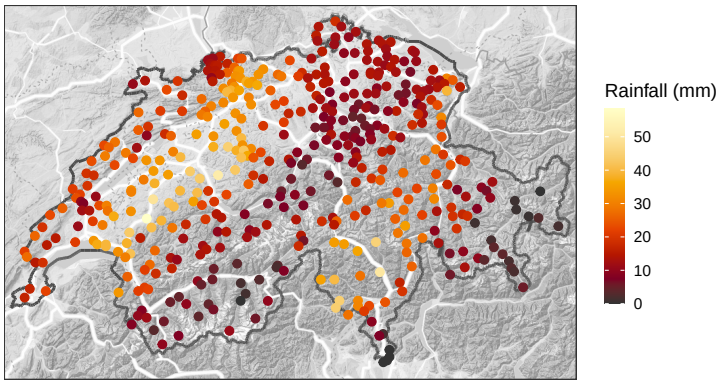

Fig. 1: Rainfall measurements at 467 locations in the Swiss territory.

In addition to univariate distributional characteristics of the response, geostatistical applications require taking into account the spatial dependence induced by the topological nature of the data generating process. Flows and external perturbations are well-known sources of complex structures of dependence in spatial data, whose impact can largely differ across quantiles. In addition, the complex conformation of the sampling domain, and the possible presence of boundary constraints, may introduce additional levels of complexity. In climate and oceanographic studies, for instance, temperature, pressure, and precipitation data often manifest local anisotropy and non-stationarity, and are influenced by wind or water streams, and by local characteristics of the geographical morphology. Moreover, such data typically manifests heteroscedastic skewed distributions, with local changes in space and time. Modelling these complex features is of prominent interest not only in environmental problems but also in various other fields, including physical and life sciences, as well as engineering.

A real data example is provided in Fig. 1, which portrays 467 rainfall measurements (in 1/10 mm units) collected in Switzerland on May 8, 1986 [14]. The data clearly manifest a strong southwest-northeast anisotropy, which might be caused by the geographical characteristics of the region, and their interactions with the local dynamics of the atmosphere. In addition, heteroscedastic noise and extreme values [14] introduce another source of variability that must be properly modelled to provide a complete description of the data distribution.

A second example is shown in Fig. 2, where we consider replicated historical oceanographic data on the sea surface electrical conductivity (in S/m scale), measured in the Gulf of Mexico around the Florida peninsula, from 1978 to 2017 [26, 83]. Data were derived from temperature, pressure and salinity observations, gathered by different data sources and aggregated on a regular grid in space and on a monthly basis in time. The behaviour of the sea surface characteristics in the area, such as conductivity, temperature and nutrient concentration, are strongly affected by the presence of the Florida coastline barrier and by the Gulf Stream, which originates in the Gulf of Mexico and transports a flow of warm water from the tropics to the North Atlantic Ocean. The presence of such non-stationary steam constitutes a big challenge from a modelling point of view, but it also provides an important source of additional information on the system.

In the spatial and quantile regression literature, many authors studied the problem of estimating nonparametric and semiparametric regression models in one- and two-dimensional domains. For example, [40, 51] considered quantile smoothing spline models with total variation regularisation, and they proposed a linear programming algorithm for parameter estimation. [11] studied quantile regression with cubic smoothing spline, and estimated the parameters with an interior point algorithm. [32] generalised the quantile smoothing approach to a bivariate setting, introducing a tensor product basis expansion. [39] extended the total variation regularisation method to spatial regression quantiles by combining the so-called penalised triogram basis expansion with a linear programming algorithm. Univariate and bivariate quantile smoothing spline, based on linear programming and interior point methods, are implemented in the `R` package `quantreg` [37]. Alternative approaches to non-linear quantile regression are the local linear estimator by [31, 79], the reproducing kernel Hilbert space estimator by [45], and the non-linear mixed model for clustered data by [27, 28]. More recently, [19] proposed a fast calibrated framework for estimating additive and mixed quantile regression models, which is implemented in the `R` package `qgam` [20]. This leverages and extends the capabilities of

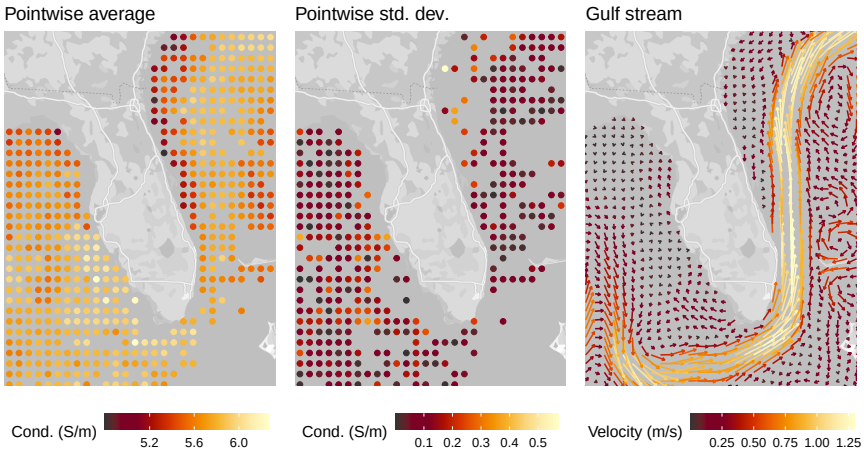

Fig. 2: Sea surface conductivity data around the Florida peninsula. Left: pointwise average of the sea surface conductivity. Middle: pointwise standard deviation of the sea surface conductivity. Right: Gulf stream velocity.

the popular `R` package `mgcv` [77], enabling a wide range of univariate and bivariate smoothers in a quantile regression context, including thin plate spline smoothing [see, e.g., 72, 76] and soap film smoothing [78]. Soap film smoothing, in particular, can handle data over complex domains. Other smoothers having similar features, though not in a quantile regression framework, are the bivariate penalised spline approach by, e.g., [43, 58, 73].

We here propose a nonparametric quantile regression model for spatially referenced data. This model has a regularising term involving a Partial Differential Equation (PDE, [17]) defined over the considered spatial sampling domain. The proposed method generalises the univariate quantile cubic spline [11] to planar regions. Moreover, it extends to conditional quantile estimation, of a spatially distributed response variable, the technique originally introduced for spatial linear regression problems by [5, 64]; see, e.g., the review in [63]. The regularising PDE allows us to incorporate structured smoothing effects in the estimation of the conditional quantile surface, taking into account anisotropy, unidirectional smoothing, and other complex and possibly non-stationary spatial effects, in a unified modelling framework; moreover, being defined on the considered spatial domain, it permits to naturally comply with its geometry. Such a PDE-driven approach also extends the Laplacian regularisation of soap film smoothing [78], while providing a more flexible alternative to thin-plate-spline [72] and multivariate splines [10] in two-dimensional quantile smoothing problems [see, e.g., 63, 64].

Unlike penalised smoothers for generalised linear models, the estimation and theoretical analysis of penalised quantile regression estimators are intrinsically complicated by the non-differentiability of the quantile loss function. This issue prevents the straightforward exploitation of standard methods based on the Taylor expansion, which must be replaced by specifically tailored derivative-free approaches. The novelty of the proposed methodology is threefold. First, we introduce a broad class of flexible, anisotropic and nonstationary, quantile regression models, based on a penalised loss criterion. In doing this, we trade off a goodness-of-fit measure and a roughness penalisation depending on the PDE specification. Secondly, we propose an innovative parameter estimation algorithm, following the Expectation-Maximisation (EM) scheme [13] under interpolation constraints. Such an approach allows to minimise the penalised quantile loss, directly overcoming the non-differentiability of the problem, without relying on linear programming [39] or loss smoothing [19]. The infinite-dimensional solution of such an optimisation is then discretised by means of finite element methods [see, e.g., 57]. Building upon this discretisation, we propose a new constraint-adjusted Generalized Cross-Validation (GCV) criterion, for smoothing parameter selection in penalised quantile regression problems, which accounts for the implicit identity constraints arising from the EM solution. The proposed constrained EM algorithm and constrained adjusted GCV selection constitute non-standard approaches that could also prove useful in more classical quantile regression settings. Finally, we provide a theoretical characterisation of the finite-dimensional PDE quantile estimators, proving existence, consistency, and asymptotic normality.

The article is organized as follows. In Section 2, we introduce the spatial quantile regression model with PDE regularisation, along with the associated infinite-dimensional estimation problem. In Section 3, we propose an appro-

priate functional EM algorithm and characterise its solution at each iteration. In Section 4, we introduce the finite element method to discretise the infinite-dimensional estimation problem. In Section 5, we study the large-sample properties of the resulting finite-dimensional estimator, proving its consistency and asymptotic normality. In Section 6, we present two simulation studies in which we compare our method to state-of-the-art alternatives, under different data generation scenarios. Finally, in Section 7, we employ our method to analyse the rainfall intensity in Switzerland and the sea surface conductivity in the Gulf of Mexico. Concluding remarks are given in Section 8.

2. Spatial quantile regression model

Consider n spatial locations $\mathbf{p}_1, \dots, \mathbf{p}_n$ collected over the bounded region $\Omega \subset \mathbb{R}^2$, with regular boundary $\partial\Omega \in C^2$. At each site $\mathbf{p}_i$ we observe a realisation y_i of a real random variable Y_i, $i \in \{1, \dots, n\}$. We assume Y_i has absolutely continuous distribution with probability density function $\pi_{Y_i|\mathbf{p}_i}(y)$, cumulative density function $\Pi_{Y_i|\mathbf{p}_i}(y)$, and quantile function $Q_{Y_i|\mathbf{p}_i}(\alpha) = \inf\{y \in \mathbb{R} : \Pi_{Y_i|\mathbf{p}_i}(y) \geq \alpha\}$, for any probability level $\alpha \in (0, 1)$. We further assume the following nonparametric spatial model for the α-quantile of Y_i

$$Q_{Y_i|\mathbf{p}_i}(\alpha) = f_\alpha(\mathbf{p}_i), \qquad \mathbf{p}_i \in \Omega, \qquad i \in \{1, \dots, n\}, \tag{1}$$

where $f_\alpha : \Omega \to \mathbb{R}$ is an unknown smooth field. Our aim is to estimate f_α, taking advantage of the available problem-specific information.

We propose to estimate the unknown spatial field f_α by minimising the penalised loss functional

$$J_\alpha(f) = \frac{1}{n}\sum_{i=1}^{n} \rho_\alpha(y_i - f(\mathbf{p}_i)) + \frac{\lambda}{2}\int_\Omega \Big(Lf(\mathbf{p}) - u(\mathbf{p})\Big)^2 \mathrm{d}\mathbf{p}, \tag{2}$$

where $\lambda > 0$ is a smoothing parameter, and $\rho_\alpha(x) = \frac{1}{2}|x| + (\alpha - \frac{1}{2})x$ is the so-called quantile check function, or pinball loss [38]. The penalised objective functional (2) trades off an appropriate goodness-of-fit criterion, the sum of quantile loss functions, and a regularisation criterion, the squared integral of the PDE misfit $Lf(\mathbf{p}) - u(\mathbf{p})$ over the spatial domain Ω, where L is a second-order differential operator and u a smooth squared-integrable function on Ω; see Section 2.1. Such a regularisation term shrinks the estimate of the quantile field toward the solution space of the considered PDE, and can be used to introduce structured smoothing effects, such as anisotropic and unidirectional nonstationary smoothing effects, as detailed in Section 2.1

2.1. Differential regularisation and structured smoothing

We here consider second-order elliptic PDEs of the form $Lf(\mathbf{p}) = u(\mathbf{p})$ in Ω, with squared integrable forcing term $u : \Omega \to \mathbb{R}$ and diffusion-transport-reaction operator L, defined as

$$Lf(\mathbf{p}) = -\mathrm{div}(\mathbf{K}\nabla f(\mathbf{p})) + \mathbf{b} \cdot \nabla f(\mathbf{p}) + cf(\mathbf{p}), \tag{3}$$

where $\mathbf{K} \in \mathbb{R}^{2\times 2}$ denotes a symmetric positive definite diffusion tensor, $\mathbf{b} \in \mathbb{R}^2$ is a transport vector and $c \geq 0$ is a scalar reaction term. Moreover, we denote by $\mathrm{div}(\cdot)$ the divergence of a vector field and by ∇ the gradient of a scalar field, so that

$$\mathrm{div}(\mathbf{K}\nabla f(\mathbf{p})) = \sum_{i,j=1}^{2} K_{ij}\frac{\partial}{\partial p_i}\frac{\partial}{\partial p_j} f(\mathbf{p}), \qquad \mathbf{b} \cdot \nabla f(\mathbf{p}) = \sum_{j=1}^{2} b_j \frac{\partial}{\partial p_j} f(\mathbf{p}).$$

The PDE parameters, namely $\mathbf{K}$, $\mathbf{b}$ and c, model respectively anisotropic diffusion, unidirectional smoothing and pointwise shrinkage toward 0 [see, e.g., 63]. The forcing term instead describes possible source effects, which constitute perturbations from the homogeneous state, corresponding to $u = 0$. Non-stationary effects in space can be considered as well by letting $\mathbf{K} = \mathbf{K}(\mathbf{p})$, $\mathbf{b} = \mathbf{b}(\mathbf{p})$, $c = c(\mathbf{p})$ and $u = u(\mathbf{p})$ vary as smooth functions over the domain Ω, as shown in [1, 5]. For instance, in Section 7, we illustrate a practical case study where a non-stationary transport term is used to account for the Gulf Stream dynamics, while modelling the sea surface conductivity around the Florida peninsula.

The simplest and most common form of differential regularisation can be obtained by considering the pure diffusion equation $\Delta f(\mathbf{p}) = 0$, where the Laplacian operator $\Delta f(\mathbf{p}) = \partial^2 f(\mathbf{p})/\partial p_1^2 + \partial^2 f(\mathbf{p})/\partial p_2^2$ measures of the local

curvature of f. This choice corresponds to a PDE with coefficients $\mathbf{K} = \mathbf{I}$, $\mathbf{b} = \mathbf{0}$, $c = 0$ and $u = 0$. The induced penalisation functional $\int_\Omega (\Delta f(\mathbf{p}))^2 \mathrm{d}\mathbf{p}$ may be seen as a generalisation of the univariate smoothing spline penalties to bivariate smoothing problems [see, e.g., 58, 64, 78]. This relationship between PDE regularised regression and alternative smoothers, such as multivariate splines [10], thin-plate-spline [72] and soap film smoothing [78] are discussed in, e.g., [63].

Additionally, we need to introduce some boundary conditions to complete the specification of the PDE, that is $Bf(\mathbf{p}) = \gamma(\mathbf{p})$ on $\partial\Omega$, where B is an appropriate differential operator, and $\gamma : \partial\Omega \to \mathbb{R}$ is a smooth non-homogeneous term on the boundary. Depending on the problem at hand, Dirichlet, Neumann, Robin, or mixed boundary conditions can be used to model the boundary behaviour of the phenomenon under study; see, e.g., the work of [1, 4, 5] for the use of such conditions on linear regression models with PDE regularisation. Specifically, we here consider Neumann boundary conditions, which are characterised by the differential operator $Bf(\mathbf{p}) = \mathbf{K}\nabla f(\mathbf{p}) \cdot \boldsymbol{\nu}(\mathbf{p})$ on $\partial\Omega$, where $\boldsymbol{\nu}$ denotes the outward normal vector of $\partial\Omega$.

To lighten the notation and improve the readability of the forthcoming mathematical expressions, hereafter we drop the dependence on $\mathbf{p}$ in all the formulas involving differential and integral operators acting on f, or any other spatial field. For instance, we write $\int_\Omega (Lf - u)^2$ in place of $\int_\Omega (Lf(\mathbf{p}) - u(\mathbf{p}))^2 \mathrm{d}\mathbf{p}$.

2.2. Infinite-dimensional estimation problem

We denote by $H^d(\Omega)$ the Sobolev space of order d, that is the space of functions in $L^2(\Omega)$ having d weak derivatives in $L^2(\Omega)$. Formally,

$$H^d(\Omega) = \{f \in L^2(\Omega) : \ D^k f \in L^2(\Omega), \ \forall |k| \leq d\},$$

where $D^k f$ is the k-th weak derivative of f [17]. The appropriate functional embedding for the estimation problem in (2) is the space $\mathcal{F}_\gamma$, defined as

$$\mathcal{F}_\gamma = \{f \in H^2(\Omega) : Bf = \gamma \text{ on } \partial\Omega\}.$$

Because of the Sobolev embedding theorem, any function $f \in \mathcal{F}_\gamma$ is continuous, since $\mathcal{F}_\gamma \subset H^2(\Omega) \subset C^0(\bar{\Omega})$ for $\Omega \subset \mathbb{R}^2$, where $C^0(\bar{\Omega})$ is the set of continuous functions over the closure $\bar{\Omega} = \Omega \cup \partial\Omega$. Therefore, assuming $f \in \mathcal{F}_\gamma$, the pointwise evaluation of f at the observed spatial locations in (2) is well-defined, as well as the differential operator in (3). The estimation problem is thus specified as follows.

Problem 1. Find $\hat{f}_\alpha \in \mathcal{F}_\gamma$ such that $\hat{f}_\alpha \in \operatorname{argmin}_{f \in \mathcal{F}_\gamma} J_\alpha(f)$.

Let us denote by $\mathcal{V}_{\gamma,\alpha} = \{\hat{f}_\alpha \in \mathcal{F}_\gamma : J_\alpha(\hat{f}_\alpha) = \inf_{f \in \mathcal{F}_\gamma} J_\alpha(f)\}$ the space collecting all the fields $\hat{f}_\alpha$ solving Problem 1. The existence of $\mathcal{V}_{\gamma,\alpha}$ is guaranteed by the following proposition.

Proposition 1. *The solution space $\mathcal{V}_{\gamma,\alpha}$ is a non-empty, closed, convex set. Moreover, any field $\hat{f}_\alpha \in \mathcal{V}_{\gamma,\alpha}$ is a global minimum of the functional* (2)*.*

We defer all the derivations and proofs to the Appendix. Differently from the strictly convex optimisation criteria discussed, e.g., by [5, 64] for simple linear models, or by [75] for generalised linear models, in the quantile regression framework here considered, the solution to Problem 1 is not guaranteed to be unique for finite samples. However, it is worth noting that all the estimators minimising (2) are global minimisers and, therefore, reach the same value of the penalised loss functional (2). In the parametric quantile regression literature, this is a known property that naturally arises by exploiting the alternative representation of Problem 1 in terms of linear programming. For details on the parametric context, we refer the reader to [36, 38].

The elements belonging to $\mathcal{V}_{\gamma,\alpha}$ can also be characterised by means of a first-order necessary condition. This leverages on the convexity and continuity of the functional $J_\alpha(f)$ without requiring any further regularity property, such as uniform differentiability of the loss function. The main idea is to exploit the behaviour of directional derivatives, i.e., the subgradient, of functional (2), in a neighbourhood of its minimum, and to use standard convex analysis results to ensure optimality [see, e.g., 60]. Intuitively, any departure from the optimum must yield an increment of the functional leading to a non-negative effect on its directional derivatives. This is formally stated in the following proposition.

Proposition 2. *Let $\hat{f}_\alpha \in \mathcal{V}_{\gamma,\alpha}$ be a minimum of* (2)*. Then, $\hat{f}_\alpha$ satisfies*

$$-\frac{1}{n}\sum_{i=1}^{n}\psi(\mathbf{p}_i)\,\dot{\rho}_\alpha(y_i-\hat{f}_\alpha(\mathbf{p}_i),-\psi(\mathbf{p}_i)) \geq -\lambda\int_\Omega (L\psi)(L\hat{f}_\alpha-u),$$

for any $\psi \in \mathcal{F}_0$, where $\dot{\rho}_\alpha(v,w)$ is the directional derivative of $\rho_\alpha(\cdot)$ calculated in $v \in \mathbb{R}$ along the direction $w \in \mathbb{R}$, which is defined as $\alpha - \mathbb{I}(v<0)$ if $v \neq 0$ and $\alpha - \mathbb{I}(w<0)$ if $v = 0$.

Proposition 2 characterises the elements of the solution space $\mathcal{V}_{\gamma,\alpha}$. However, it does not provide any direct way to find an explicit solution to the minimisation of (2). The quantile estimator in Problem 1 must then be computed via iterative algorithms, like the one proposed in the next section.

3. Functional EM algorithm

We here propose an innovative functional version of the EM algorithm [13, 49] in order to approximate optimisation Problem 1 by a sequence of simpler problems having a convenient variational representation. This provides, in the limit, an approximate solution to the estimation of $\hat{f}_\alpha \in \mathcal{V}_{\gamma,\alpha}$, because of the monotonic convergence property of the EM algorithm.

3.1. Algorithm derivation

We here give a sketch of the derivation of our functional EM algorithm. First, we recall the result by [80], which states the equivalence between the MLE of an Asymmetric Laplace model [41] and the minimum of the quantile check function. Solving Problem 1 then corresponds to maximise the penalised log-likelihood functional under the following specification:

$$Y_i = f(\mathbf{p}_i) + \varepsilon_i, \qquad \varepsilon_i \sim \mathrm{AL}(\alpha,0,1), \qquad i \in \{1,\dots,n\}, \tag{4}$$

where $\mathrm{AL}(\alpha,0,1)$ denotes the Asymmetric Laplace law with shape parameter $\alpha \in (0,1)$, location 0 and scale 1 [see, e.g., 41]. The working probability density function of ε_i is then given by $\pi_{\varepsilon_i|\mathbf{p}_i}(\varepsilon_i) = \alpha(1-\alpha)\exp\{-\rho_\alpha(\varepsilon_i)\}$, which is such that $\Pi_{\varepsilon_i|\mathbf{p}_i}(0) = \int_{-\infty}^{0}\pi_{\varepsilon_i|\mathbf{p}_i}(\varepsilon)\,\mathrm{d}\varepsilon = \alpha$. We take advantage of the location-scale mixture representation of the Laplace distribution [41], which enables us to write the i-th error component in model (4) as a Gaussian random variable with conditional distribution

$$\varepsilon_i|\nu_i \sim \mathrm{N}\left(\frac{(1-2\alpha)\nu_i}{\alpha(1-\alpha)}, \frac{2\nu_i}{\alpha(1-\alpha)}\right), \qquad \nu_i \sim \mathrm{Exp}(1), \tag{5}$$

where $\mathrm{N}(\mu,\sigma^2)$ is the univariate Gaussian law with mean $\mu \in \mathbb{R}$ and variance $\sigma^2 > 0$, while $\mathrm{Exp}(1)$ is the Exponential law with mean 1.

Combining the completed log-likelihood functional relative to model (4) and (5) with the PDE regularisation term in (2), we obtain the penalised completed log-likelihood:

$$\ell_\alpha(f;\boldsymbol{\nu},\mathbf{y}) = \sum_{i=1}^{n}\ell_\alpha(f;\nu_i,y_i) - \frac{\lambda n}{2}\int_\Omega (Lf-u)^2, \tag{6}$$

where the i-th contribution of the completed log-likelihood is equal to

$$\ell_\alpha(f;\nu_i,y_i) = -\frac{1}{2}\log\nu_i - \nu_i - \frac{\alpha(1-\alpha)}{4\nu_i}\left\{y_i - f(\mathbf{p}_i) - \frac{(1-2\alpha)}{\alpha(1-\alpha)}\nu_i\right\}^2.$$

Provided that $\exp\{\ell_\alpha(f;y_i)\} \propto \int_0^\infty \exp\{\ell_\alpha(f;\nu_i,y_i)\}\,\mathrm{d}\nu_i$ [see, e.g., 41], the penalised Laplace log-likelihood (6) can be maximised via an EM algorithm, iterating until convergence the Expectation (E) and Maximisation (M) steps. At the $(k+1)$-th iteration of the procedure, the parameter estimates are updated as follows:

$$\text{E-step:}\quad \bar{\ell}_\alpha^{(k)}(f;\mathbf{y}) = \mathrm{E}^{(k)}\{\ell_\alpha(f;\boldsymbol{\nu},\mathbf{y})\}, \qquad \text{M-step:}\quad f_\alpha^{(k+1)} = \operatorname*{argmax}_{f\in\mathcal{F}_\gamma}\ \bar{\ell}_\alpha^{(k)}(f;\mathbf{y}).$$

The expectation in the E-step is taken with respect to the conditional distribution of $\nu_1, \ldots, \nu_n$, given the observations $y_1, \ldots, y_n$, keeping fixed the values of the spatial field $f_\alpha^{(k)}$, obtained at the k-th iteration of the algorithm.

As shown by, e.g., [42, 71], a posteriori each latent factor ν_i is independently distributed as a Generalized Inverse Gaussian random variable with known parameters. We can thus use the linearity of the expected value, exploit the analytic integration of ν_i, and discard all the additive terms not depending on the unknown f, to obtain a closed-form expression for the expected penalised log-likelihood functional $\bar{\ell}_\alpha^{(k)}(f; \mathbf{y})$. Hence, at the $(k+1)$-th iteration of the algorithm, we have $\bar{\ell}_\alpha^{(k)}(f; \mathbf{y}) = -nJ_\alpha^{(k)}(f) + \text{const}$, where $J_\alpha^{(k)}(f)$ can be interpreted as a local quadratic approximation of $J_\alpha(f)$ calculated in a neighborhood of $f_\alpha^{(k)}$, that is

$$J_\alpha^{(k)}(f) = \frac{1}{2n}(\mathbf{z}^{(k)} - \mathbf{f}_n)^\top \mathbf{W}^{(k)}(\mathbf{z}^{(k)} - \mathbf{f}_n) + \frac{\lambda}{2}\int_\Omega (Lf - u)^2, \tag{7}$$

where $\mathbf{z}^{(k)} = \mathbf{y} - (1-2\alpha)|\mathbf{y} - \mathbf{f}_{\alpha,n}^{(k)}|$ is a vector of working observations, $\mathbf{W}^{(k)} = \text{diag}(\mathbf{w}^{(k)})$ is a working weighting matrix and $1/\mathbf{w}^{(k)} = 2|\mathbf{y} - \mathbf{f}_{\alpha,n}^{(k)}|$. Hereafter we use the notation $\mathbf{f}_n = (f(\mathbf{p}_1), \ldots, f(\mathbf{p}_n))^\top$ to indicate the vector containing the evaluation of any spatial field f at the n spatial locations $\mathbf{p}_1, \ldots, \mathbf{p}_n$.

It is worth noting that functional (7) in the M-step of the algorithm is well-defined only in the case where all elements of the vector $1/\mathbf{w}^{(k)}$ are non-zero. Otherwise, the calculation of $\mathbf{W}^{(k)}$ might produce diagonal entries approaching ∞. Moreover, the M-step still involves an optimisation over an infinite-dimensional space. However, differently from functional (2) in Problem 1, functional (7) is convex and quadratic, and can thus be minimised by appropriately extending the procedure discussed in [5] and [64] for simple linear regression settings. The infinite-weight issue and the characterisation of the infinite-dimensional formulation of the M-step of the algorithm are the main topics of the next section.

As a final remark, we observe that the value of the penalised objective functional (2) decreases at each iteration of the EM algorithm, or at least it does not increase, yielding a stable monotonic convergence to a stationary point. Any solution thus belongs to $\mathcal{V}_\gamma$, thanks to the convexity and coercitivity of functional (2). See, e.g., [44, Chapters 9 and 12], for more details.

3.2. Constrained formulation

In the previous section, we introduced an infinite-weight problem: indeed, there exists the possibility that some elements of $1/\mathbf{w}^{(k)}$ numerically approach zero. As a consequence, the corresponding diagonal entries of $\mathbf{W}^{(k)}$ might diverge, causing numerical instabilities in the calculation and optimisation of functional (7). Such a behaviour has already been observed in the analysis of Bayesian support vector machines by [56]. As argued also in [56], values of $1/\mathbf{w}^{(k)}$ numerically close to 0 do not indicate a pathological behaviour of the algorithm. Actually, they arise in correspondence of support vector points, that are observations for which the complementary slackness conditions are active constraints in the Karush–Kuhn–Tucker formulation of support vector machines [see, e.g., 8, Chapter 7]. See Section 3 of [56] for more details. In the quantile regression context, we observe a similar phenomenon, that is related to the activation of some implicit constraints that arise when Problem 1 is rephrased in terms of linear programming [see, e.g., 36]. Intuitively, if $1/w_i^{(k)} \to 0$, then the i-th squared pseudo-residual $w_i^{(k)}\{z_i^{(k)} - f(\mathbf{p}_i)\}^2$ will receive an infinite weight since $w_i^{(k)} \to \infty$. This behaviour implicitly enforces the constraint $z_i^{(k)} - f(\mathbf{p}_i) = 0$ to be satisfied. A way to solve this issue, and overcome potential numerical instabilities, is to explicitly take into account the constraint induced by $1/w_i^{(k)} = 0$. To this end, we decouple the contribution of the unconstrained and constrained parts of the optimisation, defining the partition $\mathbf{z}^{(k)} = \{\mathbf{z}_s^{(k)}, \mathbf{z}_{-s}^{(k)}\}$, where $\mathbf{z}_s^{(k)} = \{z_i^{(k)} : 1/w_i^{(k)} = 0\}$ and $\mathbf{z}_{-s}^{(k)} = \{z_i^{(k)} : 1/w_i^{(k)} > 0\}$. Under this setting, the objective functional (7) becomes

$$J_{-s}^{(k)}(f) = \frac{1}{2n}(\mathbf{z}^{(k)} - \mathbf{f}_n)_{-s}^\top \mathbf{W}_{-s}^{(k)}(\mathbf{z}^{(k)} - \mathbf{f}_n)_{-s} + \frac{\lambda}{2}\int_\Omega (Lf - u)^2,$$

which is always well-defined. Moreover, we define the constraint $(\mathbf{z}^{(k)} - \mathbf{f}_n)_s = \mathbf{0}$ and we suppress the α-index in the objective functional to lighten the notation. Then, the M-step of the EM algorithm can be alternatively stated as follows.

Problem 2. Find $\tilde{f}_\alpha \in \mathcal{F}_\gamma$ such that $\tilde{f}_\alpha = \text{argmin}_{f \in \mathcal{F}_\gamma} J_{-s}^{(k)}(f)$ subject to $(\mathbf{z}^{(k)} - \tilde{\mathbf{f}}_n)_s = \mathbf{0}$.

Because of the Lagrange multiplier theorem [see, e.g., 52, Chapter 13], searching a solution to Problem 2 is equivalent to minimising the Lagrangian functional

$$\mathcal{L}_\alpha^{(k)}(f, \boldsymbol{\eta}) = J_{-s}^{(k)}(f) + \boldsymbol{\eta}^\top (\mathbf{z}^{(k)} - \mathbf{f}_n)_s / n, \tag{8}$$

where $\boldsymbol{\eta} \in \mathbb{R}^{|s|}$ is a vector of Lagrange multipliers and $|s|$ is the number of active constraints. Hence, we can directly optimise (8) with respect to $\boldsymbol{\eta}$ and f, without imposing any explicit constraint.

In order to characterise the minimum of (8) in a variational formulation, we introduce the field $g = Lf - u \in L^2(\Omega)$, which represents the misfit of the PDE. We define the bilinear forms $R_1(\cdot,\cdot)$ and $R_0(\cdot,\cdot)$ and the linear operator $F(\cdot)$ as follows

$$R_1(\phi, \psi) = \int_\Omega [(\mathbf{K}\nabla\phi) \cdot \nabla\psi + (\mathbf{b} \cdot \nabla\phi)\psi + c\phi\psi], \quad R_0(\phi, \psi) = \int_\Omega \phi\psi, \quad F(\phi) = \int_\Omega u\phi + \int_{\partial\Omega} \gamma\phi, \tag{9}$$

for any pair of functions $\phi, \psi \in H^1(\Omega)$. As we mention in Section 2, we here assume for simplicity Neumann boundary conditions, but similar formulas arise when Dirichlet or Robin conditions are imposed; see, e.g., the works of [1, 4, 5] for the linear regression case. Assuming that all these quantities are well-defined, the minimiser of the Lagrangian functional (8) satisfies the following proposition.

Proposition 3. *Let $(\tilde{f}_\alpha, \tilde{\boldsymbol{\eta}}_\alpha) \in \mathcal{F}_\gamma \times \mathbb{R}^{|s|}$ be a minimum of $\mathcal{L}_\alpha^{(k)}(f, \boldsymbol{\eta})$ in (8) and let $\tilde{g}_\alpha = L\tilde{f}_\alpha - u \in H^1(\Omega)$. Then $(\tilde{f}_\alpha, \tilde{g}_\alpha, \tilde{\boldsymbol{\eta}}_\alpha)$ is the solution of the following system of first-order equations*

$$(\boldsymbol{\psi}_n)_{-s}^\top \mathbf{W}_{-s}^{(k)} (\tilde{\mathbf{f}}_{\alpha,n})_{-s} - (\boldsymbol{\psi}_n)_s^\top \tilde{\boldsymbol{\eta}}_\alpha + \lambda n R_1(\psi, \tilde{g}_\alpha) = (\boldsymbol{\psi}_n)_{-s}^\top \mathbf{W}_{-s}^{(k)} \mathbf{z}_{-s}^{(k)}, \quad R_1(\tilde{f}_\alpha, \phi) - R_0(\tilde{g}_\alpha, \phi) = F(\phi), \quad (\mathbf{z}^{(k)} - \tilde{\mathbf{f}}_{\alpha,n})_s = 0, \tag{10}$$

for any pair of test functions $\phi, \psi \in \mathcal{F}_0$.

Such a variational formulation of Problem 2 is the cornerstone to the derivation of the finite element discretisation we propose in Section 4. In order to lighten the notation hereafter we drop the index (k) associated with the iteration counter of the algorithm.

4. Finite element discretisation

In order to tackle the infinite-dimensional problem as stated in the weak formulation (10), we study a numerical solution within a proper finite-dimensional subspace. We consider a regular triangularisation $\mathcal{T}_h$ of the original spatial domain, with characteristic size h, where h is the maximum length of the triangle edges. In this way, Ω can be represented by the union of all triangles in $\mathcal{T}_h$, leading to the approximated domain Ω_h with polygonal boundary $\partial\Omega_h$. The discretisation $\mathcal{T}_h$, also called mesh, is a fundamental tool in numerical analysis and engineering, that permits to describe the geometry of possibly very complex domains, characterized by strong concavities, holes, or a curved nature.

4.1. Finite element basis expansion

Denote by $\mathcal{P}_r(\boldsymbol{\tau})$ the space of polynomial functions of order r over the triangle $\boldsymbol{\tau} \in \mathcal{T}_h$ and define by $\mathcal{F}_{\gamma,h}^r \subset H^1(\Omega_h) \cap C^0(\bar{\Omega}_h)$ the finite-dimensional subspace

$$\mathcal{F}_{\gamma,h}^r = \{f_h \in C^0(\bar{\Omega}_h) : \ f_h|_{\boldsymbol{\tau}} \in \mathcal{P}_r(\boldsymbol{\tau}) \ \ \forall \boldsymbol{\tau} \in \mathcal{T}_h, \ \ Bf_h = \gamma_h \text{ on } \partial\Omega_h\}.$$

where γ_h is the local r-th order polynomial interpolation of γ. Starting from the triangular discretisation $\mathcal{T}_h$, we can thus define locally supported polynomial functions that provide a basis $\psi_1, \ldots, \psi_{N_h}$ for $\mathcal{F}_{\gamma,h}^r$. If we consider piecewise linear functions, the elements of the basis expansion $\psi_1, \ldots, \psi_{N_h}$ have a one-to-one correspondence with the nodes of the mesh $\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_{N_h}$, that are the vertices of the triangles. The evaluation of the i-th basis on the j-th node is given by $\psi_i(\boldsymbol{\xi}_j) = \delta_{ij}$, where δ_{ij} is the Kronecker delta with $\delta_{ij} = 1$ if $i = j$ and $\delta_{ij} = 0$ if $i \neq j$, for any $i, j \in \{1, \ldots, N_h\}$. It turns out that any surface $f_h \in \mathcal{F}_{\gamma,h}^1$ is uniquely determined by its values at the nodes:

$$f_h(\mathbf{p}) = \sum_{j=1}^{N_h} f_h(\boldsymbol{\xi}_j)\, \psi_j(\mathbf{p}) = \mathbf{f}^\top \boldsymbol{\psi}(\mathbf{p}),$$

where $\boldsymbol{\psi}(\mathbf{p}) = (\psi_1(\mathbf{p}), \ldots, \psi_{N_h}(\mathbf{p}))^\top$ is the basis vector, evaluated at the point $\mathbf{p} \in \Omega$, while $\mathbf{f} = (f_h(\boldsymbol{\xi}_1), \ldots, f_h(\boldsymbol{\xi}_{N_h}))^\top$ is the coefficient vector of the basis expansion. It is worth highlighting that the mesh can be constructed independently of the n data locations $\mathbf{p}_1, \ldots, \mathbf{p}_n$. For instance, it is possible to consider a regular mesh, also when the data locations are not uniformly distributed over the spatial domain Ω. In practical applications of the method, we consider rich meshes, having a resolution fine enough to permit an accurate description of global and local characteristics of the underlying spatial signal. Consequently, we often use discretisations where some triangles do not contain any observation, as illustrated in Fig. 9. Notice that this does not invalidate the method and the well-posedness of the estimation problem holds also in this case; see Proposition 4 and Corollary 1 in Section 4.2. We refer the reader to [57] for a comprehensive introduction to finite element methods and numerical discretisation of partial differential problems.

4.2. Finite-dimensional estimator

Let $\boldsymbol{\Psi}$ be the matrix evaluation of the N_h basis functions at the n data locations:

$$\boldsymbol{\Psi} = \begin{bmatrix} \boldsymbol{\psi}(\mathbf{p}_1)^\top \\ \vdots \\ \boldsymbol{\psi}(\mathbf{p}_n)^\top \end{bmatrix} = \begin{bmatrix} \psi_1(\mathbf{p}_1) & \cdots & \psi_{N_h}(\mathbf{p}_1) \\ \vdots & & \vdots \\ \psi_1(\mathbf{p}_n) & \cdots & \psi_{N_h}(\mathbf{p}_n) \end{bmatrix},$$

so that, for any $f_h \in \mathcal{F}^1_{\gamma,h}$, we have $\mathbf{f}_n = \boldsymbol{\Psi}\mathbf{f}$. Let $\mathbf{R}_1$ be the $N_h \times N_h$ matrix evaluation of the bilinear form $R_1(\cdot,\cdot)$ over the finite element basis: $\mathbf{R}_1 = \int_{\Omega_h} [(\nabla\boldsymbol{\psi})\mathbf{K}(\nabla\boldsymbol{\psi})^\top + (\nabla\boldsymbol{\psi})(\mathbf{b}\boldsymbol{\psi}^\top) + c\boldsymbol{\psi}\boldsymbol{\psi}^\top]$, where $\nabla\boldsymbol{\psi}$ is the Jacobian matrix of the vector field $\boldsymbol{\psi}$. Similarly, we define the mass matrix $\mathbf{R}_0 = \int_{\Omega_h} \boldsymbol{\psi}\boldsymbol{\psi}^\top$, being the discretisation of $R_0(\cdot,\cdot)$. The vectors $\mathbf{u} = \int_{\Omega_h} u_h\boldsymbol{\psi}$ and $\boldsymbol{\gamma} = \int_{\partial\Omega_h} \gamma_h\boldsymbol{\psi}$ are used to build a discretisation of $F(\cdot)$. The finite element approximation of the system (10) in Proposition 3 is presented in the following proposition.

Proposition 4. *The finite element estimator* $(\tilde{\mathbf{f}}_\alpha, \tilde{\mathbf{g}}_\alpha, \tilde{\boldsymbol{\eta}}_\alpha) \in \mathbb{R}^{N_h} \times \mathbb{R}^{N_h} \times \mathbb{R}^{|s|}$ *is the solution of the linear system*

$$\begin{bmatrix} \boldsymbol{\Psi}_{-s}^\top\mathbf{W}_{-s}\boldsymbol{\Psi}_{-s} & \lambda n\mathbf{R}_1^\top & \boldsymbol{\Psi}_s^\top \\ \lambda n\mathbf{R}_1 & -\lambda n\mathbf{R}_0 & \mathbf{O} \\ \boldsymbol{\Psi}_s & \mathbf{O} & \mathbf{O} \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{f}}_\alpha \\ \tilde{\mathbf{g}}_\alpha \\ \tilde{\boldsymbol{\eta}}_\alpha \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Psi}_{-s}^\top\mathbf{W}_{-s}\mathbf{z}_{-s} \\ \lambda n(\mathbf{u}+\boldsymbol{\gamma}) \\ \mathbf{z}_s \end{bmatrix}. \tag{11}$$

Set $\mathbf{A} = \boldsymbol{\Psi}_{-s}^\top\mathbf{W}_{-s}\boldsymbol{\Psi}_{-s} + \lambda n\mathbf{R}_1^\top\mathbf{R}_0^{-1}\mathbf{R}_1$. Note that $\mathbf{A}$ is non-singular under minimal regularity conditions [4]. Set $\mathbf{B} = \boldsymbol{\Psi}_s\mathbf{A}^{-1}\boldsymbol{\Psi}_s^\top$ and $\mathbf{d} = \boldsymbol{\Psi}_{-s}^\top\mathbf{W}_{-s}\mathbf{z}_{-s} + \lambda n\mathbf{R}_1^\top\mathbf{R}_0^{-1}(\mathbf{u}+\boldsymbol{\gamma})$.

Corollary 1. *The minimiser* $(\tilde{\mathbf{f}}_\alpha, \tilde{\mathbf{g}}_\alpha, \tilde{\boldsymbol{\eta}}_\alpha)$ *that solves the linear system* (11) *exists unique. Moreover, it has the closed-form solution*

$$\begin{aligned} \tilde{\mathbf{f}}_\alpha &= \mathbf{A}^{-1}\mathbf{d} + \mathbf{A}^{-1}\boldsymbol{\Psi}_s^\top\mathbf{B}^{-1}(\mathbf{z}_s - \boldsymbol{\Psi}_s\mathbf{A}^{-1}\mathbf{d}), \\ \tilde{\mathbf{g}}_\alpha &= \mathbf{R}_0^{-1}(\mathbf{R}_1\tilde{\mathbf{f}}_\alpha - \mathbf{u} - \boldsymbol{\gamma}), \\ \tilde{\boldsymbol{\eta}}_\alpha &= \mathbf{B}^{-1}(\mathbf{z}_s - \boldsymbol{\Psi}_s\mathbf{A}^{-1}\mathbf{d}). \end{aligned} \tag{12}$$

Remark 1 (Projection onto the constrained space)**.** It is worth noting that $\mathbf{A}^{-1}\mathbf{d}$ is the solution of the unconstrained optimisation problem $\min\{J^{(k)}_{-s}(\mathbf{f}^\top\boldsymbol{\psi}) : \mathbf{f} \in \mathbb{R}^{N_h}\} = \min\{J^{(k)}_{-s}(f_h) : f_h \in \mathcal{F}^1_{\gamma,h}\}$. The constrained estimator $\tilde{\mathbf{f}}_\alpha$ is then obtained by projecting the unconstrained solution $\mathbf{A}^{-1}\mathbf{d}$ onto the constrained space $\{\mathbf{f} \in \mathbb{R}^{N_h} : \mathbf{z}_s - \boldsymbol{\Psi}_s\mathbf{f} = \mathbf{0}\}$.

Remark 2 (Convergence to the global solution)**.** Because of the monotonic convergence of the EM algorithm [44, Chapters 9 and 12], in the limit, $k \to \infty$, the finite element estimator $\tilde{f}_{\alpha,h} = \tilde{\mathbf{f}}_\alpha^\top\boldsymbol{\psi}$ converges to $\hat{f}_{\alpha,h} = \hat{\mathbf{f}}_\alpha^\top\boldsymbol{\psi}$, which is the minimiser of $J_\alpha(f_h)$ for $f_h \in \mathcal{F}^1_{\gamma,h}$.

Remark 3 (Inclusion of covariates)**.** The above results, and in particular Propositions 3 and 4, may be easily generalised to accommodate for additive and semiparametric models when also a set of fixed covariates is available.

Remark 4 (Computational complexity)**.** In (11), the computation of the system vector and matrix, and the numerical solution of the system require, respectively, $O(nN_h^{\delta/2})$, $O(nN_h^\delta)$ and $O(N_h^{1+\delta})$ floating point operations, where the coefficient $\delta \in [0,2]$ is determined by the degree of sparsity of the matrices $\boldsymbol{\Psi}$, $\mathbf{R}_0$ and $\mathbf{R}_1$. In applications, the local support of the finite element basis $\psi_1, \ldots, \psi_{N_h}$ guarantees for the system (11) high levels of sparsity, leading to an overall computational complexity approaching the lower bound $O(n) + O(N_h)$, with δ tending to 0. Refer to [2] for detailed discussions and extensive simulations on numerical aspects of spatial regression with PDE regularisation.

4.3. Smoothing parameter selection

A classical issue in penalised nonparametric regression modelling is the selection of the smoothing parameter λ, which controls the amount of regularisation enforced in the estimates. Common selection methods proposed in the quantile regression literature rely on approximate Generalized Cross-Validation (GCV) minimization [53, 54, 67, 82], likelihood maximisation [50, 59], iterative Laplace approximation [28] and Bayesian belief updating [19].

In this work, we select the value of λ that minimises the approximated GCV score. We rely on the definition of approximate GCV for quantile regression problems used by [45] and [82], which built over the original proposal of [54]. In our case, the quantile GCV takes the form

$$\mathrm{GCV}(\lambda) = \sum_{i=1}^{n} \frac{\rho_\alpha(y_i - \hat{f}_{\alpha,h}(\mathbf{p}_i))}{n - \mathrm{df}_\alpha}, \tag{13}$$

where df_α denotes a measure of the effective degrees of freedom induced by the smoother $\hat{f}_{\alpha,h}$. Notice that the GCV score (13) depends on the smoothing parameter through $\hat{f}_{\alpha,h} \equiv \hat{f}_{\alpha,h}(\lambda)$ and $\mathrm{df}_\alpha \equiv \mathrm{df}_\alpha(\lambda)$, which are implicit functions of λ.

In order to derive a closed-form expression for the effective degrees of freedom, as defined by [54], we first need to provide a convenient representation for the estimated surface at the n observed locations, $\hat{\mathbf{f}}_{\alpha,n}$. To this purpose, in Proposition 5 we show that, at the end of the optimisation, the discrete quantile estimator $\hat{\mathbf{f}}_\alpha$ in (12) can be written as a linear transformation of the pseudo-data vector $\mathbf{z}$. Then, we explore the linearity of $\hat{\mathbf{f}}_{\alpha,n}$ to derive the explicit expression of the equivalent degrees of freedom.

Define the matrix $\mathbf{S} = [\,\mathbf{S}_{-s},\ \mathbf{S}_s\,]$, where $\mathbf{S}_{-s} = \boldsymbol{\Psi}(\mathbf{I} - \mathbf{A}^{-1}\boldsymbol{\Psi}_s^\top\mathbf{B}^{-1}\boldsymbol{\Psi}_s)\mathbf{A}^{-1}\boldsymbol{\Psi}_{-s}^\top\mathbf{W}_{-s}$ and $\mathbf{S}_s = \boldsymbol{\Psi}\mathbf{A}^{-1}\boldsymbol{\Psi}_s^\top\mathbf{B}^{-1}$, and the vector $\mathbf{r} = \lambda n\boldsymbol{\Psi}\mathbf{A}^{-1}\boldsymbol{\Psi}_s^\top\mathbf{B}^{-1}\boldsymbol{\Psi}_s\mathbf{A}^{-1}\mathbf{R}_1^\top\mathbf{R}_0^{-1}(\mathbf{u} + \boldsymbol{\gamma})$.

Proposition 5. *The discrete quantile estimator at the n observed locations, $\hat{\mathbf{f}}_{\alpha,n}$, is a linear smoother of the pseudo-data vector, $\mathbf{z}$, and can be written as*

$$\hat{\mathbf{f}}_{\alpha,n} = \mathbf{S}\,\mathbf{z} + \mathbf{r} = \mathbf{S}_{-s}\mathbf{z}_{-s} + \mathbf{S}_s\mathbf{z}_s + \mathbf{r}, \tag{14}$$

Moreover, the effective degrees of freedom of $\hat{\mathbf{f}}_{\alpha,n}$ are

$$\mathrm{df}_\alpha = \mathrm{tr}(\mathbf{S}) = |s| + \mathrm{tr}\{\mathbf{A}^{-1}(\mathbf{I} - \boldsymbol{\Psi}_s^\top\mathbf{B}^{-1}\boldsymbol{\Psi}_s\mathbf{A}^{-1})(\boldsymbol{\Psi}_{-s}^\top\mathbf{W}_{-s}\boldsymbol{\Psi}_{-s})\}. \tag{15}$$

Remark 5 (Unconstrained solution)**.** The latter formulation generalises the definition of degrees of freedom for penalised linear smoothers by [72] to penalised linear smoothers subject to equality constraints. Indeed, if no active constraints were present, the set $\{i \in \mathbb{N} : 1/w_i = 0, 1 \le i \le n\}$ would be empty and, therefore, we would obtain $\mathrm{df}_\alpha = \mathrm{tr}\{\mathbf{A}^{-1}(\boldsymbol{\Psi}^\top\mathbf{W}\boldsymbol{\Psi})\}$, which is the standard definition of effective degrees of freedom for penalised weighted regression problems.

Remark 6 (Numerical implementation)**.** Note that $|s| \le N_h$, and, typically, $|s| \ll N_h$. Hence, we can efficiently compute the effective degrees of freedom, df_α, defining $\mathbf{U} = \mathbf{A}^{-1}\boldsymbol{\Psi}_s^\top$ and $\mathbf{V} = \mathbf{A}^{-1}(\boldsymbol{\Psi}_{-s}^\top\mathbf{W}_{-s}\boldsymbol{\Psi}_{-s})$ and using the identity $\mathrm{df}_\alpha = |s| + \mathrm{tr}(\mathbf{V}) - \mathrm{tr}\{\mathbf{U}(\boldsymbol{\Psi}_s\mathbf{U})^{-1}\mathbf{U}^\top(\boldsymbol{\Psi}_{-s}^\top\mathbf{W}_{-s}\boldsymbol{\Psi}_{-s})\}$. The computation of $\mathbf{U}$ via sparse solvers is of order $O(|s|N_h^{1+\delta})$, with $\delta \in [0, 2]$. The inversion of $\boldsymbol{\Psi}_s\mathbf{U}$ requires only $O(|s|^3)$ operations. Finally, we can directly approximate $\mathrm{tr}(\mathbf{V})$ in $O(N_h^{1+\delta})$ operations by using stochastic estimators of the trace without explicitly forming $\mathbf{V}$; see [2] for more details.

5. Large sample properties

We here study the asymptotic properties of the quantile estimator discussed in Section 4. We denote by $f_\alpha \in \mathcal{F}_\gamma$ the true α-quantile field of Y_i given $\mathbf{p}_i$, and we define $g_\alpha = Lf_\alpha - u \in L^2(\Omega)$ as the misfit of the PDE relative to f_α. We denote by $\mathbf{f}_\alpha$ and $\mathbf{g}_\alpha$ the evaluation vectors of f_α and g_α at the mesh knots $\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_{N_h}$. Furthermore, we define the $N_h \times N_h$ matrices

$$\mathbf{D}_{0,n} = \alpha(1-\alpha)\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\psi}(\mathbf{p}_i)\boldsymbol{\psi}(\mathbf{p}_i)^\top, \qquad \mathbf{D}_{1,n} = \frac{1}{n}\sum_{i=1}^{n}\pi_i\,\boldsymbol{\psi}(\mathbf{p}_i)\boldsymbol{\psi}(\mathbf{p}_i)^\top,$$

where the subscript n highlights the dependence on the sample size, while the constants $\pi_i = \pi_{Y_i|\mathbf{p}_i}(\mathbf{f}_\alpha^\top\boldsymbol{\psi}(\mathbf{p}_i))$, $i \in \{1, \ldots, n\}$, define a sequence of weighting values depending on the true underlying distribution of Y_i given $\mathbf{p}_i$.

Assuming that the number of bases N_h and the triangulation $\mathcal{T}_h$ are fixed, a sufficient set of regularity conditions that guarantee the convergence of the discretised estimator to its asymptotic distribution is the following.

Assumption 1. There exist positive constants π_- and π_+, such that $0 < \pi_- < \pi_{Y|\mathbf{p}}(\mathbf{f}_\alpha^\top \boldsymbol{\psi}(\mathbf{p})) < \pi_+ < \infty$ for any $\mathbf{p} \in \Omega$.

Assumption 2. There exists n^* such that for $n > n^*$, $\boldsymbol{\Psi}$ is a full-rank matrix.

Assumption 3. There exist positive definite matrices $\mathbf{D}_0$ and $\mathbf{D}_1$ such that $\mathbf{D}_{0,n} \to \mathbf{D}_0$ and $\mathbf{D}_{1,n} \to \mathbf{D}_1$ as $n \to \infty$.

Assumption 1 ensures the existence of a well-behaved asymptotic estimator, preventing the interpolated quantile field $\mathbf{f}_\alpha^\top \boldsymbol{\psi}$ lying in a region with almost null probability mass. Such a condition is easily verified for any non-degenerate continuous distribution over each point of the domain Ω. Assumption 2, along with Assumption 1, gives a sufficient condition for the non-singularity of the matrices $\mathbf{D}_{0,n}$ and $\mathbf{D}_{1,n}$. For example, Assumption 2 is satisfied when the knots of the triangulation are a subset of the data locations; it is also satisfied when, for $n > n^*$, there is at least one observation in the support of each basis function $\psi_1, \ldots, \psi_{N_h}$. Finally, Assumption 3 guarantees the existence of a non-singular limit for the matrices $\mathbf{D}_{0,n}$ and $\mathbf{D}_{1,n}$. Assumptions 1, 2, and 3 are standard requirements in the asymptotic analysis of quantile estimator [see, e.g., 36]. It is worth highlighting that the above assumptions are not necessary to guarantee the existence and well-posedness of the estimator in finite-sample problems, but only serve for asymptotic analysis. Under Assumptions 1, 2 and 3, we can now study the convergence of the finite-element estimator $\hat{\mathbf{f}}_\alpha$ toward $\mathbf{f}_\alpha$ conditionally on the mesh knots $\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_{N_h}$.

Theorem 1. *Set $\lambda = \lambda_n = O(n^{-1/2})$ with $\sqrt{n}\,\lambda_n \to \bar{\lambda}$ and define $\boldsymbol{\mu} = -\bar{\lambda}\,\mathbf{D}_1^{-1}\mathbf{R}_1^\top \mathbf{g}_\alpha$ and $\boldsymbol{\Sigma} = \mathbf{D}_1^{-1}\mathbf{D}_0\,\mathbf{D}_1^{-1}$. Then, under Assumptions 1, 2 and 3, the α-quantile estimator $\hat{\mathbf{f}}_\alpha$ is such that*

$$\sqrt{n}(\hat{\mathbf{f}}_\alpha - \mathbf{f}_\alpha) \xrightarrow{\mathrm{d}} \mathrm{N}_{N_h}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \qquad \textit{and} \qquad \mathrm{MSE}(\hat{\mathbf{f}}_\alpha) = O(\lambda^2 + n^{-1} + \lambda n^{-1}) = O(n^{-1}).$$

Additionally, if $\lambda = \lambda_n = o(n^{-1/2})$, $\sqrt{n}\,\mathrm{E}(\hat{\mathbf{f}}_\alpha - \mathbf{f}_\alpha) \to 0$.

Theorem 1 establishes the asymptotic normality of the finite-dimensional estimator $\hat{\mathbf{f}}_\alpha$ and it shows that the MSE of $\hat{\mathbf{f}}_\alpha$ reaches the optimal rate of convergence for parametric estimators. Such result can be used to perform Wald-type inference on $f_\alpha(\mathbf{p}) = \mathbf{f}_\alpha^\top \boldsymbol{\psi}(\mathbf{p})$, with the caveat discussed in Section 8.

6. Simulation studies

In this section, we present the results of two simulation studies, in order to assess the performance of the proposed method and to compare it with alternative existing approaches.

We consider heteroscedastic data generating models, defined as

$$Y_i \sim \mathrm{N}(\mu_i, \sigma_i^2), \qquad \mu_i = \mu(\mathbf{p}_i), \qquad \sigma_i = \sigma(\mathbf{p}_i), \qquad \mathbf{p}_i \in \Omega, \tag{16}$$

with $i \in \{1, \ldots, n\}$, and where $\mu : \Omega \to \mathbb{R}$ is the spatial mean surface, and $\sigma : \Omega \to \mathbb{R}_+$ is the standard deviation surface. In the first simulation, we consider three combinations of mean and standard deviation surfaces defined over a non-trivial horseshoe domain. In the second simulation setup, we consider an anisotropic specification of the mean and standard deviation fields, defined upon a simple square domain.

In each scenario, we independently simulate 100 datasets with 500 observations each, according to the generative model (16). We then estimate 5 quantile surfaces, corresponding to levels $\alpha \in \{0.1, 0.25, 0.5, 0.75, 0.9\}$. We compare the following methods:

1. SQR-PDE: the proposed spatial quantile regression with the second order differential regularisation $\int_\Omega \left(\mathrm{div}(\mathbf{K}\nabla f)\right)^2$ and homogeneous Neumann boundary conditions;
2. SOAP: quantile version of the soap film smoothing by [19] and [78];
3. H-SOAP: quantile soap film smoothing with location-scale calibration for the heteroscedasticity by [19];
4. TPS: quantile version of thin plate spline smoothing by [19] and [76];

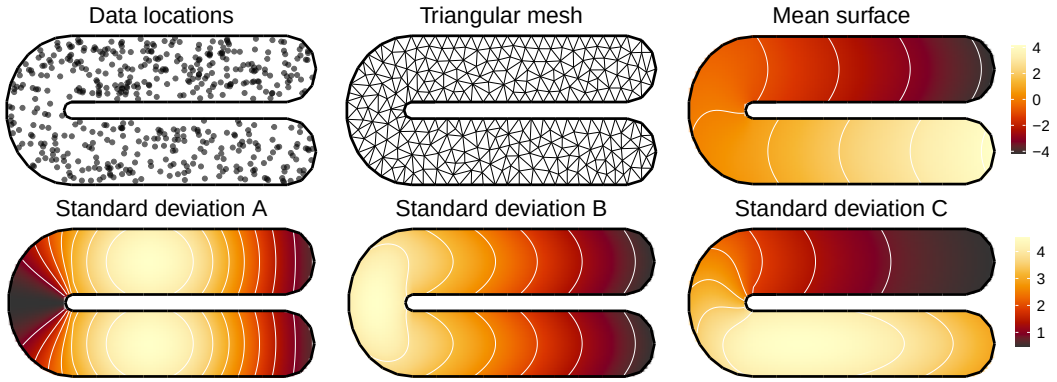

Fig. 3: First simulation setup (Section 6.1). Horseshoe domain for the first simulation study. First row: data locations, triangular discretisation of the domain and mean surface. Second row: standard deviation surfaces corresponding to three simulation settings A, B and C.

5. H-TPS: quantile thin plate spline smoothing with location-scale calibration for the heteroscedasticity by [19];

6. QSS: quantile smoothing splines with total variation regularisation by [39].

SQR-PDE is implemented using the `R` parkage `fdaPDE` [3]; SOAP, H-SOAP, TPS and H-TPS are all implemented in the `R` package `qgam` [20]; QSS is implemented in the `R` package `quantreg` [37].

The methods are compared in terms of Root Mean Squared Error (RMSE), computed as RMSE $= \{\int_\Omega (f_\alpha - \hat{f}_\alpha)^2/|\Omega|\}^{1/2}$, where the integral is approximated by a sum over a fine regular grid that covers the domain Ω, and $|\Omega|$ is the area of Ω.

6.1. First simulation setup: field over irregular domain

In the first simulation, the data is generated according to the heteroscedastic Gaussian model (16) defined over the horseshoe domain proposed by [58]; see Fig. 3. For the mean surface $\mu(\cdot)$ we consider the test function proposed by [78], shown in the bottom left panel of Fig. 3. We moreover consider three different standard deviation surfaces, $\sigma_{\text{A}}(\cdot)$, $\sigma_{\text{B}}(\cdot)$, and $\sigma_{\text{C}}(\cdot)$, shown in the right panels of Fig. 3; this corresponds to three different scenarios A, B, and C, which enable us to explore both symmetric and asymmetric heteroscedasticity schemes. Heterogeneous heteroscedastic patterns can indeed lead to quantile estimation problems with different levels of complexity.

We estimate the quantile fields using the methods mentioned before. Since no anisotropy or flows are appreciable in the synthetic data, we consider SQR-PDE model with regularisation of the isotropic Laplacian, i.e., $\mathbf{K} = \mathbf{I}$.

Fig. 4 shows results in terms of RMSE for the three scenarios (A, B and C) and the five quantile levels (10%, 25%, 50%, 75%, 90%). Fig. 5 displays the estimated 10% quantile field for all the considered estimating methods but H-SOAP and H-TPS, in one replication of the simulation experiments under setting C. We excluded H-SOAP and H-TPS from Fig. 5, since they always provide estimated surfaces very similar to SOAP and TPS, respectively, and indeed they also reach similar RMSE values, as displayed in Fig. 4. The proposed SQR-PDE has the best performance in terms of RMSE, as shown in Fig. 4. The second best model is quantile soap film smoothing. Both these methods outperform quantile thin-plate-spline and quantile smoothing splines in terms of precision. This effect is due to proper management of the domain constraints by the proposed SQR-PDE and by quantile soap film smoothing. Quantile thin-plate splines and quantile smoothing splines, instead, smooth the quantile surface not only within the domain but also across boundaries, causing undesired artefacts in a neighbourhood of the interior boundary, as it is displayed in Fig. 5.

As for the average execution time, SQR-PDE, TPS and H-TPS take around 1 second per replicate, while SOAP, H-SOAP and QSS require about 10 seconds. These times include both estimation and smoothing parameter selection. Further numerical experiments, not reported here for the sake of space, suggest a scalable behaviour of the computational time of SQR-PDE as the sample size increases. All the computations are performed on a Dell XPS 15 laptop with 4.7 gigahertz processor and 32 gigabytes of random access memory.

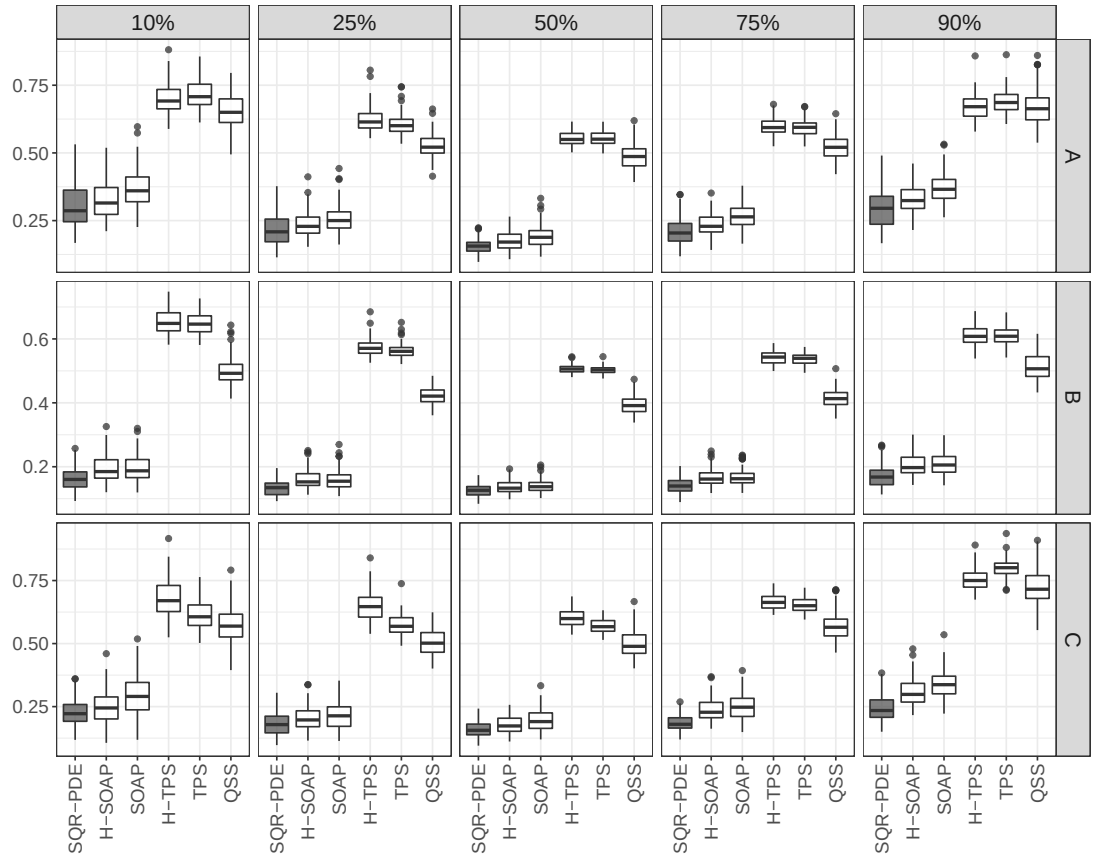

Fig. 4: First simulation setup (Section 6.1). Boxplots of the RMSE over 100 replicates of the estimated spatial quantile fields. Each column corresponds to a different quantile level (10%, 25%, 50%, 75%, 90%). Each row corresponds to a simulation setting (A, B, C).

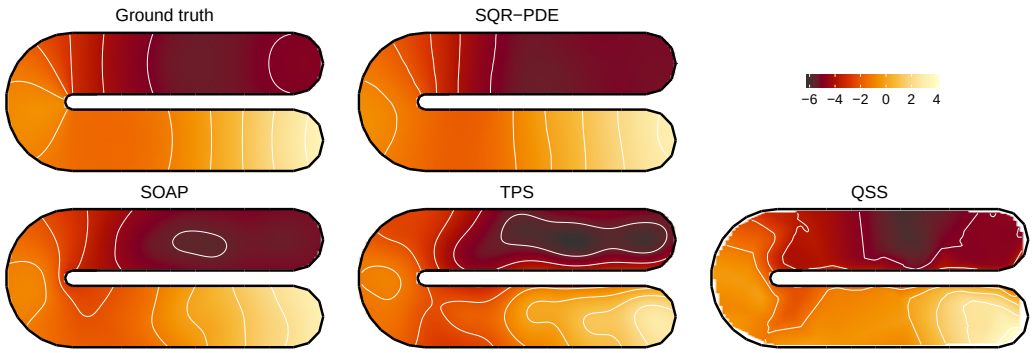

Fig. 5: First simulation setup (Section 6.1). Estimated 10% quantile fields over the horseshoe domain from simulation setting C.

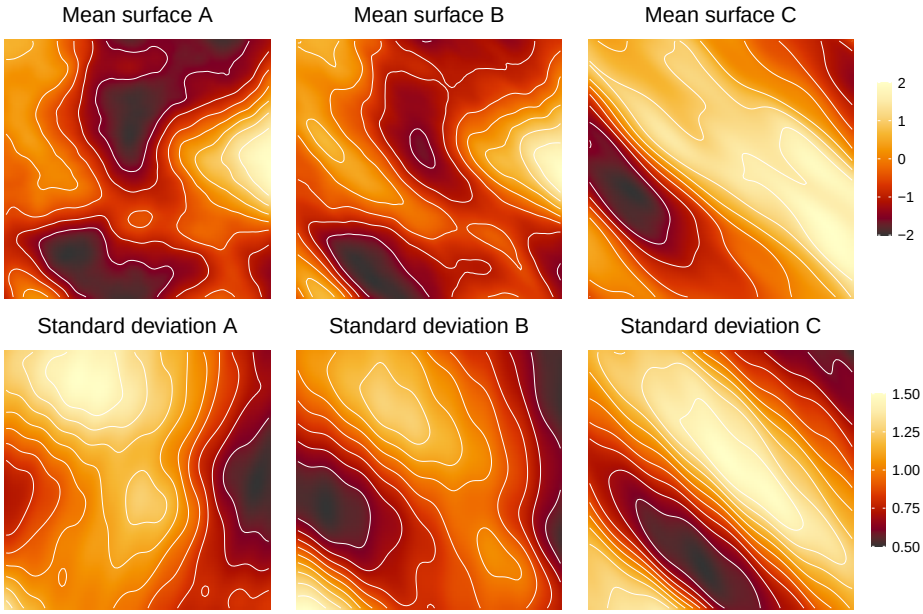

Fig. 6: Second simulation setup (Section 6.2). First row: mean surfaces. Second row: standard deviation surfaces. From left to right: scenarios A, B and C.

6.2. Second simulation setup: anisotropic field over regular domain

In the second simulation, we generate both the mean $\mu(\cdot)$ and the standard deviation $\sigma(\cdot)$ as Gaussian random fields with anisotropic covariance function over the square domain $\Omega = [0,1]^2$. In particular, we consider a Matérn covariance function with marginal variance $\tau^2 > 0$, covariance range $\rho > 0$, and smoothing parameter $\nu > 0$, given by $C(d_{ij}; \tau, \rho, \nu) = \tau^2(2^{1-\nu}/\Gamma(\nu))(2\sqrt{\nu}\, d_{ij}/\rho)^{\nu} K_{\nu}(2\sqrt{\nu}\, d_{ij}/\rho)$, where $d_{ij}^2 = (\mathbf{p}_i - \mathbf{p}_j)^{\top}\mathbf{D}\,(\mathbf{p}_i - \mathbf{p}_j)$ is the squared Mahalanobis distance between $\mathbf{p}_i$ and $\mathbf{p}_j$ calculated with respect to the anisotropy tensor $\mathbf{D} \in \mathbb{R}^{2\times 2}$, which determines the direction and intensity of the anisotropy. For the mean surface, we set $\tau^2 = 1$, $\rho = 0.3$, and $\nu = 2$. For the standard deviation surface, we set $\tau^2 = 0.3$, $\rho = 0.6$, and $\nu = 2$. The anisotropic tensor is specified according to the form: $D_{ij} = 1$ if $i = j$ and $D_{ij} = \phi$ if $i \neq j$, where the parameter $\phi \in (-1, +1)$ controls the direction and intensity of the anisotropic effect. Specifically, we consider three scenarios, A, B and C, with an increasing level of anisotropy, namely $\phi_{\mathrm{A}} = 0$, $\phi_{\mathrm{B}} = -0.3$, $\phi_{\mathrm{C}} = -0.6$.

We use the function `RFsimulate` of the `R` package `RandomFields` [65] to simulate the mean and standard deviation fields on a fine grid covering Ω; see Fig. 6. Estimation of quantile fields is performed with the same methods considered in the first simulation setup. In particular, for SQR-PDE, we consider two specifications for the differential regularisation:

7. SQR-PDE($\mathbf{I}$): spatial quantile regression with Laplacian regularisation, i.e., $\mathbf{K} = \mathbf{I}$;
8. SQR-PDE($\mathbf{K}$): spatial quantile regression with anisotropic regularisation, i.e. $\mathbf{K} \neq \mathbf{I}$.

In order to determine the optimal diffusion tensor $\mathbf{K}$ for SQR-PDE($\mathbf{K}$), we use the parameter cascading algorithm proposed by [6] in the context of spatial smoothing with anisotropic PDE regularisation. The complete estimation procedure for the anisotropic quantile model involves two steps. In the first step, we estimate the diffusion tensor $\mathbf{K}$ by minimising a squared error loss criterion [6] using the `R` package `fdaPDE` [3]. At this stage of the procedure, no quantile regression is considered, assuming that the anisotropic effect is homogeneous over the mean and all quantile surfaces. In the second step, we estimate the quantile field using the anisotropic penalty induced by the tensor $\mathbf{K}$, estimated in the first step.

Fig. 7 shows the RMSE of the various methods for the three considered scenarios (A, B and C) and different quantile levels (10%, 25%, 50%, 75%, 90%). As anisotropy increases, SQR-PDE models improve their performances

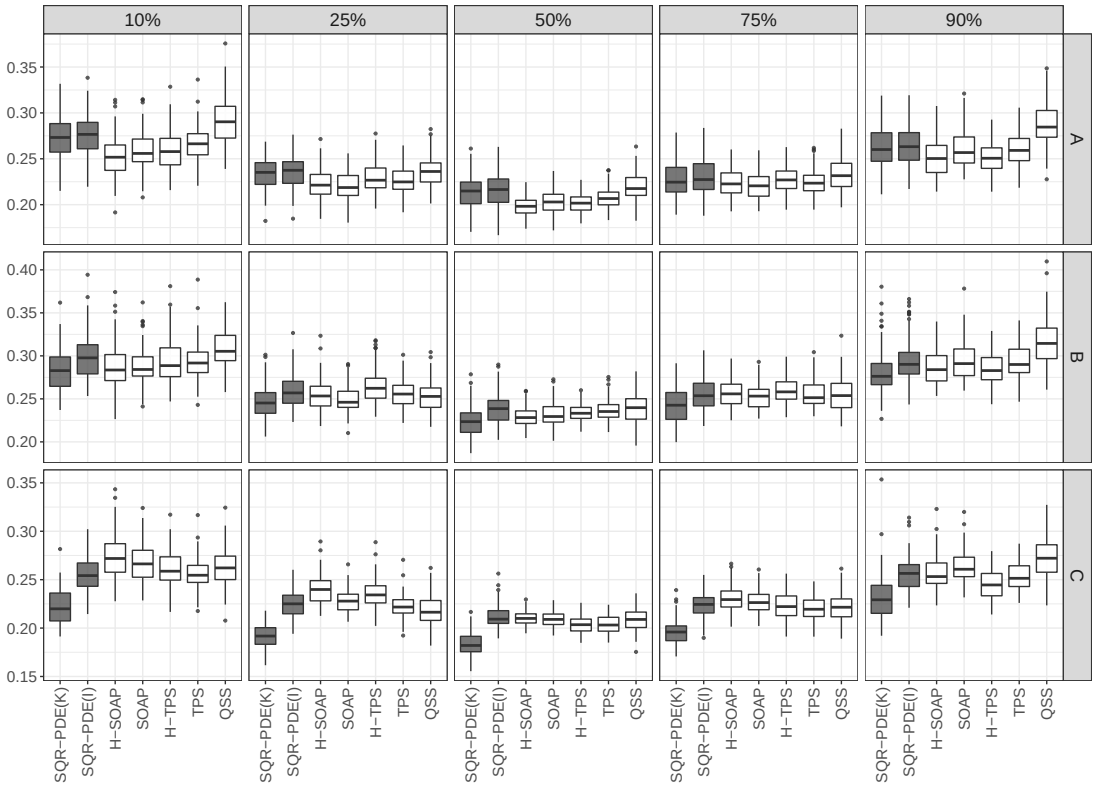

Fig. 7: Second simulation setup (Section 6.2). Boxplots of the RMSE over 100 replicates of the estimated spatial quantile fields. Each column corresponds to a quantile level (10%, 25%, 50%, 75%, 90%). Each row corresponds to a simulation setting (A, B, C).

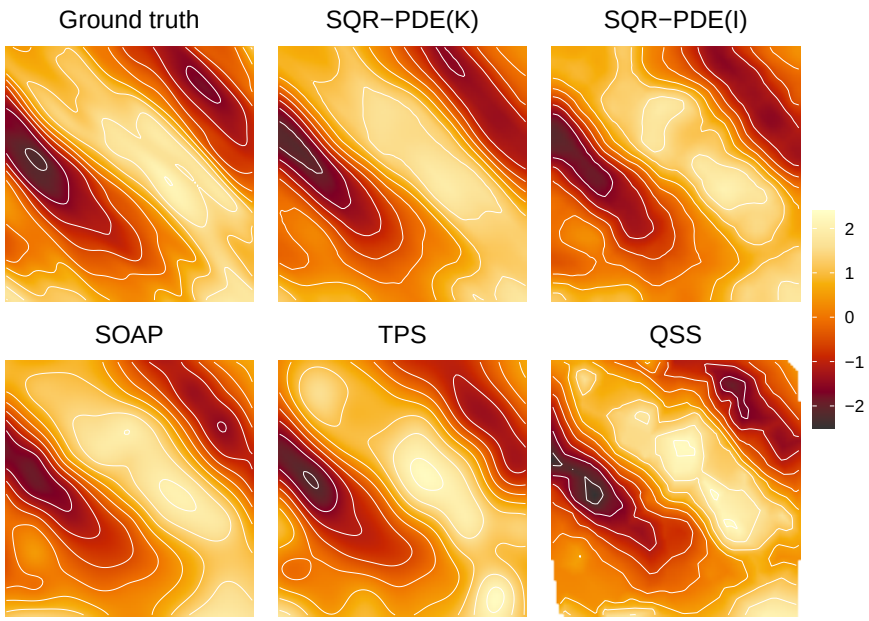

Fig. 8: Second simulation setup (Section 6.2). Estimated 90% quantile fields from simulation setting C.

compared with other methods. In scenario A, where no anisotropy is present, both anisotropic and isotropic SQR-PDE perform similarly to other methods, with SOAP and TPS providing the best estimates. When some moderate anisotropic effects are present, like in scenario B, the RMSE of SQR-PDE improves and becomes comparable or better than SOAP and TPS. Moreover, in this scenario, anisotropic SQR-PDE always reaches a lower error than isotropic SQR-PDE. Such an effect is more evident when the anisotropy is stronger, like in scenario C. Here, anisotropic SQR-PDE outperforms all other methods.

As highlighted by this simulation study, the possibility to use different regularising terms, possibly estimated from data, is an important modelling feature and may have a relevant impact on the model fit, especially when the phenomenon under study presents non-trivial spatial characteristics. In Section 7 we adopt an anisotropic SQR-PDE model for the analysis of real data, that are characterised by strongly anisotropy and a heteroscedastic locally skewed spatial distribution.

7. Real data applications

7.1. Switzerland rainfall data

We apply the proposed spatial quantile regression method to the Switzerland rainfall data, illustrated in Fig. 1 and 9, collecting 467 rainfall measurements (in 1/10 mm units) recorded on May 8, 1986. This dataset has already been analysed in, e.g., [6], [14] and [21]. In these works the authors estimated the underlying spatial field employing several geostatistical approaches, taking into account the macroscopic southwest-northeast directional anisotropy characterising the spatial mean of the underlying data-generating process. These works are restricted to mean field estimation. In our work, we instead explore quantile levels, therefore obtaining additional insights, and highlighting areas of exceedingly high or low precipitation.

Indeed, meteorological data can not be solely characterised by a spatially varying mean. Precipitation, temperature, pressure, humidity, and possibly other climatic variables, often have distributions manifesting heteroscedasticity and local skewness, as well as fat tails and extreme values. This is the case also for the Switzerland rainfall data, for which we can observe a highly heteroscedastic spatial distribution, where regions showing higher precipitation levels also have higher variability. For the sake of illustration, we first replicate the analysis proposed by [6], which developed an anisotropic spatial linear regression model based on a PDE regularisation term, indicated as SR-PDE,

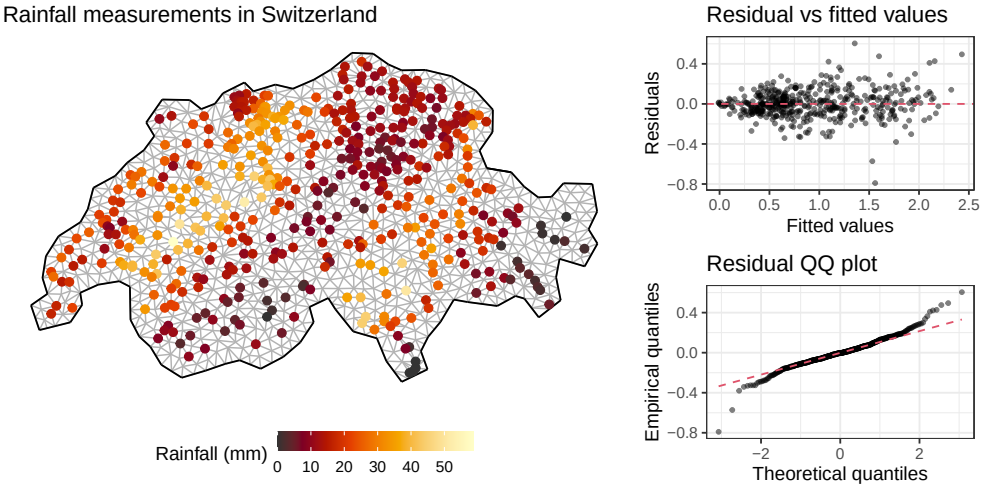

Fig. 9: Top: Switzerland rainfall data and triangular discretisation of the domain. Coloured dots indicate the measurement locations. The color of the dots is proportional to the rainfall intensity. Top right: residual versus fitted value plot. Bottom right: normal quantile-quantile plot of the residuals. The fitted values and the residuals are obtained by using the anisotropic spatial regression model by [6].

to estimate the mean field. The results are shown in Fig. 9. We observe an increase in the residual variability as the estimated mean grows, which indeed provides evidence against the usual hypothesis of homoscedastic errors. Furthermore, the normal quantile-quantile plot of the residuals highlights some deviations from the hypothesis of normal errors, with tails heavier than the Gaussian. Similar interpretations also arise when adopting different models for the underlying spatial field, like spatial regression based on kriging, radial basis expansions, and neural networks, among others. See [14] for a detailed discussion on the usage of these and other methods in the analysis of the Switzerland rainfall data.

The local specification of the proposed quantile regression makes it appropriate in the presence of heteroscedasticity. Furthermore, it is clear that our interest lies more in the tails of the distribution, which might severely differ from shifted mean surfaces, due to heteroscedasticity.

We model the Switzerland rainfall data by the proposed spatial quantile regression with stationary anisotropic regularisation given by $\int_\Omega (\mathrm{div}(\mathbf{K}\nabla f))^2$. Specifically, as described in Section 6.2, we use a two-step procedure where we first estimate the anisotropic diffusion tensor $\mathbf{K}$, as in [6], and we then use such diffusion tensor to estimate the spatial field f_α, for different quantile levels. We argue that an anisotropic diffusion tensor common to all quantile surfaces is appropriate: indeed, there is no empirical evidence to suggest a heterogeneous effect of the spatial anisotropy over the probability distribution of the rainfall, i.e., over different quantiles.

Fig. 10 shows the estimated mean and quantile surfaces at different quantile levels. In addition to a clear anisotropic effect on the mean and quantile surfaces, some interesting differences can be observed when comparing mean and quantile surfaces. The shape of the estimated quantile surfaces differs from that of the mean surface and it differs across quantiles, with deviations that are particularly evident over some localised regions. This indicates that some regions might experience more heterogeneous patterns of rainfall than others, and more extreme events, even in the presence of similar mean rainfall. The median surface is very smooth when compared to the mean surface. This is an indication that there exist local outliers and skewness, that make the mean surface less smooth than it should be. The mean surface might also be misleading, for instance suggesting excessively low precipitations in certain areas. Surfaces corresponding to percentiles more in the tails of the distribution are less smooth, as could be reasonably expected, capturing local changes in the tail of the distribution and identifying areas where more extreme events shall be expected. The 90% percentile, for instance, is characterised by several spikes in correspondence to regions manifesting a high mean level of precipitations, a further indication of the fact that limiting the analysis to the mean surface might indeed be rather misleading.

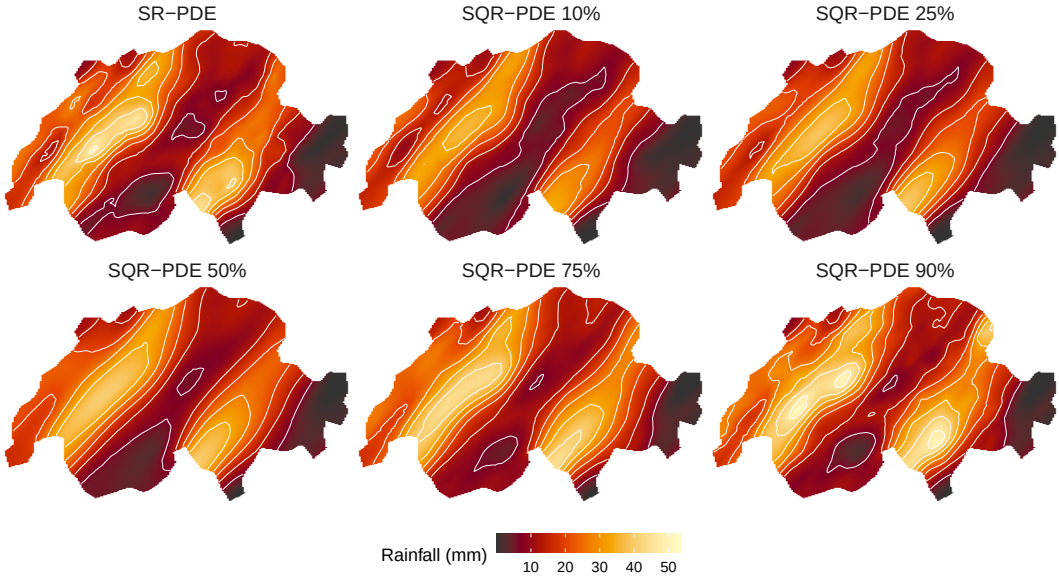

Fig. 10: The estimated mean (SR-PDE) and α-quantile (SQR-PDE) fields for the Switzerland rainfall data.

7.2. Sea surface conductivity in the Gulf of Mexico

In this second application, we consider historical data on the sea surface conductivity in the Gulf of Mexico around the Florida peninsula from 1978 to 2017 [26, 83], displayed in Fig. 11. The dataset collects 1545 aggregated measurements obtained from several data sources, and organised on a quarter-degree latitude-longitude grid in space and on a seasonal basis in time. The conductivity is obtained as a derived quantity based on temperature, pressure and salinity collected at sea surface level by buoy sensors, ship-mounted thermosalinographs, as well as satellite observations. As a consequence, the conductivity profiles can only be calculated over space-time locations for which temperature, pressure and salinity are all available, resulting in a non-uniform distribution in space of the data locations, as shown in Fig. 11.

The analysis of the conductivity profile is fundamental for the study of ocean electrodynamics and for the interpretation of ocean electric and magnetic fields. In this application, we aim at estimating the median conductivity field around the Florida peninsula, also providing spatial indicators of variability and skewness, based on robust quantile estimators obtained via the proposed SQR-PDE method.

In addition to the conductivity data, we also have information on the direction and intensity of the Gulf Stream velocity, shown in the right panel of Fig. 2. The non-stationary dynamics of the Gulf Stream have a well-known impact on the spatial distribution of the sea temperature, salinity and nutrients in the Gulf of Mexico. Thus, we want to leverage such additional information to improve the estimation of the conductivity field over the considered region.

In order to provide a suitable model for the sea surface conductivity measurements, say y_i, we assume for each quantile field the following additive decomposition: $Q_{Y_i|\mathbf{x}_i,\mathbf{p}_i}(\alpha) = \mathbf{x}_i^\top \boldsymbol{\beta}_\alpha + f_\alpha(\mathbf{p}_i)$, where $\mathbf{x}_i$ is a vector of dummy variables indicating the season of the i-th observation, $\boldsymbol{\beta}_\alpha$ is a quantile-specific vector of regression coefficients, and f_α is a quantile-specific spatial field. Additionally, we specify a non-stationary differential regularisation based on the diffusion-advection equation

$$\begin{cases} -\Delta f(\mathbf{p}) + \mathbf{b}(\mathbf{p}) \cdot \nabla f(\mathbf{p}) = 0 & \mathbf{p} \in \Omega, \\ \partial_\nu f(\mathbf{p}) = 0 & \mathbf{p} \in \partial\Omega, \end{cases} \tag{17}$$

where $\mathbf{b}(\mathbf{p})$ is a non-stationary transport term proportional to the Gulf stream velocity, and $\partial_\nu f$ is the outward normal derivative of f on the boundary. Specifically, we set $\mathbf{b}(\mathbf{p}) = \eta \mathbf{v}(\mathbf{p})$, where $\mathbf{v}(\cdot)$ is the stream velocity field shown in the right panel of Fig. 2, and $\eta \geq 0$ is a scalar parameter controlling the relative weight of the transport term $\mathbf{v} \cdot \nabla f$ relative to the diffusion term $-\Delta f$.

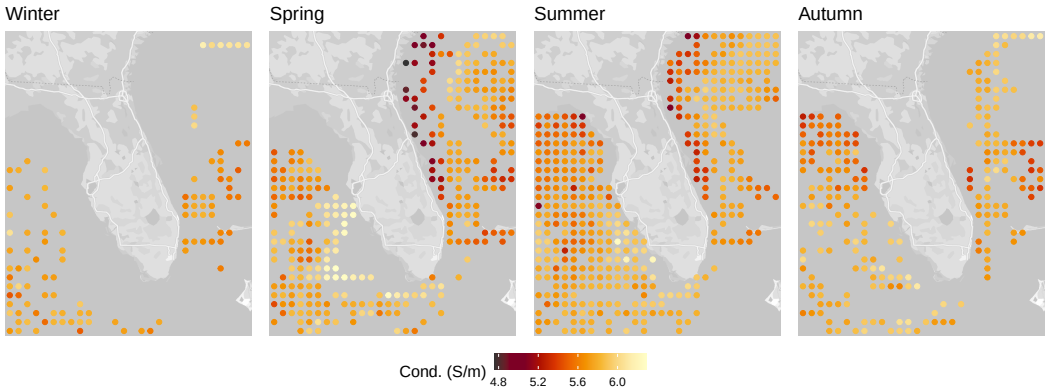

Fig. 11: Historical sea surface conductivity data in the Gulf of Mexico around the Florida peninsula. The data are aggregated on a seasonal basis in time and on a quarter latitude-longitude grid in space, where each panel corresponds to a season.

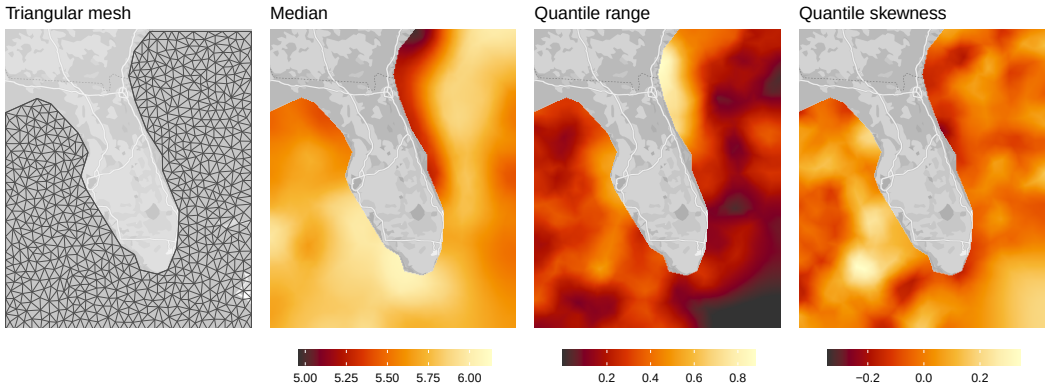

Fig. 12: Results obtained for the sea surface conductivity application. From left to right: triangular mesh, estimated median, quartile range and quantile skewness fields as defined in (18).

In order to build a finite element approximation of the quantile field f_α, we discretise the considered domain using the triangular mesh shown in the left panel of Fig. 12. Then, analogously to the two-step procedure discussed in Section 6.2, we use the parameter cascading algorithm by [6] to determine the optimal value of η. Doing so, we obtain the estimate $\hat{\eta} = 2.7181$, which suggests that the relative importance of the transport term is higher than the isotropic diffusion term in the PDE specification (17). Given such a non-stationary diffusion-transport PDE, we estimate $\boldsymbol{\beta}_\alpha$ and f_α using a generalisation of the EM algorithm proposed in Section 3, that accounts for the presence of a vector of fixed effect covariates.

We study the spatial distribution of the sea conductivity over the domain by estimating three quantile surfaces, corresponding to $\alpha = 0.1, 0.5, 0.9$. Based on such estimates, we derive the following quantile-based fields

$$\text{Median}(\mathbf{p}) = \hat{f}_{0.5}(\mathbf{p}), \quad \text{Range}(\mathbf{p}) = \hat{f}_{0.9}(\mathbf{p}) - \hat{f}_{0.1}(\mathbf{p}), \quad \text{Skewness}(\mathbf{p}) = \hat{f}_{0.9}(\mathbf{p}) - 2\hat{f}_{0.5}(\mathbf{p}) + \hat{f}_{0.1}(\mathbf{p}), \tag{18}$$

which provide robust measures of pointwise location, variability and skewness, respectively. We show the obtained fields in Fig. 12. Two high-level conductivity regions are detected: the first one is localised along the southwest coastline of the Florida peninsula, where the water stream flowing out from the Gulf of Mexico encounters the Florida boundary before entering into the Atlantic Ocean; the second one follows the south-north direction induced by the Gulf stream. The lowest conductivity is measured along the northeast Florida coastline, which is also characterised by high variability and a relatively calm water flow. More generally, comparing the left and central panels in Fig. 12, we observe that lower conductivity levels along the coastline are associated with higher variability. The lowest variability is observed in the southwest regions of the domain. This is due to the fact that only a few observations are available in

that area, leading the estimated quantiles to converge toward the median solution. The estimated skewness field in the right panel of Fig. 12 varies around zero without manifesting evident spatial patterns, which suggests that the spatial distribution of the data is almost symmetric over the domain.

8. Discussion and future works

In the present work we described a new spatial quantile regression model with PDE regularisation. We proposed an innovative functional EM algorithm, and we proved the theoretical properties of the resulting estimator.

The proposed method has very good performances, as shown by simulation studies and by the analysis of real data concerning rainfall measurements and sea surface conductivity. The considered scenarios highlight the importance of accounting for complex domain structures, for deviation from isotropic stationary fields, and for spatially heterogeneous characteristics of the distribution, such as heteroscedasticity or skewness. The proposed method offers increasingly large advantages with respect to the available alternatives as the data complexity increases, in terms of domain shape or field characteristics, such as anisotropy.

A remarkable benefit of our model formulation is its generality, which permits us to naturally accommodate several extensions, different sampling schemes, and PDE models. For instance, data recorded over areal regions can be handled by extending the proposal made by [5] in the context of spatial linear models. Data observed in space and time, as well as spatially varying functional data, could be handled by considering space-time PDEs, as shown by [1] and [7] in spatial linear regression settings. Another fascinating possibility is to adapt our PDE-based spatial quantile regression model to data lying on a curved domain. In this context the necessary theoretical background is provided by the works of [16, 46] for spatial linear regression over two-dimensional Riemannian manifolds, which could naturally be integrated into our quantile regression framework, permitting, for instance, the application of SQR-PDE to data observed over the globe. Such an extension might also be of significant interest for applications in mechanical and space engineering, as well as in life sciences.

A second relevant research direction is concerned with the joint modelling of multiple quantiles. Inspired by the recent works of [24, 25, 66] on varying coefficient modelling, we might extend our methodology in order to jointly estimate a whole family of quantile fields, as smooth functions of the confidence level α. In this context, the heteroscedastic quantile model by [32] and the constrained quantile curves by [9] provide two remarkable examples of non-crossing approaches that are worth investigating and possibly extending to the proposed spatial regression with PDE regularisation.

Another research area, which deserves relevant attention, involves finite-sample inference. Indeed, as mentioned in Section 5, Theorem 1 justifies asymptotic Wald-type inference. Wald-type approaches, however, may lead to under-conservative inference in finite-sample scenarios [see, e.g., 23, 47]. This issue has classically stimulated the development of various alternative inference approaches for nonparametric and semiparametric regression under linear and generalized linear settings [see, e.g., 30, 81]. For spatial linear regression with differential regularisation, [12, 21, 22] discuss a nonparametric inference strategy, based on innovative resampling approaches, which provides valid inference also in finite sample. Such ideas constitute promising avenues to be explored also in quantile regression settings.

Appendix

Proof of Proposition 1: We first recall that any continuous convex function defined over a convex domain attains its minimum values within its domain [see, e.g., 44, Proposition 6.5.1]. Then, proving the first statement in Proposition 1 is equivalent to showing that $\mathcal{F}_\gamma$ is a closed, convex space, and $J_\alpha(f)$ is a continuous, convex functional.

The closure and convexity of $\mathcal{F}_\gamma(\Omega) = \{f \in H^2(\Omega) : Bf = \gamma \text{ on } \partial\Omega\}$ follows from the vector space structure of $H^2(\Omega)$ and from the linearity of the differential operator B, that is $B\{\phi f+(1-\phi)g\} = \phi Bf+(1-\phi)Bg = \phi\gamma+(1-\phi)\gamma = \gamma$, for all $\phi \in [0, 1]$. The continuity and convexity of $J_\alpha(f)$ follow from the continuity and convexity of the quantile loss $\rho_\alpha\{y_i - f(\mathbf{p}_i)\}$ [see, e.g., 38] and the regularisation term $\int_\Omega (Lf - u)^2$ [see, e.g., 4].

Finally, we prove that $\mathcal{V}_{\gamma,\alpha}(\Omega)$ is a closed, convex set. To do so, we define the sublevel set of $J_\alpha(f)$ by

$$\mathcal{V}_{\gamma,\alpha}(\Omega, t) = \{f \in \mathcal{F}_\gamma(\Omega) : J_\alpha(f) \leq t\},$$

for any t such that $\mathcal{V}_\gamma(\Omega, t)$ is non-empty. Let $f, g \in \mathcal{V}_{\gamma,\alpha}(\Omega, t)$ and consider the convex combination

$$J_\alpha\{\phi f + (1-\phi)g\} \leq \phi J_\alpha(f) + (1-\phi)J_\alpha(g) \leq \phi t + (1-\phi)t = t, \qquad \forall\, \phi \in [0,1].$$

This implies the closure and convexity of $\mathcal{V}_{\gamma,\alpha}(\Omega, t)$ for any t since $J_\alpha(f)$ is a continuous functional. The closure and convexity of $\mathcal{V}_{\gamma,\alpha}(\Omega)$ immediately follows by noting that

$$\mathcal{V}_{\gamma,\alpha}(\Omega) = \{f \in \mathcal{F}_\gamma(\Omega) : J_\alpha(\hat{f}_\alpha) = \inf J_\alpha(f)\} = \mathcal{V}_\gamma(\Omega, \inf J_\alpha(f)).$$

This concludes the proof. □

Proof of Proposition 2: Let $\hat{f}_\alpha \in \mathcal{V}_{\gamma,\alpha}$ be a minimiser of functional (2). For any $t \geq 0$ and $\psi \in \mathcal{F}_0$, we have $J_\alpha(\hat{f}_\alpha) \leq J_\alpha(\hat{f}_\alpha + t\psi)$. Hence, taking the limit for $t \downarrow 0$, we get that the Gateaux directional derivative of $J_\alpha(\hat{f}_\alpha)$ along the direction ψ must be non-negative:

$$\partial_\psi J(\hat{f}_\alpha) = \frac{\partial}{\partial t} J(\hat{f}_\alpha + t\psi)\Big|_{t=0} = \lim_{t \downarrow 0} \frac{J(\hat{f}_\alpha + t\psi) - J(\hat{f}_\alpha)}{t} \geq 0.$$

Whenever $J_\alpha(\cdot)$ is a differentiable functional, the above condition collapses into the first order equation $\partial_\psi J_\alpha(\hat{f}_\alpha) = 0$ for any $\psi \in \mathcal{F}_0$. This is the case, for instance, for generalised linear models with PDE regularisation [75].

Considering the explicit functional form of $J_\alpha(\cdot)$ and applying the above condition on the Gateaux derivative of $J_\alpha(\cdot)$, we obtain

$$\partial_\psi J_\alpha(\hat{f}_\alpha) = \frac{1}{n}\sum_{i=1}^{n} \partial_\psi \rho_\alpha(y_i - \hat{f}_\alpha(\mathbf{p}_i)) + \frac{\lambda}{2}\partial_\psi \int_\Omega (L\hat{f}_\alpha - u)^2 \geq 0.$$

As shown by, e.g., [4], the directional derivative for the penalty term is given by $\partial_\psi \int_\Omega (L\hat{f}_\alpha - u)^2 = 2\int_\Omega (L\psi)(L\hat{f}_\alpha - u)$. On the other hand, the directional derivative for the quantile loss term [36] is $\partial_\psi \rho_\alpha(y_i - \hat{f}_\alpha(\mathbf{p}_i)) = -\psi(\mathbf{p}_i)\dot{\rho}_\alpha(y_i - \hat{f}_\alpha(\mathbf{p}_i), -\psi(\mathbf{p}_i))$, where $\dot{\rho}_\alpha(y_i - \hat{f}(\mathbf{p}_i), -\psi(\mathbf{p}_i))$ is the directional derivative of $\rho_\alpha(\cdot)$ calculated in $y_i - \hat{f}_\alpha(\mathbf{p}_i)$ along the direction $-\psi(\mathbf{p}_i)$. Substituting these expressions in the above inequality we obtain the final result. This concludes the proof. □

Proof of Proposition 3: Let $(\tilde{f}_\alpha, \tilde{\boldsymbol{\eta}}_\alpha) \in \mathcal{F}_\gamma \times \mathbb{R}^{|s|}$ be a minimiser of $\mathcal{L}_\alpha(f, \boldsymbol{\eta})$. Then, because of the Lagrange multiplier theorem, $(\tilde{f}_\alpha, \tilde{\boldsymbol{\eta}}_\alpha)$ must satisfy the first order conditions $(\partial/\partial\boldsymbol{\eta})\mathcal{L}_\alpha(\tilde{f}_\alpha, \tilde{\boldsymbol{\eta}}_\alpha) = 0$ and $(\partial/\partial t)\mathcal{L}_\alpha(\tilde{f}_\alpha + t\psi, \tilde{\boldsymbol{\eta}}_\alpha)\big|_{t=0} = 0$, for all $\psi \in \mathcal{F}_0$. The first equation just enforces the constraints, being

$$\frac{\partial}{\partial \boldsymbol{\eta}}\mathcal{L}(\tilde{f}_\alpha, \tilde{\boldsymbol{\eta}}_\alpha) = \frac{1}{n}\frac{\partial}{\partial \boldsymbol{\eta}}\{\boldsymbol{\eta}^\top(\mathbf{z} - \tilde{\mathbf{f}}_{\alpha,n})_s\}\Big|_{\boldsymbol{\eta}=\tilde{\boldsymbol{\eta}}_\alpha} = \frac{1}{n}(\mathbf{z} - \tilde{\mathbf{f}}_{\alpha,n})_s = 0.$$

The second equation requires that the Gateaux directional derivatives of $\mathcal{L}_\alpha(f, \boldsymbol{\eta})$ with respect to f get nullified in all the possible directions $\psi \in \mathcal{F}_0$.

Observing that $\mathcal{L}_\alpha(f, \boldsymbol{\eta})$ is quadratic in f, taking the directional derivative and equating to zero, we obtain the first-order variational equation

$$(\boldsymbol{\psi}_n)_{-s}^\top \mathbf{W}_{-s}\,(\tilde{\mathbf{f}}_{\alpha,n})_{-s} + \lambda n \int_\Omega (L\psi)(L\tilde{f}_\alpha) = (\boldsymbol{\psi}_n)_{-s}^\top \mathbf{W}_{-s}\,\mathbf{z}_{-s} + (\boldsymbol{\psi}_n)_{-s}^\top \tilde{\boldsymbol{\eta}}_\alpha + \lambda n \int_\Omega (L\psi)u, \tag{19}$$

which must hold for any $\psi \in \mathcal{F}_0$. We denote by $G(\psi, \tilde{f}_\alpha)$ and $F(\tilde{f}_\alpha)$ the left and right sides of the above equation (19), respectively; in this way, we can write the first order condition in the equivalent form

$$G(\psi, \tilde{f}_\alpha) = F(\psi), \qquad \forall \psi \in \mathcal{F}_0(\Omega). \tag{20}$$

If the parameters of the PDE are such that, for any $u \in L^2(\Omega)$, there exists a unique solution f of the PDE $Lf = u$, which, moreover, satisfies $f \in H^2(\Omega)$ [4, Assumption 2], then $G(\cdot,\cdot)$ is a symmetric, continuous, coercive bilinear map, and $F(\cdot)$ is a continuous linear operator [4, Theorem 2]. As a consequence, thanks to the Lax-Milgram lemma [see, e.g., 57, Section 3.4.1], there exists a unique solution $\tilde{f}_\alpha \in \mathcal{F}_\gamma$ to equation (20).

The proof is concluded by noting that the Euler-Lagrange equation (19) is equivalent to the system

$$(\boldsymbol{\psi}_n)_{-s}^\top \mathbf{W}_{-s}\,(\tilde{\mathbf{f}}_{\alpha,n})_{-s} + \lambda n \int_\Omega (L\psi)\tilde{g}_\alpha = (\boldsymbol{\psi}_n)_{-s}^\top \mathbf{W}_{-s}\,\mathbf{z}_{-s} + (\boldsymbol{\psi}_n)_{-s}^\top \tilde{\boldsymbol{\eta}}_\alpha, \qquad \int_\Omega (L\tilde{f}_\alpha)\phi - \int_\Omega \tilde{g}_\alpha\phi = \int_\Omega u\phi,$$

for any pair of test functions $\psi, \phi \in \mathcal{F}_0(\Omega)$, where $\tilde{g}_\alpha = L\tilde{f}_\alpha - u$. Hence, we can integrate by parts and use the definition of $R_1(\cdot,\cdot)$, $R_0(\cdot,\cdot)$ and $F(\cdot)$ in (10) in order to recognise the weak variational formulation in Proposition 3. □

Proof of Proposition 4: From Proposition 3, any solution $(\tilde{f}_{\alpha,h}, \tilde{g}_{\alpha,h}, \tilde{\boldsymbol{\eta}}_\alpha)$ in the finite element space must satisfy the system of equations in (10). Thanks to the finite element discretisation, we have $(\boldsymbol{\psi}_n)_{-s} = \boldsymbol{\Psi}_{-s}\boldsymbol{\psi}$, $(\tilde{\mathbf{f}}_{\alpha,n})_{-s} = \boldsymbol{\Psi}_{-s}\tilde{\mathbf{f}}_\alpha$, $(\tilde{\mathbf{f}}_{\alpha,n})_s = \boldsymbol{\Psi}_s\tilde{\mathbf{f}}_\alpha$, $R_0(\psi_h, \tilde{g}_{\alpha,h}) = \boldsymbol{\psi}^\top\mathbf{R}_0^\top\tilde{\mathbf{g}}_\alpha$, $R_1(\tilde{f}_{\alpha,h}, \phi_h) = \tilde{\mathbf{f}}_\alpha^\top\mathbf{R}_0^\top\boldsymbol{\phi}$, $F(\phi_h) = (\mathbf{u}+\boldsymbol{\gamma})^\top\boldsymbol{\phi}$. Therefore, the above system can be written as

$$\boldsymbol{\psi}^\top\boldsymbol{\Psi}_{-s}^\top\mathbf{W}_{-s}\boldsymbol{\Psi}_{-s}\tilde{\mathbf{f}}_\alpha - \boldsymbol{\psi}^\top\boldsymbol{\Psi}_s^\top\tilde{\boldsymbol{\eta}}_\alpha + \lambda n\boldsymbol{\psi}^\top\mathbf{R}_1^\top\tilde{\mathbf{g}}_\alpha = \boldsymbol{\psi}^\top\boldsymbol{\Psi}_{-s}^\top\mathbf{W}_{-s}\mathbf{z}_{-s}, \quad \tilde{\mathbf{f}}_\alpha^\top\mathbf{R}_1^\top\boldsymbol{\phi} - \tilde{\mathbf{g}}_\alpha^\top\mathbf{R}_0\boldsymbol{\phi} = (\mathbf{u}+\boldsymbol{\gamma})^\top\boldsymbol{\phi}, \quad (\mathbf{z} - \boldsymbol{\Psi}_s\tilde{\mathbf{f}}_\alpha) = \mathbf{0},$$

for any pair of vectors $\boldsymbol{\psi}, \boldsymbol{\phi} \in \mathbb{R}^{N_h}$. Since any function $f_h \in \mathcal{F}_{\gamma,h}$ is uniquely determined by its values on the nodes, i.e., by its coefficient vector, solving the above system is equivalent to finding $(\tilde{\mathbf{f}}_\alpha, \tilde{\mathbf{g}}_\alpha, \tilde{\boldsymbol{\eta}}_\alpha) \in \mathbb{R}^{N_h} \times \mathbb{R}^{N_h} \times \mathbb{R}^{|s|}$ solving linear system (11). This concludes the proof. □

Proof of Corollary 1: Let us start by considering the finite element system (11) in Proposition 4. Solving the second equation with respect to $\tilde{\mathbf{g}}_\alpha$, we have

$$\tilde{\mathbf{g}}_\alpha = \mathbf{R}_0^{-1}(\mathbf{R}_1\tilde{\mathbf{f}}_\alpha - \mathbf{u} - \boldsymbol{\gamma}) = \mathbf{R}_0^{-1}\mathbf{R}_1\tilde{\mathbf{f}}_\alpha - \mathbf{R}_0^{-1}(\mathbf{u}+\boldsymbol{\gamma}).$$

If we substitute the expression for $\tilde{\mathbf{g}}_\alpha$ in the first equation in (11), we get

$$(\boldsymbol{\Psi}_{-s}^\top\mathbf{W}_{-s}\boldsymbol{\Psi}_{-s} + \lambda n\mathbf{R}_1^\top\mathbf{R}_0^{-1}\mathbf{R}_1)\,\tilde{\mathbf{f}}_\alpha = \boldsymbol{\Psi}_{-s}^\top\mathbf{W}_{-s}\mathbf{z}_{-s} + \lambda n\,\mathbf{R}_1^\top\mathbf{R}_0^{-1}(\mathbf{u}+\boldsymbol{\gamma}) - \boldsymbol{\Psi}_s^\top\tilde{\boldsymbol{\eta}}_\alpha.$$

Then, using the definition of $\mathbf{A}$ and $\mathbf{d}$ and solving for $\tilde{\mathbf{f}}_\alpha$, we obtain $\tilde{\mathbf{f}}_\alpha = \mathbf{A}^{-1}(\mathbf{d} - \boldsymbol{\Psi}_s^\top\tilde{\boldsymbol{\eta}}_\alpha)$. By plugging this expression for $\tilde{\mathbf{f}}_\alpha$ in (11) and solving for $\tilde{\boldsymbol{\eta}}_\alpha$, we get $\boldsymbol{\Psi}_s\mathbf{A}^{-1}(\mathbf{d} - \boldsymbol{\Psi}_s^\top\tilde{\boldsymbol{\eta}}_\alpha) = \mathbf{z}_s$, which implies

$$\tilde{\boldsymbol{\eta}}_\alpha = (\boldsymbol{\Psi}_s\mathbf{A}^{-1}\boldsymbol{\Psi}_s^\top)^{-1}(\boldsymbol{\Psi}_s\mathbf{A}^{-1}\mathbf{d} - \mathbf{z}_s).$$

Finally, we use the definition of $\mathbf{B}$ and we substitute the above expression for $\tilde{\boldsymbol{\eta}}_\alpha$ in $\tilde{\mathbf{f}}_\alpha = \mathbf{A}^{-1}(\mathbf{d} - \boldsymbol{\Psi}_s^\top\tilde{\boldsymbol{\eta}}_\alpha)$:

$$\tilde{\mathbf{f}}_\alpha = \mathbf{A}^{-1}(\mathbf{d} - \boldsymbol{\Psi}_s^\top\tilde{\boldsymbol{\eta}}_\alpha) = \mathbf{A}^{-1}\{\mathbf{d} - \boldsymbol{\Psi}_s^\top\mathbf{B}^{-1}(\boldsymbol{\Psi}_s\mathbf{A}^{-1}\mathbf{d} - \mathbf{z}_s)\} = \mathbf{A}^{-1}\mathbf{d} + \mathbf{A}^{-1}\boldsymbol{\Psi}_s^\top\mathbf{B}^{-1}(\mathbf{z}_s - \boldsymbol{\Psi}_s\mathbf{A}^{-1}\mathbf{d}).$$

This concludes the proof. □

Proof of Proposition 5: The linear predictor is obtained by pre-multiplying the coefficient vector $\hat{\mathbf{f}}_\alpha$ by its design matrix $\boldsymbol{\Psi}$, namely

$$\hat{\mathbf{f}}_{\alpha,n} = \boldsymbol{\Psi}\hat{\mathbf{f}}_\alpha = \boldsymbol{\Psi}\mathbf{A}^{-1}\mathbf{d} + \boldsymbol{\Psi}\mathbf{A}^{-1}\boldsymbol{\Psi}_s^\top\mathbf{B}^{-1}(\mathbf{z}_s - \boldsymbol{\Psi}_s\mathbf{A}^{-1}\mathbf{d}) = \boldsymbol{\Psi}\mathbf{A}^{-1}(\mathbf{I} - \boldsymbol{\Psi}_s^\top\mathbf{B}^{-1}\boldsymbol{\Psi}_s\mathbf{A}^{-1})\mathbf{d} + \boldsymbol{\Psi}\mathbf{A}^{-1}\boldsymbol{\Psi}_s^\top\mathbf{B}^{-1}\mathbf{z}_s.$$

Define $\mathbf{C} = \mathbf{I} - \boldsymbol{\Psi}_s^\top\mathbf{B}^{-1}\boldsymbol{\Psi}_s\mathbf{A}^{-1}$ and recall that $\mathbf{d} = \boldsymbol{\Psi}_{-s}^\top\mathbf{W}_{-s}\mathbf{z}_{-s} + \lambda n\mathbf{R}_0^{-1}(\mathbf{u}+\boldsymbol{\gamma})$, then

$$\begin{aligned}\hat{\mathbf{f}}_{\alpha,n} &= \boldsymbol{\Psi}\mathbf{A}^{-1}\mathbf{C}\{\boldsymbol{\Psi}_{-s}^\top\mathbf{W}_{-s}\mathbf{z}_{-s} + \lambda n\mathbf{R}_0^{-1}(\mathbf{u}+\boldsymbol{\gamma})\} + \boldsymbol{\Psi}\mathbf{A}^{-1}\boldsymbol{\Psi}_s^\top\mathbf{B}^{-1}\mathbf{z}_s \\ &= \underbrace{\boldsymbol{\Psi}\mathbf{A}^{-1}\mathbf{C}\boldsymbol{\Psi}_{-s}^\top\mathbf{W}_{-s}}_{\mathbf{S}_{-s}}\mathbf{z}_{-s} + \underbrace{\boldsymbol{\Psi}\mathbf{A}^{-1}\boldsymbol{\Psi}_s^\top\mathbf{B}^{-1}}_{\mathbf{S}_s}\mathbf{z}_s + \underbrace{\lambda n\boldsymbol{\Psi}\mathbf{A}^{-1}\mathbf{C}\mathbf{R}_1^\top\mathbf{R}_0^{-1}(\mathbf{u}+\boldsymbol{\gamma})}_{\mathbf{r}}.\end{aligned}$$

Therefore, the smoothing matrix $\mathbf{S} = [\,\mathbf{S}_{-s}, \mathbf{S}_s\,]$ may be defined according to a block-structure, separating the constrained and unconstrained smoothing factors, so that $\mathbf{S}\,\mathbf{z} = \mathbf{S}_{-s}\mathbf{z}_{-s} + \mathbf{S}_s\mathbf{z}_s$.

According to [54, 68], the effective degrees of freedom of the linear smoother $\hat{\mathbf{f}}_{\alpha,n}$ are defined as $\mathrm{df}_\alpha = \mathrm{tr}(\mathbf{S})$. Notice that the diagonal block of $\mathbf{S}$ corresponding to index set s is given by $\mathbf{S}_{s,s} = \boldsymbol{\Psi}_s\mathbf{A}^{-1}\boldsymbol{\Psi}_s^\top\mathbf{B}^{-1}$ and, similarly, $\mathbf{S}_{-s,-s} = \boldsymbol{\Psi}_{-s}\mathbf{A}^{-1}\mathbf{C}\,\boldsymbol{\Psi}_{-s}^\top\mathbf{W}_{-s}$. Then, thanks to the block-structure of $\mathbf{S}$ and the cyclic permutation property of the trace, we have

$$\begin{aligned}\mathrm{tr}(\mathbf{S}) &= \mathrm{tr}(\boldsymbol{\Psi}_{-s}\mathbf{A}^{-1}\mathbf{C}\,\boldsymbol{\Psi}_{-s}^\top\mathbf{W}_{-s}) + \mathrm{tr}(\boldsymbol{\Psi}_s\mathbf{A}^{-1}\boldsymbol{\Psi}_s^\top\mathbf{B}^{-1}) = \mathrm{tr}(\mathbf{A}^{-1}\mathbf{C}\,\boldsymbol{\Psi}_{-s}^\top\mathbf{W}_{-s}\boldsymbol{\Psi}_{-s}) + \mathrm{tr}(\boldsymbol{\Psi}_s\mathbf{A}^{-1}\boldsymbol{\Psi}_s^\top\mathbf{B}^{-1}) \\ &= \mathrm{tr}(\mathbf{A}^{-1}\mathbf{C}\,\boldsymbol{\Psi}_{-s}^\top\mathbf{W}_{-s}\boldsymbol{\Psi}_{-s}) + \mathrm{tr}(\mathbf{B}\mathbf{B}^{-1}) = \mathrm{tr}(\mathbf{A}^{-1}\mathbf{C}\,\boldsymbol{\Psi}_{-s}^\top\mathbf{W}_{-s}\boldsymbol{\Psi}_{-s}) + |s|.\end{aligned}$$

This concludes the proof. □

Proof of Theorem 1: Let $J_\alpha(\mathbf{f}) = J_\alpha(f_h)$ be the finite element discretisation of the objective functional in (2) with $f_h = \mathbf{f}^\top \boldsymbol{\psi} \in \mathcal{F}^1_{\gamma,h}$, which is given by $J_\alpha(\mathbf{f}) = (1/n)\sum_{i=1}^n \rho_\alpha(y_i - \boldsymbol{\psi}_i^\top \mathbf{f}) + (\lambda/2)P(\mathbf{f})$, where $P(\mathbf{f}) \equiv P(f_h) = \int_{\Omega_h} (Lf_h - u_h)^2$ denotes the discretisation of the differential penalisation term and corresponds to $P(\mathbf{f}) = (\mathbf{R}_1\mathbf{f} - \mathbf{u} - \boldsymbol{\gamma})^\top \mathbf{R}_0^{-1}(\mathbf{R}_1\mathbf{f} - \mathbf{u} - \boldsymbol{\gamma})$.

Following the convexity argument proposed by, e.g., [34, 35, 55] in the context of parametric linear quantile regression, we introduce the quantile residual $\varepsilon_i = y_i - \boldsymbol{\psi}_i^\top \mathbf{f}_\alpha$ and the reparametrisation $\boldsymbol{\delta} = \sqrt{n}\,(\mathbf{f} - \mathbf{f}_\alpha)$, so that $y_i - \boldsymbol{\psi}_i^\top \mathbf{f} = \varepsilon_i - \boldsymbol{\psi}_i^\top \boldsymbol{\delta}/\sqrt{n}$. Notice that $\boldsymbol{\psi}^\top \mathbf{f}_\alpha$ is the finite element interpolation of f_α on the mesh knots. Similarly, $\boldsymbol{\psi}^\top \mathbf{g}_\alpha$ is the interpolation of g_α and, thanks to the second identity in (12), we have $\mathbf{g}_\alpha = \mathbf{R}_0^{-1}(\mathbf{R}_1\mathbf{f}_\alpha - \mathbf{u} - \boldsymbol{\gamma})$. Furthermore, we define the reparametrised scaled objective functional

$$nJ_\alpha(\boldsymbol{\delta}/\sqrt{n} + \mathbf{f}_\alpha) = S_\alpha(\boldsymbol{\delta}) + \frac{\lambda}{2} nP(\boldsymbol{\delta}/\sqrt{n} + \mathbf{f}_\alpha) = \sum_{i=1}^n \{\rho_\alpha(\varepsilon_i - \boldsymbol{\psi}_i^\top \boldsymbol{\delta}/\sqrt{n}) - \rho_\alpha(\varepsilon_i)\} + \frac{\lambda}{2} nP(\boldsymbol{\delta}/\sqrt{n} + \mathbf{f}_\alpha).$$

It is easy to verify that minimising $J_\alpha(\cdot)$ with respect to $\boldsymbol{\delta}$ is equivalent to minimising $J_\alpha(\cdot)$ with respect to $\mathbf{f}$, and the solution corresponds to $\hat{\boldsymbol{\delta}}_\alpha = \sqrt{n}\,(\hat{\mathbf{f}}_\alpha - \mathbf{f}_\alpha)$. Therefore, we can study the asymptotic properties of $\hat{\boldsymbol{\delta}}_\alpha$ to infer the limiting behaviour of $\hat{\mathbf{f}}_\alpha$.

We consider the Knight's identity [34], which permits us to represent $S_\alpha(\boldsymbol{\delta})$ using the additive decomposition

$$S_\alpha(\boldsymbol{\delta}) = S'_\alpha(\boldsymbol{\delta}) + S''_\alpha(\boldsymbol{\delta}) = \frac{1}{\sqrt{n}} \sum_{i=1}^n (\boldsymbol{\psi}_i^\top \boldsymbol{\delta})\{\alpha - \mathbb{I}(\varepsilon_i < 0)\} + \sum_{i=1}^n \int_0^{\boldsymbol{\psi}_i^\top \boldsymbol{\delta}/\sqrt{n}} \{\mathbb{I}(\varepsilon_i < t) - \alpha\}\,\mathrm{d}t.$$

We will now study the asymptotic behaviour of $S'_\alpha(\boldsymbol{\delta})$ and $S''_\alpha(\boldsymbol{\delta})$ to find a convenient limiting representation of $S_\alpha(\boldsymbol{\delta})$. First, observing that $\mathbb{I}(\varepsilon_i < 0)$, $i \in \{1, \ldots, n\}$, is a sequence of independent Bernoulli random variables with success probability α, we have

$$\mathrm{E}\{S'_\alpha(\boldsymbol{\delta})\} = \frac{1}{\sqrt{n}} \sum_{i=1}^n (\boldsymbol{\psi}_i^\top \boldsymbol{\delta})\{\alpha - \Pi_{\varepsilon_i|\mathbf{p}_i}(0)\} = 0, \qquad \mathrm{Var}\{S'_\alpha(\boldsymbol{\delta})\} = \alpha(1-\alpha)\frac{1}{n}\sum_{i=1}^n (\boldsymbol{\psi}_i^\top \boldsymbol{\delta})^2 = \boldsymbol{\delta}^\top \mathbf{D}_{0,n} \boldsymbol{\delta}.$$

Therefore, using the central limit theorem, we get $S'_\alpha(\boldsymbol{\delta}) = \boldsymbol{\delta}^\top \boldsymbol{\eta}_n + o_{\mathrm{p}}(1)$, where $\boldsymbol{\eta}_n \sim \mathrm{N}_{N_h}(\mathbf{0}, \mathbf{D}_{0,n})$. The second term has expected value

$$\mathrm{E}\{S''_\alpha(\boldsymbol{\delta})\} = \sum_{i=1}^n \mathrm{E} \int_0^{\boldsymbol{\psi}_i^\top \boldsymbol{\delta}/\sqrt{n}} \{\mathbb{I}(\varepsilon_i < t) - \alpha\}\,\mathrm{d}t = \sum_{i=1}^n \int_0^{\boldsymbol{\psi}_i^\top \boldsymbol{\delta}/\sqrt{n}} \{\Pi_{\varepsilon_i|\mathbf{p}_i}(t) - \alpha\}\,\mathrm{d}t.$$

By a second order Taylor expansion around $\boldsymbol{\delta} = 0$, we obtain

$$\mathrm{E}\{S''_\alpha(\boldsymbol{\delta})\} = \frac{1}{2n}\sum_{i=1}^n \pi_{\varepsilon_i|\mathbf{p}_i}(0)(\boldsymbol{\psi}_i^\top \boldsymbol{\delta})^2 + \frac{1}{n^{3/2}} \sum_{i=1}^n O(|\boldsymbol{\psi}_i^\top \boldsymbol{\delta}|^3) = \frac{1}{2}\boldsymbol{\delta}^\top \mathbf{D}_{1,n} \boldsymbol{\delta} + O(\|\boldsymbol{\delta}\|^3/\sqrt{n}),$$

where, thanks to Assumption 1, $\pi_i = \pi_{\varepsilon_i|\mathbf{p}_i}(0) = \pi_{y_i|\mathbf{p}_i}(\boldsymbol{\psi}_i^\top \mathbf{f}_\alpha)$ is bounded away from 0 and ∞ for any $i \in \{1, \ldots, n\}$ and the residual term $O(\|\boldsymbol{\delta}\|^3/\sqrt{n})$ converges to 0 for any value of $\boldsymbol{\delta}$. As shown in, e.g., [34, 36, 55] the variance of $S''_\alpha(\boldsymbol{\delta})$ can be upper-bounded by

$$\mathrm{Var}\{S''_\alpha(\boldsymbol{\delta})\} \leq \frac{1}{\sqrt{n}} \max_{i=1,\ldots,n} |\boldsymbol{\psi}_i^\top \boldsymbol{\delta}|\, \mathrm{E}\{S''_\alpha(\boldsymbol{\delta})\} \leq \frac{1}{\sqrt{n}} \sup_{\mathbf{p}\in\Omega} |\boldsymbol{\psi}(\mathbf{p})|\, \|\boldsymbol{\delta}\|\, \mathrm{E}\{S''_\alpha(\boldsymbol{\delta})\} = O(n^{-1/2}),$$

thus $S''_\alpha(\boldsymbol{\delta})$ converges in probability to its mean and can be represented as $S''_\alpha(\boldsymbol{\delta}) = \boldsymbol{\delta}^\top \mathbf{D}_{1,n}\boldsymbol{\delta}/2 + o_{\mathrm{p}}(1)$.

We now consider the non-stochastic penalty term, which is given by

$$\begin{aligned} nP(\boldsymbol{\delta}/\sqrt{n} + \mathbf{f}_\alpha) &= n\{\mathbf{R}_1(\boldsymbol{\delta}/\sqrt{n} + \mathbf{f}_\alpha) - (\mathbf{u} + \boldsymbol{\gamma})\}^\top \mathbf{R}_0^{-1}\{\mathbf{R}_1(\boldsymbol{\delta}/\sqrt{n} + \mathbf{f}_\alpha) - (\mathbf{u} + \boldsymbol{\gamma})\} \\ &= n\{\boldsymbol{\delta}^\top \mathbf{R}_1^\top \mathbf{R}_0^{-1}\mathbf{R}_1\boldsymbol{\delta}/n + 2\boldsymbol{\delta}^\top \mathbf{R}_1^\top \mathbf{R}_0^{-1}(\mathbf{R}_1\mathbf{f}_\alpha - \mathbf{u} - \boldsymbol{\gamma})/\sqrt{n}\} + n\,\mathbf{g}_\alpha^\top \mathbf{R}_0 \mathbf{g}_\alpha \\ &= \boldsymbol{\delta}^\top \mathbf{R}_1^\top \mathbf{R}_0^{-1}\mathbf{R}_1\boldsymbol{\delta} + 2\sqrt{n}\,\boldsymbol{\delta}^\top \mathbf{R}_1^\top \mathbf{g}_\alpha + n\,\mathbf{g}_\alpha^\top \mathbf{R}_0 \mathbf{g}_\alpha, \end{aligned}$$

where $n\,\mathbf{g}_\alpha^\top \mathbf{R}_0 \mathbf{g}_\alpha$ is a constant term not depending on $\boldsymbol{\delta}$.

Thanks to the asymptotic expansions of $S'_\alpha(\boldsymbol{\delta})$ and $S''_\alpha(\boldsymbol{\delta})$ and to the equivalent reformulation of the penalty term $nP(\boldsymbol{\delta}/\sqrt{n} + \mathbf{f}_\alpha)$, for n sufficiently large, we can write

$$\begin{aligned} nJ_\alpha(\boldsymbol{\delta}/\sqrt{n} + \mathbf{f}_\alpha) &= (1/2)\,\boldsymbol{\delta}^\top \mathbf{D}_{1,n} \boldsymbol{\delta} - \boldsymbol{\delta}^\top \boldsymbol{\eta}_n + (\lambda/2)\,\boldsymbol{\delta}^\top \mathbf{R}_1^\top \mathbf{R}_0^{-1} \mathbf{R}_1 \boldsymbol{\delta} + \lambda\sqrt{n}\,\boldsymbol{\delta}^\top \mathbf{R}_1^\top \mathbf{g}_\alpha + o_{\mathrm{p}}(1) \\ &= (1/2)\,\boldsymbol{\delta}^\top (\mathbf{D}_{1,n} + \lambda \mathbf{R}_1^\top \mathbf{R}_0^{-1} \mathbf{R}_1)\boldsymbol{\delta} - \boldsymbol{\delta}^\top (\boldsymbol{\eta}_n - \lambda\sqrt{n}\,\mathbf{R}_1^\top \mathbf{g}_\alpha) + o_{\mathrm{p}}(1). \end{aligned}$$

Assumption 2 guarantees, for n large enough, that the matrix $\mathbf{D}_{1,n}$ is positive definite, so that $nJ_\alpha(\boldsymbol{\delta}/\sqrt{n} + \mathbf{f}_\alpha)$ is strictly convex in $\boldsymbol{\delta}$ and has a unique asymptotic minimiser $\hat{\boldsymbol{\delta}}_\alpha$. Then, the asymptotic estimator $\hat{\boldsymbol{\delta}}_\alpha$ may be obtained as the solution to the first-order equation $(\mathbf{D}_{1,n} + \lambda \mathbf{R}_1^\top \mathbf{R}_0^{-1} \mathbf{R}_1)\boldsymbol{\delta} = (\boldsymbol{\eta}_n - \lambda\sqrt{n}\,\mathbf{R}_1^\top \mathbf{g}_\alpha) + o_{\mathrm{p}}(1)$, that is

$$\hat{\boldsymbol{\delta}}_\alpha = (\mathbf{D}_{1,n} + \lambda \mathbf{R}_1^\top \mathbf{R}_0^{-1} \mathbf{R}_1)^{-1} (\boldsymbol{\eta}_n - \lambda\sqrt{n}\,\mathbf{R}_1^\top \mathbf{g}_\alpha) + o_{\mathrm{p}}(1).$$

As a consequence [33], $\hat{\boldsymbol{\delta}}_\alpha$ has an asymptotically normal distribution with mean and variance

$$\begin{aligned} \mathrm{E}(\hat{\boldsymbol{\delta}}_\alpha) &= -\lambda\sqrt{n}\,(\mathbf{D}_{1,n} + \lambda \mathbf{R}_1^\top \mathbf{R}_0^{-1} \mathbf{R}_1)^{-1} \mathbf{R}_1^\top \mathbf{g}_\alpha + o(1), \\ \mathrm{Var}(\hat{\boldsymbol{\delta}}_\alpha) &= (\mathbf{D}_{1,n} + \lambda \mathbf{R}_1^\top \mathbf{R}_0^{-1} \mathbf{R}_1)^{-1}\, \mathbf{D}_{0,n} (\mathbf{D}_{1,n} + \lambda \mathbf{R}_1^\top \mathbf{R}_0^{-1} \mathbf{R}_1)^{-1} + o(1). \end{aligned} \tag{21}$$

To obtain a non-exploding bias for the estimator $\hat{\mathbf{f}}_\alpha$, we require that $\sqrt{n}\lambda \to \bar{\lambda}$ for some finite value $\bar{\lambda} > 0$, which implies $\lambda \to 0$. Under this condition, the large-sample bias and variance of the estimator $\hat{\mathbf{f}}_\alpha$ are $\mathrm{Bias}(\hat{\mathbf{f}}_\alpha) = \mathrm{E}(\hat{\boldsymbol{\delta}}_\alpha)/\sqrt{n}$ and $\mathrm{Var}(\hat{\mathbf{f}}_\alpha) = \mathrm{Var}(\hat{\boldsymbol{\delta}}_\alpha)/n$. Following [21], we expand the bias term with a second-order Taylor approximation around $\lambda = 0$, which leads to

$$\mathrm{Bias}(\hat{\mathbf{f}}_\alpha) = -\lambda\Big\{\mathbf{D}_{1,n}^{-1} - \lambda \mathbf{D}_{1,n}^{-1} (\mathbf{R}_1^\top \mathbf{R}_0^{-1} \mathbf{R}_1) \mathbf{D}_{1,n}^{-1} + O(\lambda^2)\Big\} \mathbf{R}_1^\top \mathbf{g}_\alpha + o(n^{-1/2}),$$

that is, $\mathrm{Bias}(\hat{\mathbf{f}}_\alpha) = O(\lambda) + o(n^{-1/2})$. In the same way, the variance can be expanded as

$$\mathrm{Var}(\hat{\mathbf{f}}_\alpha) = \Big\{\mathbf{D}_{1,n}^{-1} \mathbf{D}_{0,n} \mathbf{D}_{1,n}^{-1} - 2\lambda \mathbf{D}_{1,n}^{-1} (\mathbf{R}_1^\top \mathbf{R}_0^{-1} \mathbf{R}_1)\, (\mathbf{D}_{1,n}^{-1} \mathbf{D}_{0,n} \mathbf{D}_{1,n}^{-1}) + O(\lambda^2)\Big\} + o(n^{-1}),$$

which leads to $\mathrm{Var}(\hat{\mathbf{f}}_\alpha) = O(n^{-1}) + O(\lambda n^{-1})$. Considering the Taylor expansions of $\mathrm{Bias}(\hat{\mathbf{f}}_\alpha)$ and $\mathrm{Var}(\hat{\mathbf{f}}_\alpha)$, using the asymptotic linearity of the estimator, and setting $\lambda\sqrt{n} \to \bar{\lambda}$, the asymptotic distribution of $\hat{\mathbf{f}}_\alpha$ can be written as

$$\sqrt{n}\,(\hat{\mathbf{f}}_\alpha - \mathbf{f}_\alpha) \xrightarrow{\mathrm{d}} \mathrm{N}_{N_h}(-\bar{\lambda}\,\mathbf{D}_1^{-1} \mathbf{R}_1^\top \mathbf{g}_\alpha,\, \mathbf{D}_1^{-1} \mathbf{D}_0\, \mathbf{D}_1^{-1}).$$

If we further assume that $\lambda = o(n^{-1/2})$, that is $\bar{\lambda} = 0$, the estimator $\sqrt{n}\,(\hat{\mathbf{f}}_\alpha - \mathbf{f}_\alpha)$ is asymptotically unbiased.

Finally, we prove the consistency of the estimator $\hat{\mathbf{f}}_\alpha$ by studying the limiting behaviour of the Mean Squared Error (MSE) and leveraging the results obtained for $\mathrm{Bias}(\hat{\mathbf{f}}_\alpha)$ and $\mathrm{Var}(\hat{\mathbf{f}}_\alpha)$:

$$\mathrm{MSE}(\hat{\mathbf{f}}_\alpha) = \mathrm{Var}(\hat{\mathbf{f}}_\alpha) + \mathrm{Bias}_n(\hat{\mathbf{f}}_\alpha)\, \mathrm{Bias}(\hat{\mathbf{f}}_\alpha)^\top = O(\lambda^2) + O(n^{-1}) + O(\lambda n^{-1}).$$

Then, the proof is concluded by noting that, for any $\lambda = O(n^{-1/2})$, we obtain a convergent MSE with limiting rate $O(n^{-1})$. □

Acknowledgments. We are grateful to the Associate Editor and Referees for insightful comments. 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 Ministry for University and Research (MUR). L.M. Sangalli also acknowledges the MUR research project Dipartimento di Eccellenza 2023–2027, Dipartimento di Matematica, Politecnico di Milano, Italy. A. Farcomeni acknowledges the PRIN research project 2022LANNKC CUP E53D23005810006. A. Farcomeni has been partially funded by the PRIN funding scheme of the Italian Ministry of University and Research (Grant No. 2022LANNKC CUP E53D23005810006). M. Bernardi acknowledges partial funding by the BERN BIRD2222 01 - BIRD 2022 grant from the University of Padua, and by the European Union - Next Generation EU, Mission 4 Component 2 - CUP C53D23002580006 via the MUR-PRIN grant 2022SMNNKY.

References

[1] E. Arnone, L. Azzimonti, F. Nobile, L. M. Sangalli, Modeling spatially dependent functional data via regression with differential regularization, J. Multivariate Anal. 170 (2019) 275–295.

[2] E. Arnone, C. De Falco, L. Formaggia, G. Meretti, L. M. Sangalli, Computationally efficient techniques for spatial regression with differential regularization, Int. J. Comput. Math. 100 (2023) 1971–1991.

[3] E. Arnone, L. M. Sangalli, E. Lila, J. Ramsay, L. Formaggia, G. Ardenghi, A. Clemente, A. Colli, A. Colombo, L. Colombo, C. de Falco, E. Dall'Acqua, G. Ferla, L. Ghilotti, J. Kim, M. Massardi, G. Meretti, G. Perin, C. Pigolotti, A. Poiatti, G. M. Rinaldi, S. Spaziani, A. Vicini, `fdaPDE`: Functional Data Analysis and Partial Differential Equations (PDE); Statistical Analysis of Functional and Spatial Data, Based on Regression with PDE Regularization, 2022. R package version 1.1.8.

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

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

[6] M. S. Bernardi, M. Carey, J. O. Ramsay, L. M. Sangalli, Modeling spatial anisotropy via regression with partial differential regularization, J. Multivariate Anal. 167 (2018) 15–30.

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

[8] C. M. Bishop, Pattern Recognition and Machine Learning, Information Science and Statistics, Springer, New York, 2006.

[9] H. D. Bondell, B. J. Reich, H. Wang, Noncrossing quantile regression curve estimation, Biometrika 97 (2010) 825–838.

[10] C. de Boor, A Practical Guide to Splines, volume 27 of Applied Mathematical Sciences, Springer-Verlag, New York-Berlin, 1978.

[11] R. J. Bosch, Y. Ye, G. G. Woodworth, A convergent algorithm for quantile regression with smoothing splines, Comput. Statist. Data Anal. 19 (1995) 613–630.

[12] M. Cavazzutti, E. Arnone, F. Ferraccioli, L. Finos, L. M. Sangalli, Sign-flip inference for spatial regression with differential regularization, Stat 13 (2024) e711.

[13] A. P. Dempster, N. M. Laird, D. B. Rubin, Maximum likelihood from incomplete data via the EM algorithm, J. Roy. Statist. Soc. Ser. B 39 (1977) 1–38. With discussion.

[14] G. Dubois, J. Malczewski, M. De Cort, Mapping radioactivity in the environment: Spatial interpolation comparison 97, Office for Official Publications of the European Communities, Luxembourg, 2003.

[15] W. Ehm, T. Gneiting, A. Jordan, F. Krüger, Of quantiles and expectiles: consistent scoring functions, Choquet representations and forecast rankings, J. R. Stat. Soc. Ser. B. Stat. Methodol. 78 (2016) 505–562.

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

[17] L. C. Evans, Partial differential equations, volume 19 of Graduate Studies in Mathematics, American Mathematical Society, Providence, RI, second edition, 2010.

[18] A. Farcomeni, M. Geraci, C. Viroli, Directional quantile classifiers, Journal of Computational and Graphical Statistics 31 (2022) 907–916.

[19] M. Fasiolo, S. N. Wood, M. Zaffran, R. Nedellec, Y. Goude, Fast Calibrated Additive Quantile Regression, J. Amer. Statist. Assoc. 116 (2021) 1402–1412.

[20] M. Fasiolo, S. N. Wood, M. Zaffran, R. Nedellec, Y. Goude, `qgam`: Bayesian nonparametric quantile regression modeling in `R`, J. Stat. Softw. 100 (2021) 1–31.

[21] F. Ferraccioli, L. M. Sangalli, L. Finos, Some first inferential tools for spatial regression with differential regularization, J. Multivariate Anal. 189 (2022) Paper No. 104866, 12.

[22] F. Ferraccioli, L. M. Sangalli, L. Finos, Nonparametric tests for semiparametric regression models, TEST 32 (2023) 1106–1130.

[23] D. A. Freedman, On the so-called "Huber sandwich estimator" and "robust standard errors", Amer. Statist. 60 (2006) 299–302.

[24] P. Frumento, M. Bottai, Parametric modeling of quantile regression coefficient functions, Biometrics 72 (2016) 74–84.

[25] P. Frumento, M. Bottai, I. Fernández-Val, Parametric modeling of quantile regression coefficient functions with longitudinal data, J. Amer. Statist. Assoc. 116 (2021) 783–797.

[26] H. Garcia, T. Boyer, O. Baranova, R. Locarnini, A. Mishonov, A. Grodsky, C. Paver, K. Weathers, I. Smolyar, J. Reagan, D. Seidov, M. Zweng, World Ocean Atlas 2018: Product Documentation, Technical Report, NOAA National Centers for Environmental Information, 2019. A. Mishonov, Technical Editor.

[27] M. Geraci, Additive quantile regression for clustered data with an application to children's physical activity, J. R. Stat. Soc. Ser. C Appl. Stat. (2019) 1071–1089.

[28] M. Geraci, Modelling and estimation of nonlinear quantile regression with clustered data, Comput. Statist. Data Anal. 136 (2019) 30–46.

[29] T. Gneiting, Quantiles as optimal point forecasts, International Journal of forecasting 27 (2011) 197–207.

[30] R. J. Gray, Spline-based tests in survival analysis, Biometrics 50 (1994) 640–652.

[31] M. Hallin, Z. Lu, K. Yu, Local linear spatial quantile regression, Bernoulli 15 (2009) 659–686.

[32] X. He, P. Ng, S. Portnoy, Bivariate quantile smoothing splines, J. R. Stat. Soc. Ser. B Stat. Methodol. 60 (1998) 537–550.

[33] N. L. Hjort, D. Pollard, Asymptotics for minimisers of convex processes, arXiv preprint arXiv:1107.3806 (2011).

[34] K. Knight, Limiting distributions for L_1 regression estimators under general conditions, Ann. Statist. 26 (1998) 755–770.

[35] K. Knight, W. Fu, Asymptotics for lasso-type estimators, Ann. Statist. 28 (2000) 1356–1378.

[36] R. Koenker, Quantile Regression, volume 38 of Econometric Society Monographs, Cambridge University Press, Cambridge, 2005.

[37] R. Koenker, `quantreg`: Quantile Regression, 2021. R package version 5.86.

[38] R. Koenker, G. Bassett, Jr., Regression quantiles, Econometrica 46 (1978) 33–50.

[39] R. Koenker, I. Mizera, Penalized triograms: total variation regularization for bivariate smoothing, J. R. Stat. Soc. Ser. B Stat. Methodol. 66 (2004) 145–163.

[40] R. Koenker, P. Ng, S. Portnoy, Quantile smoothing splines, Biometrika 81 (1994) 673–680.

[41] S. Kotz, T. J. Kozubowski, K. Podgórski, The Laplace Distribution and Generalizations. A Revisit with Applications to Communications, Economics, Engineering, and Finance, Birkhäuser Boston, Inc., Boston, MA, 2001.
[42] H. Kozumi, G. Kobayashi, Gibbs sampling methods for Bayesian quantile regression, J. Stat. Comput. Simul. 81 (2011) 1565–1578.
[43] M.-J. Lai, L. Wang, Bivariate penalized splines for regression, Statist. Sinica 23 (2013) 1399–1417.
[44] K. Lange, Optimization, volume 95 of Springer Texts in Statistics, Springer, New York, second edition, 2013.
[45] Y. Li, Y. Liu, J. Zhu, Quantile regression in reproducing kernel Hilbert spaces, J. Amer. Statist. Assoc. 102 (2007) 255–268.
[46] E. Lila, J. A. D. Aston, L. M. Sangalli, Smooth principal component analysis over two-dimensional manifolds with an application to neuroimaging, Ann. Appl. Stat. 10 (2016) 1854–1879.
[47] C. J. M. Maas, J. J. Hox, Robustness issues in multilevel regression analysis, Statist. Neerlandica 58 (2004) 127–137.
[48] P. McCullagh, J. A. Nelder, Generalized Linear Models. Second edition, Monographs on Statistics and Applied Probability, Chapman & Hall, London, 1989.
[49] G. J. McLachlan, T. Krishnan, The EM algorithm and extensions, Wiley Series in Probability and Statistics, Wiley-Interscience [John Wiley & Sons], Hoboken, NJ, second edition, 2008.
[50] V. M. R. Muggeo, F. Torretta, P. H. C. Eilers, M. Sciandra, M. Attanasio, Multiple smoothing parameters selection in additive regression quantiles, Stat. Model. 21 (2021) 428–448.
[51] P. T. Ng, An algorithm for quantile smoothing splines, Comput. Statist. Data Anal. 22 (1996) 99–118.
[52] J. Nocedal, S. J. Wright, Numerical optimization, Springer Series in Operations Research and Financial Engineering, Springer, New York, second edition, 2006.
[53] B. Nortier, Automated Smoothing Parameter Estimation for Quantile Additive Models, Ph.D. thesis, University of Bristol, 2021.
[54] D. Nychka, G. Gray, P. Haaland, D. Martin, M. O'Connell, A nonparametric regression approach to syringe grading for quality improvement, J. Amer. Statist. Assoc. 90 (1995) 1171–1178.
[55] D. Pollard, Asymptotics for least absolute deviation regression estimators, Econometric Theory 7 (1991) 186–199.
[56] N. G. Polson, S. L. Scott, Data augmentation for support vector machines, Bayesian Anal. 6 (2011) 1–23.
[57] A. Quarteroni, Numerical Models for Differential Problems, volume 16 of MS&A. Modeling, Simulation and Applications, Springer, Cham, third edition, 2017.
[58] T. Ramsay, Spline smoothing over difficult regions, J. R. Stat. Soc. Ser. B Stat. Methodol. 64 (2002) 307–319.
[59] P. T. Reiss, L. Huang, Smoothness selection for penalized quantile regression splines, Int. J. Biostat. 8 (2012) Art. 10, front matter+25.
[60] R. T. Rockafellar, Convex analysis, Princeton Landmarks in Mathematics, Princeton University Press, Princeton, NJ, 1997. Reprint of the 1970 original, Princeton Paperbacks.
[61] Y. Romano, E. Patterson, E. Candes, Conformalized quantile regression, Adv. Neural. Inf. Process. Syst. 32 (2019).
[62] J. Romero, A. Madrid, J. Angulo, Quantile-based spatiotemporal risk assessment of exceedances, Stoch. Environ. Res. Risk Assess. 32 (2018) 2275–2291.
[63] L. M. Sangalli, Spatial regression with partial differential equation regularisation, International Statistical Review (2021).
[64] L. M. Sangalli, J. O. Ramsay, T. O. Ramsay, Spatial spline regression models, J. R. Stat. Soc. Ser. B. Stat. Methodol. 75 (2013) 681–703.
[65] M. Schlather, A. Malinowski, M. Oesting, D. Boecker, K. Strokorb, S. Engelke, J. Martini, F. Ballani, O. Moreva, J. Auel, P. J. Menck, S. Gross, U. Ober, C. Berreth, K. Burmeister, J. Manitz, P. Ribeiro, R. Singleton, P. Ben, R Core Team, `RandomFields`: Simulation and Analysis of Random Fields, 2017. R package version 3.1.50.
[66] S. K. Schnabel, P. H. C. Eilers, Simultaneous estimation of quantile curves using quantile sheets, AStA Adv. Stat. Anal. 97 (2013) 77–87.
[67] W. Shin, M. Kim, Y. Jung, Efficient information-based criteria for model selection in quantile regression, J. Korean Statist. Soc. 51 (2022) 245–281.
[68] B. W. Silverman, Some aspects of the spline smoothing approach to nonparametric regression curve fitting, J. Roy. Statist. Soc. Ser. B 47 (1985) 1–52. With discussion.
[69] Y. Sun, H. J. Wang, M. Fuentes, Fused adaptive lasso for spatial and temporal quantile function estimation, Technometrics 58 (2016) 127–137.
[70] K. M. Tan, L. Wang, W.-X. Zhou, High-dimensional quantile regression: convolution smoothing and concave regularization, J. R. Stat. Soc. Ser. B. Stat. Methodol. 84 (2022) 205–233.
[71] Y. Tian, M. Tian, Q. Zhu, Linear quantile regression based on EM algorithm, Comm. Statist. Theory Methods 43 (2014) 3464–3484.
[72] G. Wahba, Spline models for observational data, volume 59 of CBMS-NSF Regional Conference Series in Applied Mathematics, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1990.
[73] L. Wang, G. Wang, M.-J. Lai, L. Gao, Efficient estimation of partially linear models for data on complicated domains by bivariate penalized splines over triangulations, Statist. Sinica 30 (2020) 347–369.
[74] L. Wang, Y. Wu, R. Li, Quantile regression for analyzing heterogeneity in ultra-high dimension, J. Amer. Statist. Assoc. 107 (2012) 214–222.
[75] M. Wilhelm, L. M. Sangalli, Generalized spatial regression with differential regularization, J. Stat. Comput. Simul. 86 (2016) 2497–2518.
[76] S. N. Wood, Thin plate regression splines, J. R. Stat. Soc. Ser. B Stat. Methodol. 65 (2003) 95–114.
[77] S. N. Wood, Generalized Additive Models. An Introduction with R, Second edition, Texts in Statistical Science Series, CRC Press, Boca Raton, FL, 2017.
[78] S. N. Wood, M. V. Bravington, S. L. Hedley, Soap film smoothing, J. R. Stat. Soc. Ser. B Stat. Methodol. 70 (2008) 931–955.
[79] K. Yu, M. C. Jones, Local linear quantile regression, J. Amer. Statist. Assoc. 93 (1998) 228–237.
[80] K. Yu, R. A. Moyeed, Bayesian quantile regression, Statist. Probab. Lett. 54 (2001) 437–447.
[81] Y. Yu, D. Ruppert, Penalized spline estimation for partially linear single-index models, J. Amer. Statist. Assoc. 97 (2002) 1042–1054.
[82] M. Yuan, GACV for quantile smoothing splines, Comput. Statist. Data Anal. 50 (2006) 813–829.
[83] M. Zweng, J. Reagan, D. Seidov, T. Boyer, R. Locarnini, H. Garcia, A. Mishonov, O. Baranova, K. Weathers, C. Paver, I. Smolyar, World Ocean Atlas 2018, Volume 2: Salinity, Technical Report, NOAA National Centers for Environmental Information, 2019. A. Mishonov, Technical Editor, NOAA Atlas NESDIS 82.

MOX Technical Reports, last issues

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

97/2024 Ferro, N.; Mezzadri, F.; Carbonaro, D.; Galligani, E.; Gallo, D.; Morbiducci, U.; Chiastra, C.; Perotto, S.
Designing novel vascular stents with enhanced mechanical behavior through topology optimization of existing devices

96/2024 Brivio, S.; Fresca, S.; Manzoni, A.
PTPI-DL-ROMs: Pre-trained physics-informed deep learning-based reduced order models for nonlinear parametrized PDEs

93/2024 Conti, P.; Kneifl, J.; Manzoni, A.; Frangi, A.; Fehr, J.; Brunton, S.L.; Kutz, J.N.
VENI, VINDy, VICI - a variational reduced-order modeling framework with uncertainty quantification

94/2024 Franco, N.R.; Fresca, S.; Tombari, F.; Manzoni, A.
Deep Learning-based surrogate models for parametrized PDEs: handling geometric variability through graph neural networks

95/2024 Zacchei, F.; Rizzini, F.; Gattere, G.; Frangi, A.; Manzoni, A.
Neural networks based surrogate modeling for efficient uncertainty quantification and calibration of MEMS accelerometers

91/2024 Ciaramella, G.; Kartmann, M.; Mueller, G.
Solving Semi-Linear Elliptic Optimal Control Problems with L1-Cost via Regularization and RAS-Preconditioned Newton Methods

Castiglionea, C.; Arnonec, E.; Bernardi, M.; Farcomeni, A.; Sangalli, L.M.
PDE regularised spatial quantile regression

88/2024 Regazzoni, F.; Poggesi, C.; Ferrantini, C.
Elucidating the cellular determinants of the end-systolic pressure-volume relationship of the heart via computational modelling

85/2024 Brivio, S.; Franco, Nicola R.; Fresca, S.; Manzoni, A.
Error estimates for POD-DL-ROMs: a deep learning framework for reduced order modeling of nonlinear parametrized PDEs enhanced by proper orthogonal decomposition

90/2024 Tomasetto, M.; Arnone, E.; Sangalli, L.M.
Modeling anisotropy and non-stationarity through physics-informed spatial regression