MOX-Report No. 32/2025

Exploring nitrogen dioxide spatial concentration via physics-informed multiple quantile regression

De Sanctis, M.F.; Di Battista, I.; Arnone, E.; Castiglione, C.; Palummo, A.; Bernardi, M.; Ieva, F.; Sangalli, L.M.

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

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

Published in Environmental and Ecological Statistics, 2025
Publisher version available at https://doi.org/10.1007/s10651-025-00664-8

Exploring nitrogen dioxide spatial concentration via physics-informed multiple quantile regression

Marco F. De Sanctis[1†], Ilenia Di Battista[1†], Eleonora Arnone[2], Cristian Castiglione[3], Alessandro Palummo[1], Mauro Bernardi[4], Francesca Ieva[1], Laura M. Sangalli[1*]

[1*]MOX, Dipartimento di Matematica, Politecnico di Milano, Italy.
[2]Dipartimento di Management, Università degli Studi di Torino, Italy.
[3]Bocconi Institute for Data Science and Analytics, Bocconi University.
[4]Dipartimento di Scienze Statistiche, Università degli Studi di Padova.

*Corresponding author(s). E-mail(s): laura.sangalli@polimi.it;
†These authors contributed equally to this work.

Abstract

Understanding the spatial distribution of air pollutants, such as nitrogen dioxide (NO_2), is crucial for assessing environmental and health impacts, particularly in densely populated and industrialized regions. This paper introduces a novel method for estimating multiple spatial quantiles, ensuring the monotonicity of the resulting estimates. The proposed model builds upon recent advancements in quantile regression, and incorporates physical information of the phenomenon under analysis, to address the challenges posed by anisotropy, non-stationarity and skewness, typically observed in environmental data. For instance, in the study of air pollutants concentration, the model permits the inclusion of information concerning air-circulation, and in particular the physics of wind streams, which strongly influences the pollutant concentration. Moreover, the monotone estimation of the quantile maps yields a fully nonparametric reconstruction of the pollutant probability density function, at any spatial location. This in turn enables the construction of probability maps, that quantify the likelihood of exceeding regulatory thresholds set by policymakers, offering valuable information for environmental monitoring policies, aimed at mitigating the adverse effects of air pollution.

Keywords: Air pollution, nonparametric regression, simultaneous quantile estimation.

1 Introduction

Investigating quantiles of phenomena with complex spatial characteristics is highly relevant across numerous disciplines, especially in environmental applications. In particular, examining a sequence of multiple quantiles is fundamental to flexibly characterizing the distribution of the variable at hand. Moreover, this is essential when investigating the tail of the distribution, that is crucial whenever studying high-impact, low-probability events, such as high concentrations of air pollutants, or particularly high temperatures, where the emphasis is on tail characteristics, rather than the central tendency of the phenomenon. An example is provided in Figure 1, which illustrates the daily maximum concentrations of nitrogen dioxide (NO_2) in Lombardy region, on the 11^{th} January 2019. Data are collected from the 84 monitoring stations, managed by Agenzia Regionale per la Protezione dell'Ambiente (ARPA Lombardia, 2024), and located on the territory on the basis of population density. As typically happens in environmental phenomena, these data are characterized by strong heteroscedasticity, skewness, and kurtosis. In the case of NO_2, the distribution of its concentration deviates significantly from the Gaussian, making spatial linear regression methods unsuitable for analysis. Certain areas within the domain, in particular, display anomalously high NO_2 values. Those spikes can be captured by estimating quantiles of the right tail of the distribution, which is of particular concern to scientists monitoring health conditions of the local population. Indeed, as reported, e.g., by the European Environmental Agency, inhaling NO_2 is linked to numerous health issues, ranging from eye and throat irritation to acute respiratory conditions, impaired lung function, and an elevated risk of premature death. Beyond its detrimental effects on human health, NO_2 is also toxic to vegetation, contributing to forest damage and decreased crop

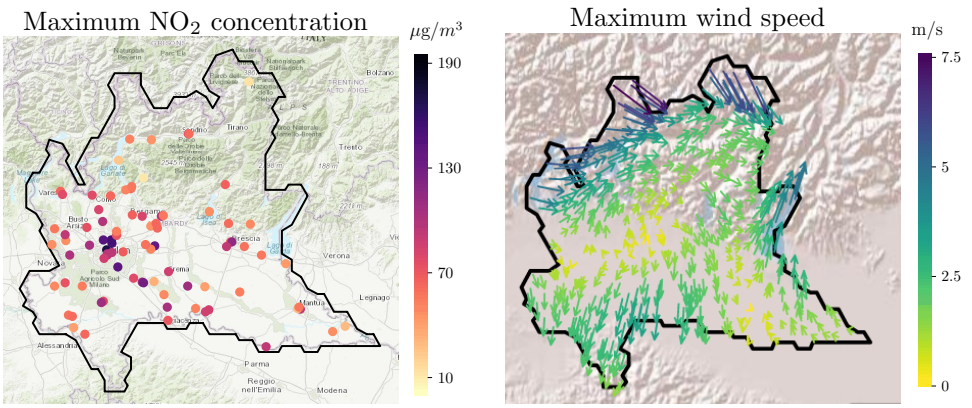


Fig. 1: Left: daily maximum of NO_2 concentration measurements, collected at the 84 control units managed by Agenzia Regionale per la Protezione dell'Ambiente (ARPA), on 11^{th} January 2019. Right: maximum wind speed measurements in Lombardy, collected over 119 ARPA meteorological stations, on 11^{th} January 2019; the direction and the length of the arrows indicate the local direction and intensity of the wind stream.

productivity; see, e.g., the report published by the European Environment Agency (2018).

Additionally, challenges arising from the environmental context are often marked by strong anisotropy or non-stationarity, that complicate the spatial pattern observed in the data. For instance, the dispersion of air pollutants, such as NO_2, is heavily influenced by wind streams, that play a crucial role in disseminating the pollutants, thereby shaping their spatial pattern. The right panel of Figure 1 shows the wind stream field, on the same day of the NO_2 measurements in the left-hand panel, highlighting the strong non-stationarity of such stream effect. The embedding of the physics of wind streams in the statistical modeling of air pollutant's concentration is highly desirable, as it is crucial to appropriately account for these complexities, and this can in turn advance the accuracy of estimation methods, and lead to meaningful insights on the complex phenomena under study.

In this work, we propose a new spatial multiple quantile estimation method, that enjoys some fundamental and advantageous modeling features. One key feature is the guarantee of monotonicity of the estimated quantiles, that in turn enables the study of tail behaviors; this is achieved by imposing appropriate non-crossing constraints between consecutive quantiles. An additional strength is the ability to include the available physical information on the underlying phenomenon, such as the presence of the wind streams in the study of air pollutants. This is achieved through a flexible penalized regression approach, where the penalty involves a Partial Differential Equation, that encodes the available problem-specific information, thus enabling the modeling of a wide range of physical phenomena, ultimately enhancing the accuracy of the model. Additionally, the proposed method can handle data observed over spatial regions with complex shapes, such as land regions or water bodies with irregular coastlines, or curved surfaces, which is fundamental when the phenomenon under study is influenced by the conformation of the domain.

The remainder of this introduction is organized as follows. In Section 1.1 we briefly review spatial data analysis approaches that permit to include some problem-specific physical information. In Section 1.2 we discuss the crossing problem, that arises in the estimation of multiple quantiles, and may prevent the coherent study of tail behaviors. Finally, in Section 1.3 we briefly outline some key aspects of our proposal.

1.1 Physics-informed spatial statistics approaches

In recent years, there has been an increasing interest in the statistical literature for physics-informed modeling of spatial and spatio-temporal data. A powerful way to include physical information in statistical models is through differential equations, and in particular through Partial Differential Equations (PDE), when the quantities of interest have spatial or spatio-temporal dependence. A well-developed literature in the spatio-temporal framework is the one of hierarchical models, in which a time-depending PDE is used to inform the spatio-temporal variation; see, e.g., Wikle (2003), Wikle and Hooten (2010), Cressie and Wikle (2011), Kuhnert (2014), Richardson (2017) and Hefley et al. (2017). Additionally, since the seminal work of Lindgren et al. (2011), the powerful stochastic PDE (sPDE) approach has been extended in various directions (see, e.g., the review in Lindgren et al., 2022), including the ability to work

over irregularly shaped domains (see, e.g., Bakka et al., 2019) and the integration of physics information (see, e.g., Carrizo-Vergara et al., 2022; Clarotto et al., 2024). In particular, Clarotto et al. (2024) proposes a spatio-temporal sPDE model for the estimation of the solar radiation, that considers an advection-diffusion PDE to account for the effect of wind stream.

Another class of physics-informed spatial data analysis methods is the PDE-regularized approach reviewed in Sangalli (2021). We refer the reader to, e.g., Azzimonti et al. (2015) and Arnone et al. (2019), for applications to problems in the life sciences, and to, e.g., Tomasetto et al. (2024) for application to environmental problems. For instance, Tomasetto et al. (2024) shows an application to rainfall data, considering the presence of wind, and a second application to the study of oceanographic measurements, where the underlying physics is that of ocean currents. This PDE-regularized approach offers greater flexibility compared to traditional geostatistical methods. Firstly, rather than explicitly imposing a specific structure on the spatial correlation of the observations, the spatial correlation is implicitly induced by the PDE regularization term, enabling the modeling of strong forms of anisotropy and non-stationarity. Moreover, the regularizing PDE permits the integration of physical knowledge into the modeling process, which can be crucial in real data applications, especially when the observed data are scarse or clustered in space. Additionally, as highlighted in various studies (see, e.g., Sangalli, 2021; Tomasetto et al., 2024) this method permits the accurate modeling of data scattered over spatial regions with complicated geometries, including domains with irregular boundaries and curved regions.

However, the aforementioned studies target mean estimation, rather than quantile estimation. A first approach to model spatial quantiles, including physical information, is the PDE-regularized spatial quantile regression method recently proposed by Castiglione et al. (2025). This approach is though limited to the estimation of spatial quantiles only for a single-quantile level at a time.

1.2 Estimation of non-crossing quantiles

As previously anticipated, investigating tail behaviors requires the computation of several, closely spaced, quantiles. This prevents the use of single-quantile estimation procedures, since they do not guarantee the monotonicity of the quantiles. Such issue, commonly referred to as the quantile crossing problem (see, e.g., Bassett and Koenker, 1982; He, 1997), has attracted a lot of interest in the literature, in more classical univariate regression settings (without any spatial dependence). In particular, He (1997) introduced non-crossing constraints by restricting Koenker and Bassett (1978)'s model, but this results in a location-scale model often too restrictive for real-world applications (Neocleous and Portnoy, 2008). Methods like Mammen (1991) employ isotonization to project estimates onto monotone functions. Other techniques parametrize the space of monotone quantile planes (see, e.g., Yang and Tokdar, 2015), or embed non-crossing constraints in estimation, such as the kernel methods of Takeuchi et al. (2006), or parametric and nonparametric models by Bondell et al. (2010). Other works, like Liu and Wu (2009) and Liu and Wu (2011), introduced constrained schemes, with

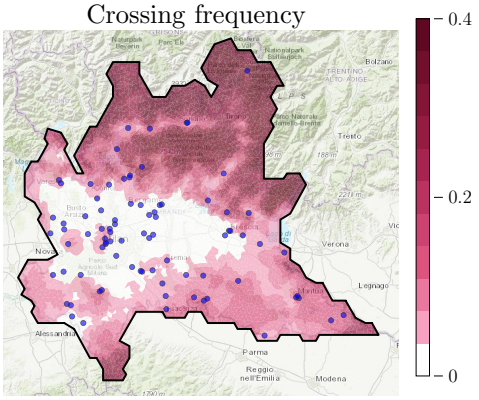


Fig. 2: Frequency of quantile crossing. The results are obtained using the single-quantile estimation procedure in Castiglione et al. (2025), on a grid of 27 quantiles, from 1% to 99%. The violation of the monotonicity constraints, highlighted in reddish tones, is more pronounced in the North-East region, where data scarcity is more severe.

the latter imposing regression coefficient constraints. This extensive literature has so far almost exclusively focused on univariate data analysis settings.

In the context of spatially dependent data, the quantile crossing problem becomes even more severe. Indeed, quantile crossing is particularly pronounced in the spatial regression, where quantiles are likely to intersect in regions of the domain where observations are sparse. Figure 2 provides an example where the scarcity of monitoring stations in the northeastern part of the region is particularly notable, causing the quantile crossing. A first attempt to tackle the problem of estimating space-varying quantiles is offered by Reich et al. (2011). However, as noted in Das and Ghosal (2017) and in Deb et al. (2024), the estimation is conducted independently at each spatial location, failing to account for the correlation between nearby points, making this model unsuitable for data that exhibit complex spatial patterns.

A possible approach for estimating monotone space-varying quantiles involves post-processing strategies applied on single-quantile spatial estimates. For instance Fasiolo et al. (2021a) present single-quantile spatial models, based on thin plate spline (TPS; see Wood, 2003) and on the soap film smoothing (SOAP; see Wood et al., 2008), that can be subsequently refined, to ensure monotonicity of the quantiles, using the approach detailed in Chernozhukov et al. (2010), which rearranges the estimates permuting the consecutive crossing quantiles. Among the spatial quantile regression models, for which it is possible to apply the rearrangement procedure of Chernozhukov et al. (2010) to guarantee the non-crossing, we also mention the neural network model proposed by Cannon (2018), and the parametric extreme value regression model proposed in Youngman (2019).

Alternatively, some non-crossing quantile regression models, not originally designed to account for spatially correlated data, can incorporate spatial information by employing a spatial smoother. The SOAP smoother can for instance be used as regressor in the Quantile Regression Coefficient Modeling (QRCM), developed in Frumento et al.

(2021) and in Sottile and Frumento (2023). The latter is a linear quantile regression model where the non-crossing is imposed in the optimization algorithm, through appropriate constraints on the parametric regression coefficients.

1.3 Our proposal

In this paper, we introduce an innovative method for addressing the lack of monotonicity in conditional quantiles of spatially distributed data. The proposed method generalizes the nonparametric approach introduced in Castiglione et al. (2025), and permits to integrate prior information about the underlying phenomenon to simultaneously estimate multiple quantiles, guaranteeing monotonicity of the estimates. Instead of considering non-crossing constraints, we show that the estimation problem can equivalently be expressed as a more convenient unconstrained problem, where the non-crossing constraints have been replaced by appropriate penalization terms.

Moreover, with respect to Castiglione et al. (2025), we enrich the model by also including space-varying covariates, through an appropriate parametric term in a semiparametric framework. In the application to NO_2 concentration analysis, this parametric term enables us to investigate the impact of key factors, such as anthropogenic influences and the geomorphology of the territory, on pollutant levels.

The considered estimation problem is non-quadratic, and it does not enjoy a closed-form solution. For this reason, we develop an appropriate iterative estimation procedure for its solution, that leverages the theory of Maximization-Minimization algorithms (see, e.g., Lange, 2016). The estimation procedure is implemented with a specific attention to computational efficiency, and exploits a spatial discretization based on the Finite Element Method: see, e.g., Quarteroni (2017) and Brezzi and Fortin (1991). The proposed method is implemented in the `fdaPDE` library (Palummo et al., 2025).

The article is organized as follows. In Section 2, we introduce the proposed method of Multiple Quantile Spatial Regression with Partial Differential Equation regularization (MQSR-PDE). In Section 3, we present the estimation procedure based on a Maximization-Minimization strategy and a Finite Element discretization. In Section 4 we report some simulation studies, that compare the accuracy of the proposed model with existing methods, showing the comparative advantages of the proposed approach. In Section 5, we demonstrate the application of the proposed model to the study of the dispersion of NO_2 in the Lombardy region. Finally, in Section 6, we summarize the paper's contributions and we discuss potential future research directions. The mathematical proofs of the theoretical results presented throughout the article are deferred to the Appendix.

2 Physics-Informed multiple quantile regression

Consider a set of spatial locations $\{\mathbf{p}_i\}_{i=1}^n$ within a bounded domain $\mathcal{D} \subset \mathbb{R}^2$. At these locations, we observe realizations $\{y_i\}_{i=1}^n$ of a real-valued random variable Y, with absolutely continuous distribution. Moreover, at the same spatial locations, we also observe a set of q covariates $\{\mathbf{x}_i\}_{i=1}^n$. Our goal is to estimate the conditional quantile maps, for a grid of r probability levels $\{\alpha_j\}_{j=1}^r \subset (0,1)$, ensuring their monotonicity.

For simplicity of exposition, in Section 2.1 we start by introducing the estimation functional relative to the single α_j-conditional quantile. In Section 2.2, we then consider multiple quantiles. In this section, we first introduce a multiple quantile estimation problem with non-crossing constraints, that ensure monotonicity of the quantiles; then we propose an equivalent, but computationally more convenient, penalized multiple quantile estimation problem, that avoids non-crossing constraints, while preserving monotonicity.

2.1 Semiparametric Physics-Informed α_j-th quantile estimation

We model the space-varying α_j-th conditional quantile of Y_i, for $j = 1, \dots, r$, as

$$Q_{Y_i|\mathbf{p}_i,\mathbf{x}_i}(\alpha_j) = \mathbf{x}_i^\top \boldsymbol{\beta}_j + f_j(\mathbf{p}_i), \quad j = 1, \dots, r,$$

where $\boldsymbol{\beta}_j \in \mathbb{R}^q$ quantifies the linear effect of the covariates on the response at the α_j-th quantile, and $f_j : \mathcal{D} \to \mathbb{R}$ is a nonparametric term that captures the spatial pattern of the α_j-th quantile. We here assume that the covariate matrix X has full rank and does not include the constant term, which is instead incorporated into the nonparametric component f. We search the nonparametric terms $\{f_j\}_{j=1}^r$ in the space $H^2(\mathcal{D})$, where $H^d(\mathcal{D}) = \{f \in L^2(\mathcal{D}) : D^k f \in L^2(\mathcal{D}),\ \forall |k| \leq d\}$ denotes the Sobolev space of order d, where $D^k f$ is the k-th weak derivative of f.

Extending to a semiparametric setting the purely nonparametric problem explored by Castiglione et al. (2025), we consider the following penalized estimation functional for the α_j-th conditional quantile:

$$J_{\alpha_j}(\boldsymbol{\beta}_j, f_j) = \frac{1}{n} \sum_{i=1}^{n} \rho_{\alpha_j}(y_i - \mathbf{x}_i^\top \boldsymbol{\beta}_j - f_j(\mathbf{p}_i)) \ + \lambda_j P(f_j), \tag{1}$$

where $\rho_\alpha(t) := 0.5|t| + (\alpha - 0.5)t$ denotes the pinball loss function, introduced by Koenker and Bassett (1978) for classical univariate quantile regression settings, while $P(f_j)$ is an appropriate penalty term. In particular, we here consider the Physics-Informed penalty

$$P(f_j) = \int_{\mathcal{D}} (Lf_j(\mathbf{p}) - u_j(\mathbf{p}))^2 \, d\mathbf{p},$$

where $Lf_j - u_j$ is a Partial Differential Equation (PDE), defined over the spatial domain $\mathcal{D}$, that encapsulates the available problem-specific information about the underlying phenomenon. Specifically, we here consider linear second order elliptic differential operators L of the form

$$Lf = -\nabla \cdot (K \nabla f) + \mathbf{b} \cdot \nabla f + cf, \tag{2}$$

where $\nabla = (\frac{\partial}{\partial p_1}, \frac{\partial}{\partial p_2})^\top$, $K \in \mathbb{R}^{2\times 2}$ is a symmetric positive definite matrix, that controls the anisotropy of the underlying phenomenon, $\mathbf{b} \in \mathbb{R}^2$ models unidirectional effects, while $c \in \mathbb{R}$ controls the shrinkage of the field f towards the zero function.

Finally, the space varying forcing term $u_j : \mathcal{D} \to \mathbb{R}$ modulates possible exogenous sources in the phenomenon under study. In this work, for simplicity of exposition, we consider homogeneous forcing terms, i.e., null functions u_j, referring the reader to, e.g., Azzimonti et al. (2014) for the case of nonhomogeneous forcing terms.

The three terms in the differential operator L in (2), that are called respectively diffusion, transport and reaction terms, permit to encode the available problem-specific information of the phenomenon under study. The coefficients K, $\mathbf{b}$, c can be function of the space, permitting the modeling of different non-stationary effects. For example, in the application to NO_2 data considered in Section 5, we include the wind stream information via a non-stationary transport term $\mathbf{b}(\mathbf{p})$, shown in the right panel of Figure 1, where the vector field $\mathbf{b}(\mathbf{p})$ encodes the local direction and speed of the wind. Encoding this information in the statistical model is fundamental to achieve physical meaningful estimates, especially in areas where no NO_2 measurements are available, like for example in the northern part of Lombardy. Other examples of the use of physics-informed penalties in the statistical analysis of environmental and ecological data are offered by, e.g., Castiglione et al. (2025) and Tomasetto et al. (2024), with application to rainfall data and to oceanographic measurements.

Moreover the PDE parameters K, $\mathbf{b}$ and c may depend on unknown hyperparameters that need to be estimated from the data. To this end, we adopt the profiling estimation technique considered by Bernardi et al. (2018) and Tomasetto et al. (2024), based on the parameter cascading algorithm (see, e.g., Ramsay et al., 2007; Xun et al., 2013). For example, in the application to the NO_2 estimation considered in Section 5, we employ the parameter cascading approach to estimate the relative strength of the natural diffusion of the pollutant in the air, encoded in an isotropic and stationary second-order term $-\nabla \cdot (I \nabla f)$, with respect to the unidirectional spreading driven by the wind, enclosed in a non-stationary first-order term $\mathbf{b} \cdot \nabla f$, where $\mathbf{b}$ is the vector field in Figure 1.

It should be pointed out that f does not need to solve the regularizing PDE $Lf - u$, as the PDE enters in the statistical model through a regularization term, whose relative weight with respect to the data loss ρ_{α_j} is controlled by the positive smoothing parameter λ_j. The smoothing parameters $\{\lambda_j\}_{j=1}^r$ are selected on the basis of data-driven criteria, as detailed in Section 4, and they vary according to the quantile order α_j. This approach allows for different regularizations tailored to the distinct quantile fields based on the observed data.

Finally, the PDE is coupled with boundary conditions. For the sake of simplicity, we here consider homogeneous Neumann boundary conditions, i.e., $K \nabla f \cdot \nu = 0$ on $\partial\mathcal{D}$. Various other forms of boundary conditions can also be specified, as detailed, e.g., in Arnone et al. (2019) for linear regression problems.

2.2 Semiparametric Physics-Informed multiple quantile estimation

Set $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_r)^\top$ and denote by $\boldsymbol{f}$ the collection of the spatial fields $f_1, \dots, f_r$, corresponding to the r quantile levels $\{\alpha_j\}_{j=1}^r \subset (0, 1)$.

A direct approach to estimate the set of quantiles $\{Q_{Y_i|\mathbf{p}_i,\mathbf{x}_i}(\alpha_j)\}_{j=1}^r$, while ensuring monotonicity of the estimated quantiles, would consist in minimizing the sum of

the loss functionals in (1), over the r quantile levels $\{\alpha_j\}_{j=1}^r$, subject to non-crossing constraints. Namely, we could consider the following constrained estimation problem

$$\begin{cases} \min\limits_{(\boldsymbol{\beta}_1,f_1),\dots,(\boldsymbol{\beta}_r,f_r)\in\mathbb{R}^q\times\mathcal{F}_0} \left\{\frac{1}{n}\sum_{i,j=1}^{n,r}\rho_{\alpha_j}(y_i-\mathbf{x}_i^\top\boldsymbol{\beta}_j-f_j(\mathbf{p}_i)) \;+\sum_{j=1}^r\lambda_j P(f_j)\right\} \\ \mathbf{x}_i^\top(\boldsymbol{\beta}_{j+1}-\boldsymbol{\beta}_j)+(f_{j+1}(\mathbf{p}_i)-f_j(\mathbf{p}_i))\geq\varepsilon, \qquad \forall i,j, \end{cases} \tag{3}$$

where $\mathcal{F}_0=\{f\in H^2(\mathcal{D}) : K\nabla f\cdot\nu=0 \text{ on } \partial\mathcal{D}\}$ is the space of functions in $H^2(\mathcal{D})$ satisfying the considered boundary conditions, and $\varepsilon>0$ is a tolerance parameter that guarantees the well-posedness of the inequality constraints. On the other hand, incorporating monotonicity constraints into the minimization problem significantly increases the computational complexity of the estimation procedure (see, e.g., Nocedal and Wright, 2006).

For this reason, we here propose to replace the constrained estimation problem (3) by an unconstrained estimation problem, using appropriate penalty functions. We do so by uplifting to the infinite-dimensional optimization problem here considered an approach classically followed for simpler optimization problems over scalar variables (see, e.g., Nocedal and Wright, 2006). Specifically, we propose to minimize the penalized loss functional

$$J_P(\boldsymbol{\beta},\boldsymbol{f})=\frac{1}{n}\sum_{i,j=1}^{n,r}\rho_{\alpha_j}(y_i-\mathbf{x}_i^\top\boldsymbol{\beta}_j-f_j(\mathbf{p}_i))+\sum_{j=1}^r\lambda_j P(f_j)+\gamma\sum_{i,j=1}^{n,r-1}\mathcal{C}_{i,j}, \tag{4}$$

where the penalty $\mathcal{C}_{i,j}=\max\{0,\varepsilon-\mathbf{x}_i^\top(\boldsymbol{\beta}_{j+1}-\boldsymbol{\beta}_j)-(f_{j+1}(\mathbf{p}_i)-f_j(\mathbf{p}_i))\}$ controls the crossing between the two consecutive quantiles of levels (α_j,α_{j+1}), at location $\mathbf{p}_i$, and the positive parameter γ controls the strength of such penalty. Notice that no contribution is added in the estimation functional J_P for those locations $\mathbf{p}_i$ and pair of quantile levels (α_j,α_{j+1}) for which the constraint is already fulfilled, as in this case the penalty does not become effective. Moreover, as shown in Nocedal and Wright (2006) for optimization problems over scalar variables, it can be shown that, in the limit for $\gamma\to\infty$, the solution to (4) converges to that of the constrained estimation problem (3).

3 Model estimation

The minimum of functional (4) is not available in closed analytical form, due to the non-quadratic terms ρ_α and $\sum_{i,j=1}^{n,r-1}\mathcal{C}_{i,j}$, as well as the complexity introduced by the PDE penalty. For this reason, in Section 3.1, we propose an iterative algorithm, that permits to derive an approximate solution to (4), by appropriate leveraging an extensive theoretical framework available for constrained optimization problems (see, e.g., Lange, 2016). Additionally, to deal with the PDE penalty and the possibly complex conformation of the spatial domain $\mathcal{D}$, we need to introduce an appropriate numerical discretization. In particular, in Section 3.2 we briefly outline a Finite Element approach that offers an efficient numerical discretization of the considered problem.

3.1 Maximization-Minimization algorithm

We here describe an iterative procedure, that approximates the minimum of the estimation functional (4), by solving a sequence of simpler optimization problems, each one enjoying a convenient characterization of its solution. Unfortunately, the Expectation-Maximization (EM) procedure described in Castiglione et al. (2025) for single-quantile estimation cannot be directly extended to this case, as it cannot accommodate the inclusion of the novel crossing penalty term. For this reason, here we develop an appropriate Maximization-Minimization (MM) strategy. Indeed, MM algorithms are particularly convenient when dealing with convex optimization problems characterized by multiple non-quadratic penalty terms and inequality constraints. We refer the reader to, e.g., Wu and Lange (2010), Sun et al. (2017) and Lange and Zhou (2022), for an overview of the MM approach and its relationship with the EM algorithm, and to Hunter and Lange (2000) for a first use of MM optimization in the context of quantile regression of univariate variables.

In brief, the general idea of MM algorithms is to replace the original optimization problem by an iterative minimization of a surrogate functional $\mathcal{G}$, which approximates J_P and it is easier to optimize. Denote by $\boldsymbol{\eta} = (\boldsymbol{\beta}, \boldsymbol{f})$ the completed vector of regression parameters. Then, the surrogate functional $\mathcal{G}$ should satisfy the two following conditions:

$$\begin{cases} \mathcal{G}(\hat{\boldsymbol{\eta}}^{(k-1)}) = J_P(\hat{\boldsymbol{\eta}}^{(k-1)}), & \text{tangency condition,} \\ \mathcal{G}(\boldsymbol{\eta}) \geq J_P(\boldsymbol{\eta}), \quad \forall \boldsymbol{\eta}, & \text{dominance condition,} \end{cases} \tag{5}$$

where $\hat{\boldsymbol{\eta}}^{(k-1)}$ is the estimate of $\boldsymbol{\eta}$ at the generic $(k-1)$-th step of the algorithm. The conditions in (5) ensure that, for $\hat{\boldsymbol{\eta}}^{(k)}$ such that

$$\mathcal{G}(\hat{\boldsymbol{\eta}}^{(k)}) \leq \mathcal{G}(\hat{\boldsymbol{\eta}}^{(k-1)}),$$

we have that

$$J_P(\hat{\boldsymbol{\eta}}^{(k)}) \leq J_P(\hat{\boldsymbol{\eta}}^{(k-1)}). \tag{6}$$

This condition, known as the descendent property, guarantees that, for a convex objective functional J_P and k large enough, the estimate $\hat{\boldsymbol{\eta}}^{(k)}$ converges to the global minimum of the objective functional J_P (see, e.g., Lange, 2016).

Appendices A.1 and A.2 define in details all the quantities involved in the proposed MM strategy for the minimization of (4). We here outline the general strategy of the MM algorithm, deriving some theoretical results. We introduce the weight matrix $W \in \mathbb{R}^{r(q+n)\times r(q+n)}$ and the vector $\boldsymbol{\xi} \in \mathbb{R}^{r(q+n)}$, that contain the absolute residuals of the model. Then, we define the matrix $T \in \mathbb{R}^{r(q+n)\times r(q+n)}$ and the vector $\mathbf{v} \in \mathbb{R}^{r(q+n)}$, that depend on the crossing quantile values. For the generic k-th step of the algorithm, we define the functional $\mathcal{G}$ as:

$$\begin{aligned} \mathcal{G}(\hat{\boldsymbol{\eta}}^{(k)}) = & \frac{1}{2}\, \hat{\boldsymbol{\eta}}^{(k-1)\,\top} W \hat{\boldsymbol{\eta}}^{(k-1)} - \hat{\boldsymbol{\eta}}^{(k-1)\,\top} (\boldsymbol{\xi} + \gamma \mathbf{v}) \\ & + \frac{\gamma}{2} \left(\hat{\boldsymbol{\eta}}^{(k-1)\,\top} T \hat{\boldsymbol{\eta}}^{(k-1)} \right) + \sum_{j=1}^{r} \lambda_j P(\hat{f}_j^{(k-1)}). \end{aligned} \tag{7}$$

The following proposition states that $\mathcal{G}$ in (7) is a surrogate for the functional J_P in equation (4).

Proposition 1. *The functional $\mathcal{G}$ in (7) satisfies the conditions (5) with respect to the functional J_P in equation (4). The functional $\mathcal{G}$ in (7) can thus be taken as a surrogate for J_P within an MM procedure.*

All proofs are deferred to the Appendix A.2. Owing to the descendent property stated in (6), at each iteration of the algorithm, we aim to solve

$$\hat{\boldsymbol{\eta}}^{(k)} = \underset{\boldsymbol{\eta}\,:\,\eta_1,..,\eta_r \in \mathbb{R}^q \times \mathcal{F}_0}{\arg\min}\ \mathcal{G}(\boldsymbol{\eta}), \tag{8}$$

which now reduces to a quadratic optimization problem. To characterize problem (8), we first define $\boldsymbol{g}$ as the collection $\{g_1, ..., g_r\}$, where $g_j = Lf_j - u_j \in L^2(\mathcal{D})$, for $j = 1, \ldots, r$, is the field representing the misfit of the solution f_j to the PDE. Moreover, we define the bilinear forms $R_0(\cdot,\cdot)$ and $R_1(\cdot,\cdot)$ as

$$R_0(\phi,\psi) = \int_{\mathcal{D}} \phi\psi, \quad R_1(\phi,\psi) = \int_{\mathcal{D}} [(K\nabla\phi)\cdot\nabla\psi + (\mathbf{b}\cdot\nabla\phi)\psi + c\phi\psi], \tag{9}$$

with $\phi, \psi \in H^1(\mathcal{D})$. Denote by $\mathbf{e}^{(r)}$ the unit vector of length r, and by $\otimes$ the Kronecker product. Set $\boldsymbol{\psi}_{rn} = \mathbf{e}^{(r)} \otimes (\psi(\mathbf{p}_1), \ldots, \psi(\mathbf{p}_n))$, and denote by $\hat{\boldsymbol{f}}_n^{(k)}$ the collection of the r nonparametric estimates at the spatial locations, i.e., $(\hat{f}_1^{(k)}(\mathbf{p}_1), \ldots, \hat{f}_1^{(k)}(\mathbf{p}_n), \ldots, \hat{f}_r^{(k)}(\mathbf{p}_n))$. Additionally, denote by W_β, W_f and by T_β, T_f the blocks of the weight matrix W and the quantile crossing matrix T associated with the parametric and nonparametric terms of the model; similarly, denote by $\boldsymbol{\xi}_\beta$, $\boldsymbol{\xi}_f$ and by $\mathbf{v}_\beta$, $\mathbf{v}_f$, the blocks of the vectors $\boldsymbol{\xi}$ and $\mathbf{v}$ associated with the parametric and nonparametric terms of the model. Finally, let $\hat{\boldsymbol{\eta}}^{(k)}$, with $\hat{\eta}_j^{(k)} \in \mathbb{R}^q \times \mathcal{F}_0$, for all $j = 1, \ldots, r$, denote the solution of (8), i.e., the minimum of $\mathcal{G}(\boldsymbol{\eta})$, and $\boldsymbol{g}$ the corresponding PDE misfit. The following proposition characterizes the solution to problem (8) as a system of first order equations.

Proposition 2. *The couple $(\hat{\boldsymbol{\eta}}^{(k)}, \hat{\boldsymbol{g}}^{(k)})$ is the solution of the following system of first order equations*

$$\begin{cases} W_\beta \hat{\boldsymbol{\beta}}^{(k)} - \boldsymbol{\xi}_\beta + \gamma(T_\beta \hat{\boldsymbol{\beta}}^{(k)} - \mathbf{v}_\beta) + C^\top \hat{\boldsymbol{f}}_n^{(k)} = 0 \\ \boldsymbol{\psi}_{rn}^\top W_f \hat{\boldsymbol{f}}_n^{(k)} - \boldsymbol{\psi}_{rn}^\top \boldsymbol{\xi}_f + \gamma(\boldsymbol{\psi}_{rn}^\top T_f \hat{\boldsymbol{f}}_n^{(k)} - \boldsymbol{\psi}_{rn}^\top \mathbf{v}_f) + \boldsymbol{\psi}_{rn}^\top C \hat{\boldsymbol{\beta}}^{(k)} + \sum_{j=1}^r \lambda_j R_1(\psi, \hat{g}_j^{(k)}) = 0 \\ R_1(\phi, \hat{f}_j^{(k)}) - R_0(\phi, \hat{g}_j^{(k)}) = 0 \quad \forall j = 1, \ldots, r. \end{cases}$$

The system of equations in Proposition 2 comprises coupled elliptic PDEs in variational form. This infinite-dimensional estimation problem lacks a closed-form analytical solution, necessitating a discretization into a suitable finite-dimensional space, as detailed in the following section.

3.2 Discretization and computational aspects

To discretize the infinite-dimensional estimation problem in Proposition 2 and the corresponding estimator $\hat{\boldsymbol{\eta}}^{(k)} = (\hat{\boldsymbol{\beta}}^{(k)}, \hat{\boldsymbol{f}}^{(k)})$, we employ the Finite Element Method, following the blueprint of Ramsay (2002), Sangalli et al. (2013) and Azzimonti et al. (2014), in the context of simpler smoothing and linear regression problems. This technique guarantees a sparse representation of the discretization matrices, ensuring high computational efficiency. Moreover, it enables us to accurately deal with spatial domains $\mathcal{D}$ having nontrivial geometries, as shown, e.g., in Sangalli et al. (2013).

Let $\mathcal{D}_h$ be a triangulation of the spatial domain $\mathcal{D}$, and $\mathbb{P}_s$ the space of polynomials of order s, where $s \in \mathbb{N}_0$. All the functions in the problem are expressed using a set of finite element basis $\{\psi \in C^0(\bar{\mathcal{D}}_h) : \psi|_T \in \mathbb{P}_s, \forall T \in \mathcal{D}_h\}$, consisting of globally continuous functions over the triangulated domain $\mathcal{D}_h$, that are polynomials of order s once restricted to any element of the triangulation. We here consider linear basis functions and, in this case, the number of basis functions coincides with N, the number of vertices of $\mathcal{D}_h$. Moreover, let $\boldsymbol{\beta}_h$, $\mathbf{f}_h$, and $\mathbf{g}_h$ represent the finite element approximations over $\mathcal{D}_h$ of $\hat{\boldsymbol{\beta}}^{(k)}$, $\hat{\boldsymbol{f}}^{(k)}$ and $\hat{\boldsymbol{g}}^{(k)}$, respectively. We define $\Psi := \{\psi_j(\mathbf{p}_i)\}_{i,j=1}^{n,N}$ as the matrix of basis function evaluations at the spatial locations $\{\mathbf{p}_i\}_{i=1}^n$, $\boldsymbol{\psi} = (\psi_1, \ldots, \psi_N)$, and $R_0 := \int_{\mathcal{D}} \boldsymbol{\psi}\boldsymbol{\psi}^\top$ and $R_1 := \int_{\mathcal{D}} [\nabla\boldsymbol{\psi}^\top K \nabla\boldsymbol{\psi} + \nabla\boldsymbol{\psi}^\top \mathbf{b}\boldsymbol{\psi}^\top + c\boldsymbol{\psi}\boldsymbol{\psi}^\top]$, as the mass and stiffness matrices, that are the discretizations of the bilinear forms in (9); see, e.g., Quarteroni et al. (2014). Moreover, we set

$$\Lambda = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_r \end{bmatrix}, \quad \tilde{R}_0 = \Lambda \otimes R_0, \quad \tilde{R}_1 = \Lambda \otimes R_1, \quad \tilde{\Psi} = I_r \otimes \Psi, \quad \tilde{X} = I_r \otimes X,$$

where I_r is the identity matrix of dimension r, and X is the design matrix of the model whose i-th row is $\mathbf{x}_i$. With these definitions, the discrete counterpart of the system in Proposition 2 can be written as

$$\begin{bmatrix} W_\beta + \gamma T_\beta & C^\top \tilde{\Psi} & 0 \\ \tilde{\Psi}^\top C & \tilde{\Psi}^\top (W_f + \gamma T_f) \tilde{\Psi} & \tilde{R}_1^\top \\ 0 & \tilde{R}_1 & -\tilde{R}_0 \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}_h \\ \mathbf{f}_h \\ \mathbf{g}_h \end{bmatrix} = \begin{bmatrix} \boldsymbol{\xi}_\beta + \gamma \boldsymbol{v}_\beta \\ \tilde{\Psi}^\top (\boldsymbol{\xi}_f + \gamma \boldsymbol{v}_f) \\ 0 \end{bmatrix}. \tag{10}$$

By appropriately manipulating the matrix system in (10), we can leverage the Sherman-Morrison-Woodbury decomposition (see Woodbury, 1950) to ensure an efficient solution to Equation (10). As detailed in Arnone et al. (2023), which explored such approach for a simpler spatial linear regression problem, this technique is particularly advantageous in the semiparametric setting, allowing fast computations. For full details on these computational aspects, we refer to Appendix A.3.

4 Simulation study

We now present three simulation studies that compare the performance of the proposed Multiple Quantile Spatial Regression with Partial Differential Equation regularization

(MQSR-PDE), implemented in the `fdaPDE` library (Palummo et al., 2025), to alternative methods, in various simulation settings. Specifically, in Section 4.1, we examine data generated from a heteroscedastic Gaussian model; in Section 4.2, we consider data generated from a heteroscedastic Skew-T process; finally, in Section 4.3, we consider the same data generation setting of Section 4.2 but we also include a spatially varying covariate.

The goal of the simulation studies is to conduct an in-depth analysis of the data distribution, by fitting a dense sequence of quantiles, with particular focus on the right tail. For this reason, we choose a fine grid of quantile levels between the 90% and 99%. Specifically, we explore the following sequence of quantile levels:

$$\alpha = (1\%, 2\%, 3\%, 5\%, 10\%, 25\%, 50\%, 75\%, 90\%, 91\%, 92\%, \dots, 97\%, , 98\%, 99\%).$$

Denoting by Q_j the true α_j-th quantile field and by $\hat{Q}_j$ the corresponding estimate, the competing methods are compared in terms of Root Mean Squared Error (RMSE),

$$RMSE(\hat{Q}_j) = \sqrt{\frac{1}{N}\sum_{k=1}^{N}(Q_j(\mathbf{p}_k) - \hat{Q}_j(\mathbf{p}_k))^2}, \quad \forall j = 1, \dots, r.$$

calculated over the 50 different replica of the experiment, on a regular lattice of $N = 2500$ points.

4.1 Simulation 1: heteroscedastic Gaussian process

We generate data according to the following heteroscedastic Gaussian model

$$Y_i \sim \mathcal{N}(\mu(\mathbf{p}_i), \sigma^2(\mathbf{p}_i)), \quad i = 1, \dots, n, \tag{11}$$

where $\mu(\mathbf{p})$ and $\sigma^2(\mathbf{p})$ are Gaussian random fields, generated using the `grf` function of the `R` package `geoR`. In particular, for both μ and σ, we specify the anisotropic covariance function, through a Matérn model, setting the anisotropy intensity to 5 and the angle to $\pi/4$. We consider $n = 169$ spatial locations, $\mathbf{p}_1, \dots, \mathbf{p}_n$, randomly scattered in the unit square domain $[0, 1] \times [0, 1]$. Figure 3 shows the sampled data, alongside with the mean and standard deviation fields.

4.1.1 Comparison among PDE-regularized quantile models

In this subsection, we compare various PDE-regularized quantile regression methods, all implemented in the `fdaPDE` library (Palummo et al., 2025). Specifically, we evaluate the performance of the proposed MQSR-PDE model for simultaneous quantile estimation, as detailed in Section 3, against the single-quantile model presented by Castiglione et al. (2025), denoted by QSR-PDE. Additionally, we also consider the rearranged version of the single-quantile estimates, where rearrangement is performed using the post-processing procedure discussed by Chernozhukov et al. (2010), in order to ensure non-crossing; the latter method is denoted by QSR-PDE-R. For all

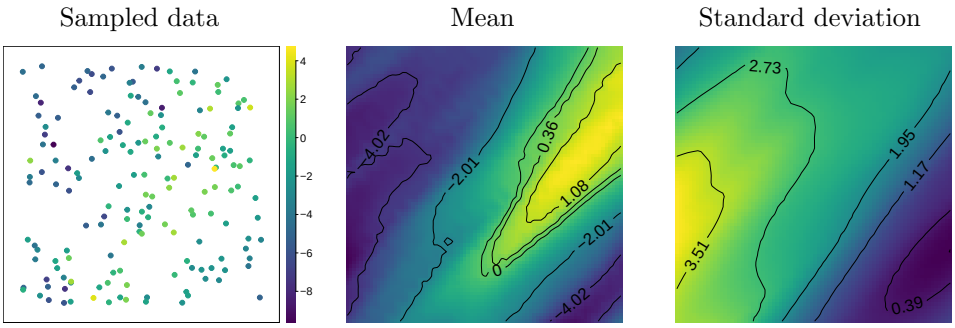


Fig. 3: Simulation 1 (Section 4.1): heteroscedastic Gaussian process. Example of the sampled data (left panel), mean (central panel), and standard deviation (right panel) of the Gaussian process used for the data generation.

three models, we incorporate a purely diffusive differential regularization term, with unknown hyperparameters, to account for the strong anisotropy present in the data. These unknown hyperparameters, that characterize the intensity and direction of the anisotropy, are estimated using the Parameter Cascading algorithm, as discussed in Section 2.1. Finally, to highlight the advantages of incorporating anisotropic diffusion over a simpler isotropic estimate, we also include a comparison with the multiple quantile model that employs in the regularizing term an isotropic diffusion; the latter model is denoted by MQSR-ISO. All models are based on a regular triangulation of the square domain, having 625 discretization nodes; the smoothing parameter λ is chosen via Generalized Cross-Validation (GCV; see, e.g., Nychka et al., 1995; Yuan, 2006; Li et al., 2007).

Figure 4 shows the RMSE of the compared methods, for a subset of the explored quantiles. Focusing first on the comparison between the anistropic and isotropic multiple quantile regression estimators, MQSR-PDE and MQSR-ISO, we observe that MQSR-PDE achieves significantly lower error than MQSR-ISO, both over the central quantiles and in the tails of the distribution. This result underscores the beneficial effect of incorporating a diffusive differential regularization term with unknown hyperparameters, to capture the anisotropy in the data and obtain more accurate estimates, across the entire range of quantiles. Turning to the comparison among the anisotropic estimators, multiple (MQSR-PDE), rearranged (QSR-PDE-R) and single (QSR-PDE), the top panel of the figure demonstrates that, for quantiles ranging from the 5-th to 75-th percentile, there are no significant differences. This is expected: indeed, since the quantile levels up to the 75-th percentile are spaced further apart, the crossing penalties do not become effective. As commented in Section 2.2, this lack of significant differences of the estimates across quantiles levels that are further apart is a positive feature of the proposed model: indeed, whenever crossing does not naturally occur, the inclusion of non-crossing penalties does not significantly influence the estimation problem, and the obtained estimates have the same accuracy of the single-quantile estimates, which, as shown by Castiglione et al. (2025), possess desirable asymptotic properties and perform well in finite samples. Focusing on the bottom panel of

Figure 4, we observe that the advantage of the non-crossing techniques, MQSR-PDE and QSR-PDE-R, becomes significant in the right tail of the distribution, where the quantile maps are estimated across a dense grid of quantile levels, and consequently quantile crossing is more likely to occur. Specifically, we can see as the proposed MQSR-PDE model outperforms both the single-quantile estimates, QSR-PDE, as well as its rearranged version, QSR-PDE-R, attaining significant lower values of RMSE.

Overall, the MQSR-PDE methodology, which integrates anisotropy information with the simultaneous estimation of multiple quantile levels, yields the most accurate estimates. This finding is further supported by the results of univariate pairwise Wilcoxon tests, conducted for each quantile in the right tail of the distribution, which are presented in the bottom panel of Figure 4.

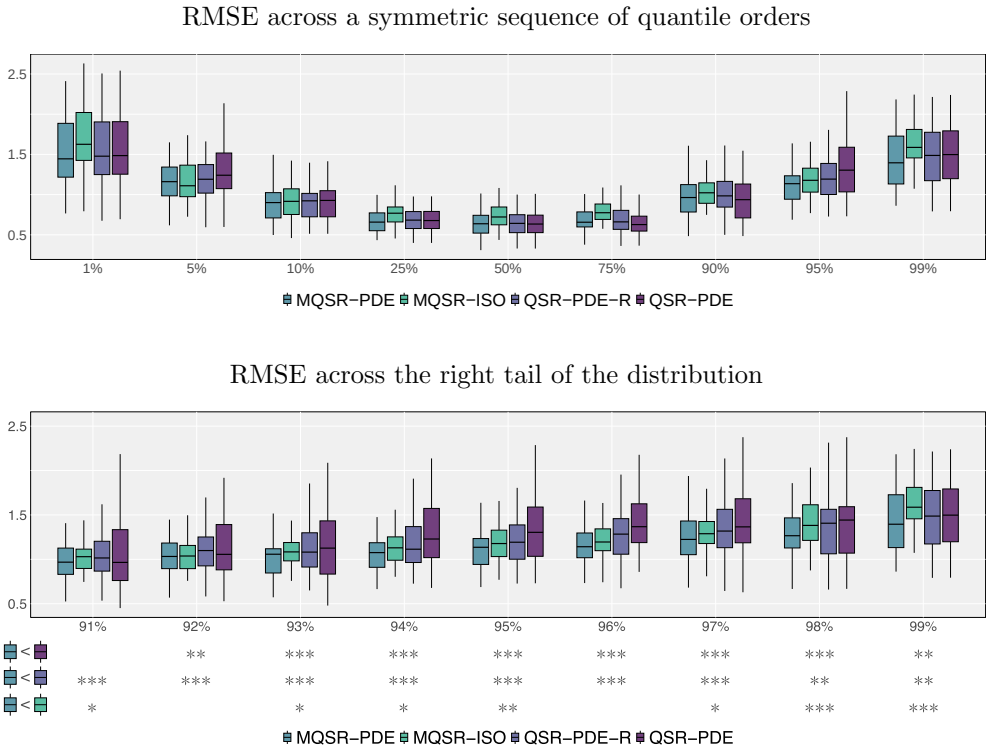


Fig. 4: Simulation 1 (Section 4.1): heteroscedastic Gaussian process. Boxplots of RMSE obtained by different PDE-regularized quantile regression methods: the proposed multiple quantile regression model with anisotropic regularizing term (MQSR-PDE); the multiple quantile regression model with isotropic regularizing term (MQSR-ISO); the single-quantile regression model with anisotropic regularizing term by Castiglione et al. (2025) (QSR-PDE); the latter model combined with the post-processing procedure discussed by Chernozhukov et al. (2010) to ensure non-crossing of the quantiles (QSR-PDE-R). Top panel: RMSE for a symmetric sequence of quantile levels. Bottom panel: RMSE for quantile levels from 91% to 99%. The bottom panel also reports, for each quantile level of the right tail of the distribution, the results of pairwise Wilcoxon tests verifying that MQSR-PDE estimates have significant lower RMSE than MQSR-ISO estimates, QSR-PDE-R estimates, and of QSR-PDE estimates. Legend: . : $0.05 <$ p-value < 0.1; * : $0.01 <$ p-value < 0.05; ** : $0.001 <$ p-value < 0.01; *** : p-value < 0.001.

Alongside the accuracy comparison, we also measured the computational times took by MQSR-PDE, MQSR-ISO, QSR-PDE-R and QSR-PDE models. All methods are executed on an Intel(R) Core(TM) Ultra 9 185H processor. We obtain an average time of 267.62 seconds (with a standard deviation 7.06 seconds) for the QSR-PDE method, while the MQSR-PDE method has a mean overhead to ensure

non-crossing estimates of 11.83%, and the rearrangement method introduces almost no additional computational cost. Finally, when comparing the two multiple methodologies MQSR-PDE and MQSR-ISO, we note that the estimation of the anisotropy tensor in MQSR-PDE results in a mean overhead of 11.71% compared to the isotropic model. In conclusion, these results highlight the advantage of using the proposed MQSR-PDE model for simultaneous quantile estimation in terms of accuracy, at a modest increment of the computational cost.

4.1.2 Comparison with state-of-the-art non-crossing quantile methods

In this subsection, we compare the proposed MQSR-PDE model with alternative methods from the literature. Specifically, we compare with the quantile models based on thin-plate-spline and on SOAP film smoothing, proposed by Fasiolo et al. (2021a), and rearranged following the strategy outlined in Chernozhukov et al. (2010); these methods, denoted by the acronyms TPS-R and SOAP-R, are implemented in the `R` package `qgam` (see Fasiolo et al., 2021b). Additionally, we consider the Quantile Regression Coefficients Modeling (QRCM) model proposed by Frumento et al. (2021), which employs non-crossing constraints; this technique is implemented in the `R` package `qrcm` (see Frumento and Bottai, 2016; Sottile and Frumento, 2023). We also considered the Monotone Composite Quantile Regression Neural Network model proposed in Cannon (2018). However, this method proved to be unstable, in all the simulation settings explored, probably due to the scarcity of the sampled data. Consequently, we decided to exclude this methodology from further comparisons. We could not compare with the other multiple quantile regression methodologies mentioned in Section 1 due to lack publicly available code.

The proposed MQSR-PDE is run as detailed in Section 4.1.1. For TPS-R we use 150 basis functions, while for SOAP-R we use 75 basis, as this is the upper limit for which the model runs without halting the execution. Likewise for MQSR-PDE, also the smoothing parameters of TPS-R and SOAP-R are selected using Generalized Cross-Validation.Lastly, QRCM utilizes 5 basis functions, since employing a larger number of basis functions results in unstable solutions.

Figure 5 shows true quantile fields for $\alpha = 50\%$, 95% and 99%, and the mean estimated quantile fields, provided by the competing methods. Qualitatively, the proposed MQSR-PDE model appears to provide the best estimates, with an advantage over the competing methods that is particularly appreciable at the 99% level, where the other methods struggle to capture the pattern of the true quantile. Figure 6 shows the RMSE values obtained by the competing methods, on a subset of the estimated quantiles. We observe that the methods yield more similar results in the central portion of the distribution, while more pronounced differences emerge at extreme quantile levels. Specifically, when examining central quantiles, such as the median, we find that QRCM, that enforces non-crossing constraints, provides less accurate estimates compared to the other methods. The performances of this method improve instead when moving towards tail quantiles. On the contrary, methods based on rearrangement, such as SOAP-R and TPS-R, which do not directly enforce non-crossing constraints,

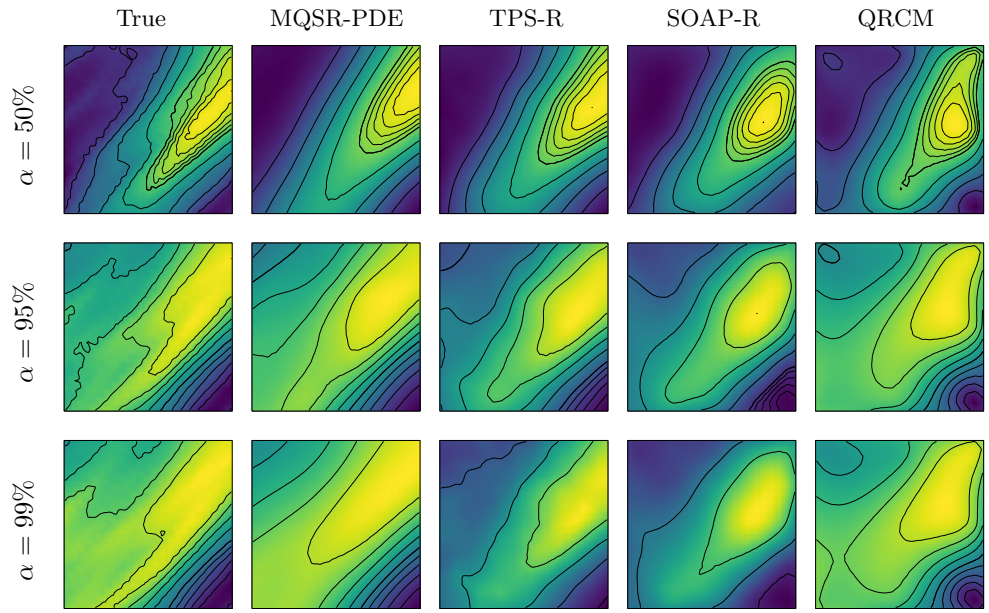


Fig. 5: Simulation 1 (Section 4.1): heteroscedastic Gaussian process. Estimated quantile maps for some of the considered quantile levels. The first column presents the true quantile field, and the subsequent columns display the mean estimates for each of the competing methods: the proposed multiple quantile spatial regression model (MQSR-PDE); the quantile regression based on a thin-plate-spline with rearrangement (TPS-R); the quantile regression based on a Soap film smoothing with rearrangement (SOAP-R); the Quantile Regression Coefficient Modeling (QRCM).

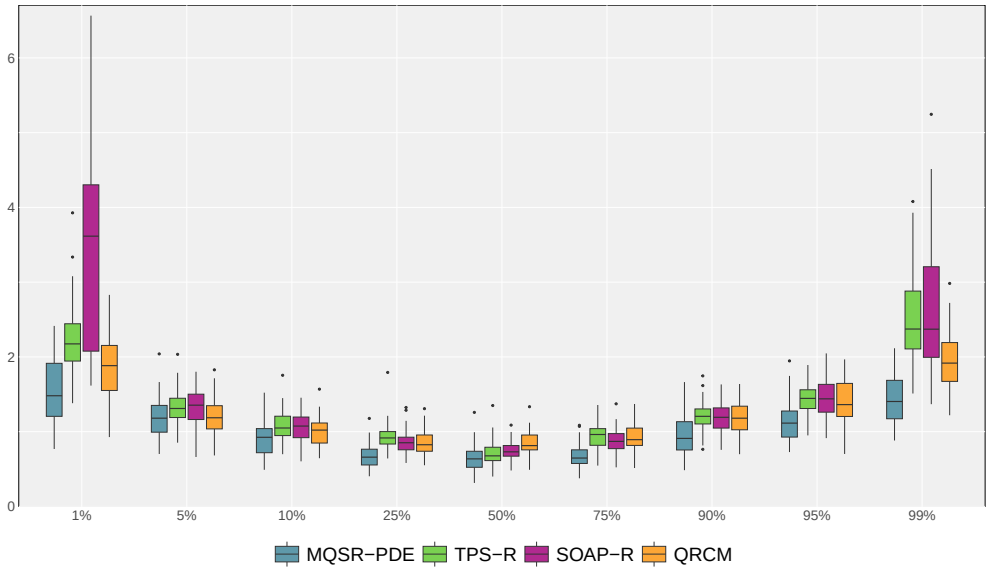
struggle to provide accurate estimation of tail quantiles, leading to higher RMSE values for lower and upper quantiles. This is likely due to the fact that crossing becomes more pronounced in the tails of the distribution, where a simple rearrangement is insufficient. In contrast, the proposed MQSR-PDE provides accurate estimates both at central quantiles and at tail quantiles, achieving significantly lower RMSE values across all the considered quantile levels.

Together with the accuracy results, we also evaluate the computational time of the different methods, with the same computer specifications detailed in Section 4.1.1. The results are presented in the table of Figure 6. In particular, MQSR-PDE, TPS-R, and SOAP-R exhibit comparable computational times, as they all rely on smoothing techniques. In contrast, QRCM stands out as significantly faster, but this comes at the cost of substantially lower accuracy across all quantile levels.

4.2 Simulation 2: heteroscedastic Skew-T process

In this simulation, we generate data using a heteroscedastic Skew-T process, to assess the performance of our proposed model under conditions characterized by skewness

RMSE



Computational times

Method	**Mean (sec)**	**Std dev (sec)**
MQSR-PDE	299.25	6.91
TPS-R	137.54	19.20
SOAP-R	258.60	11.02
QRCM	12.59	17.08

Fig. 6: Simulation 1 (Section 4.1): heteroscedastic Gaussian process. Top: boxplots of the RMSE of the estimates provided by competing methods, that are the same as in Figure 5. Bottom: mean and standard deviation of the computing times for the different methods on a Intel(R) Core(TM) Ultra 9 185H processor.

and heavy tails. The data are produced according to the following model

$$Y_i \sim St(\xi(\mathbf{p}_i), \Omega^2(\mathbf{p}_i), \delta, \nu), \quad i = 1, \dots, n, \tag{12}$$

where the location parameter $\xi(\mathbf{p})$ and the scale parameter $\Omega^2(\mathbf{p})$ are Gaussian random fields, both employing the same anisotropic covariance function used in the Gaussian simulation setting described in Section 4.1. The shape parameter δ and the degrees of freedom ν are set to $\delta = 4$ and $\nu = 10$, respectively, to produce asymmetric data exhibiting a heavy right tail. We consider $n = 153$ spatial locations, $\mathbf{p}_1, \dots, \mathbf{p}_n$, in the unit square domain $[0, 1] \times [0, 1]$, sampled using a cluster Matérn process in order to ensure a sparse sampling design. An example of the sampled data is displayed in Figure 7, alongside with the location and scale fields of the Skew-T process.

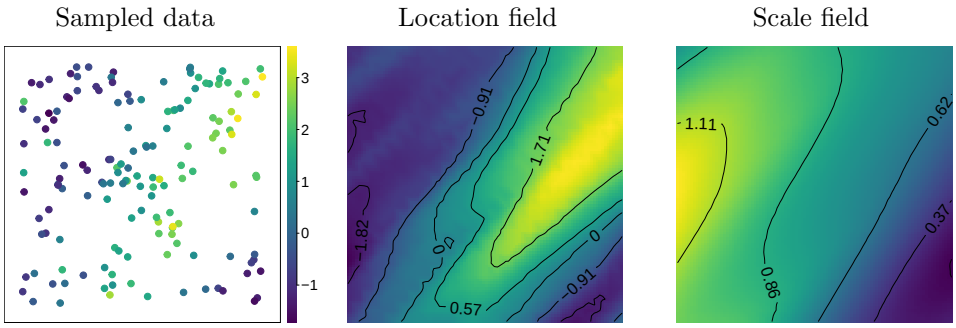


Fig. 7: Simulation 2 (Section 4.2): heteroscedastic Skew-T process. Example of the sampled data (left panel), location field (central panel), and scale field (right panel) of the Skew-T process used for the data generation.

We run the competing methods as in Section 4.1, with the sole difference that SOAP-R now use 50 basis functions, as this is the upper limit for which the model runs without halting the execution. Figure 8 displays the RMSE values obtained from the competing methods across a symmetric grid of estimated quantiles. Notably, at the extreme low quantile level $\alpha = 1\%$, all methods utilizing the rearrangement technique fail to accurately capture the true quantile field, as indicated by the high RMSE values. In contrast, the MQSR-PDE model consistently achieves the most accurate estimates, with the lowest RMSE values observed across all quantile levels. We refer to Appendix A.4 for a visualization of the quantile estimates and an additional comparison of the proposed MQSR-PDE methodology with its single counterpart (QSR-PDE) and its rearranged version (QSR-PDE-R).

4.3 Simulation 3: heteroscedastic Skew-T process with covariate

In this simulation study, we consider the same data generation process described in Section 4.2, but add a spatially varying covariate. Specifically, data are generated as follows:

$$Y_i \sim x(\mathbf{p}_i)\beta + St(\xi(\mathbf{p}_i), \Omega^2(\mathbf{p}_i), \delta, \nu), \quad i = 1, \dots, n, \tag{13}$$

where $\beta = 1$ and the covariate $x(\mathbf{p}_i)$ is generated using a Gaussian Random Field with exponential covariance model, using the R package `grf`. An example of the resulting sampled data, together with the covariate field, is reported in Figure 9.

Figure 10 presents the RMSE of the semiparametric field estimates for some of the computed quantiles, demonstrating that the proposed method consistently outperforms the alternatives across all examined quantile levels.

It should be pointed out that, as mentioned in the Introduction, rearrangement procedure makes it impossible to disentangle the contributions of the parametric and nonparametric terms. Instead, the proposed MQSR-PDE offers interpretable estimates of the parametric and nonparametric estimates, while respecting the non-crossing of

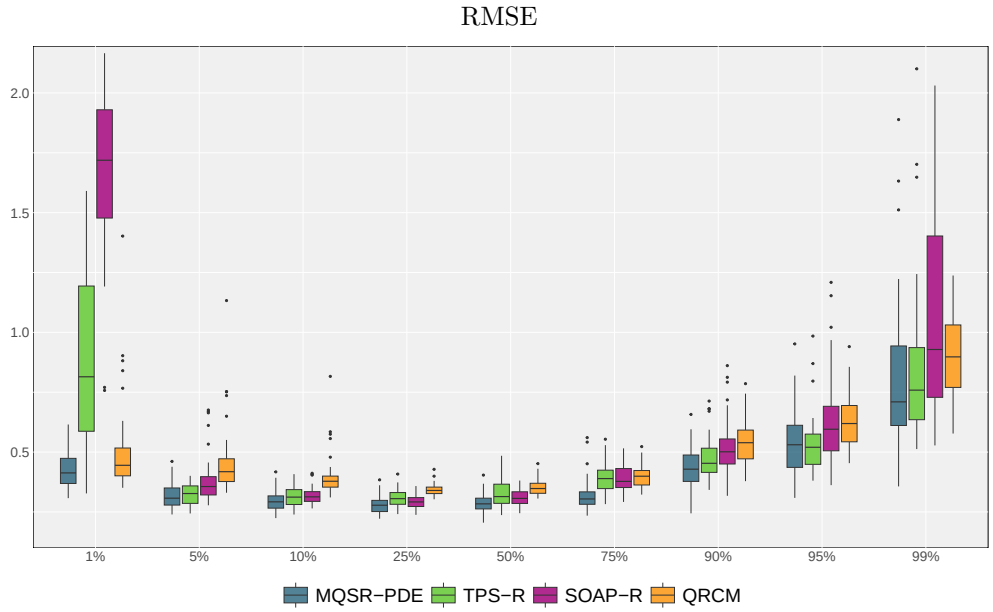


Fig. 8: Simulation 2 (Section 4.2): heteroscedastic Skew-T process. Boxplots of the RMSE for the estimates provided by the competing methods, including the proposed MQSR-PDE, the rearranged version of quantile thin-plate-spline smoothing (TPS-R), the rearranged version of quantile SOAP film smoothing (SOAP-R), and the Quantile Regression Coefficients Modeling (QRCM).

the quantiles. To assess the quality of the estimates of the parametric and nonparametric terms provided by the proposed MQSR-PDE, in Figure 11 we compare the estimates of these two terms, with those that can be obtained by the single quantile models, namely TPS and SOAP (thus not controlling the quantile crossing), and those obtained by QRCM. The boxplots of the estimated β coefficients, in the top panel of Figure 11, shows that the proposed MQSR-PDE model returns estimates of the parametric term comparable with those that can be obtained by single quantile models. An increased variability is observed for the highest quantile levels for all methods. The boxplots of the RMSE for the nonparametric component, shown in the bottom panel of the same figure, highlight instead that proposed MQSR-PDE model provides estimates of the nonparametric term that outperform those obtained by single quantile models.

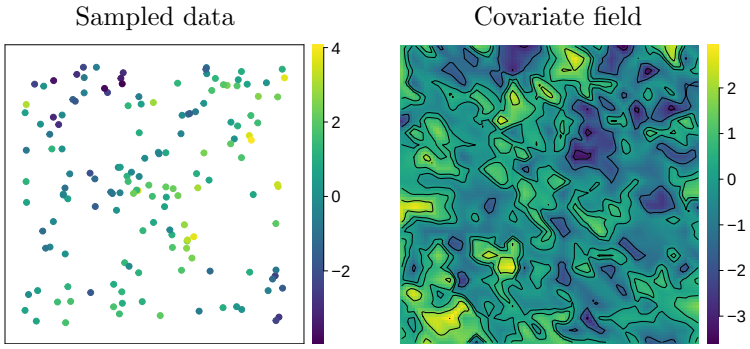


Fig. 9: Simulation 3 (Section 4.3): heteroscedastic Skew-T process with covariate. Example of the sampled data (left panel) and covariate field (right panel).

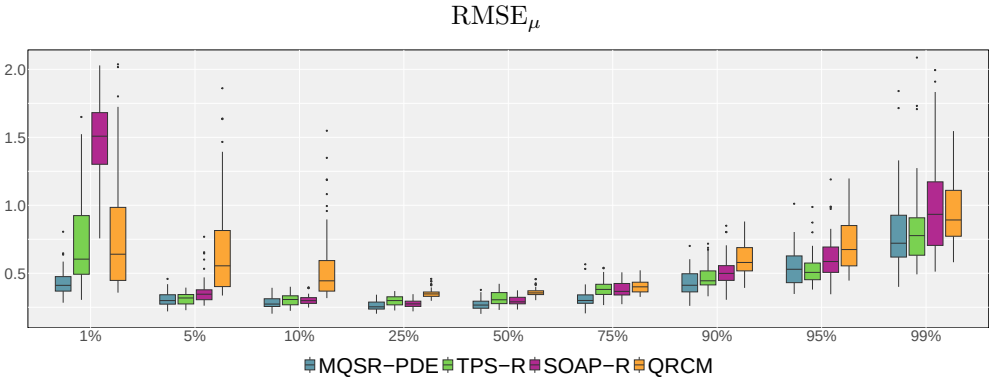


Fig. 10: Simulation 3 (Section 4.3): heteroscedastic Skew-T process with covariate. Boxplots of the RMSE for the semiparametric term provided by the proposed MQSR-PDE, the rearranged version of quantile thin-plate-spline smoothing (TPS-R), the rearranged version of quantile SOAP film smoothing (SOAP-R), and the Quantile Regression Coefficients Modeling (QRCM).

5 NO_2 spatial concentration in Lombardy region

We now apply the proposed MQSR-PDE methodology to the analysis of the NO_2 concentration in the Lombardy region, in northern Italy, exploiting the semiparametric structure of the proposed model, in order to also explore the role of some leading geographical and anthropogenic factors. Data refer to 11^{th} January 2019, and are publicly available from Regione Lombardia (Open Data, 2024).

As commented in the Introduction, it is known that exposure to high concentrations of NO_2 has harmful effects on the population, leading to the development of severe diseases and, ultimately, to premature death. The purpose of this study is to estimate the quantiles of the spatial distribution of NO_2, targeting high probability levels, with

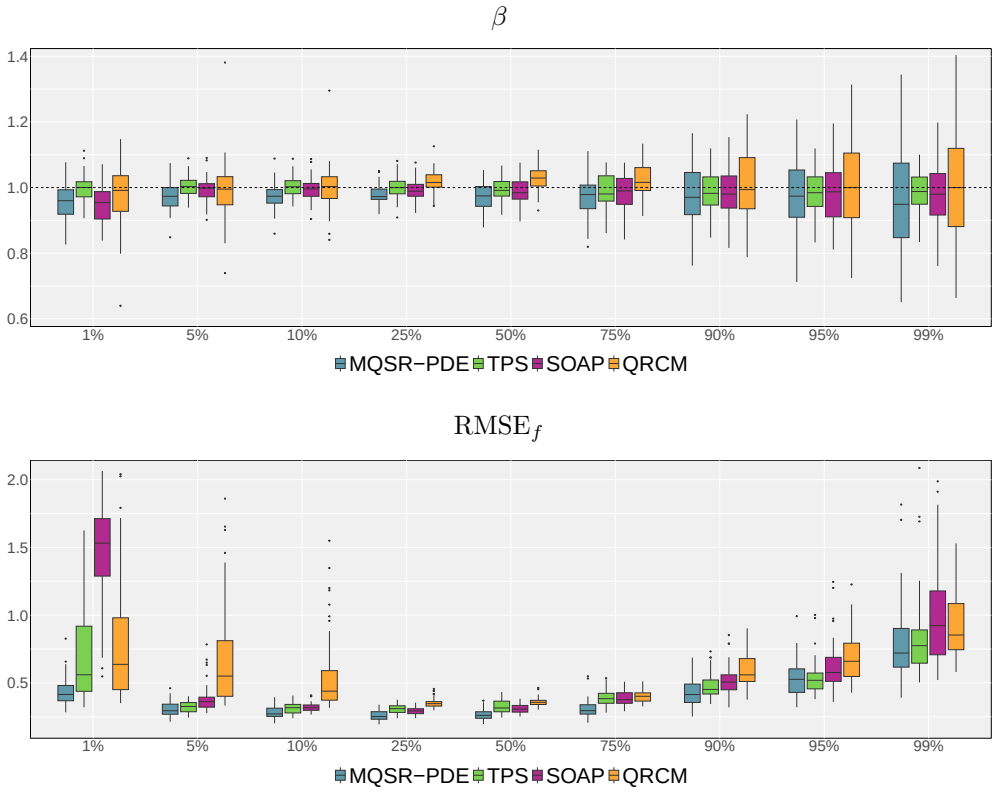


Fig. 11: Simulation 3 (Section 4.3): heteroscedastic Skew-T process with covariate. Boxplots of the estimated regression coefficient (top panel), and RMSE for the non-parametric term (bottom panel), provided by the proposed MQSR-PDE, the quantile thin-plate-spline smoothing (TPS), the quantile SOAP film smoothing (SOAP-R), and the Quantile Regression Coefficients Modeling (QRCM).

the goal of accurately reconstructing the right tail of the distribution. The proposed model is well-suited for this task, enabling the simultaneous estimation of a sequence of closely spaced quantiles, while ensuring monotonicity.

The Lombardy region, situated in the Po Valley, is one of the most critical areas for air quality in Europe, due to both geographical and anthropogenic factors. The region's morphology is particularly unfavorable: the Po Valley is landlocked, since enclosed by the Alps to the North and West, and the Apennines to the South, with a unique opening to the Adriatic Sea on the East. This creates limited atmospheric circulation, which worsens air quality. During winter, this issue is intensified by thermal inversion, a phenomenon that occurs in areas with low air circulation, when the ground is cold. This results in a stable separation of cooler air near the ground and warmer air above, trapping pollutants and allowing them to accumulate.

Since air circulation is a key factor affecting NO_2 concentration, we leverage the ability of the proposed model to incorporate knowledge on the physics of the phenomenon, by including air circulation in the physics-informed penalty detailed in Section 2.1. Wind data are gathered from the 119 ARPA meteorological stations, scattered throughout the whole territory. The correspondent wind field, registered on the same day as the NO_2 records, is depicted in Figure 1. In particular, we consider a physics-informed penalty with a transport term, having as transport vector $\mathbf{b}$ the wind stream in Figure 1; we combine this transport term with an isotropic diffusion term, which instead models the natural isotropic diffusion of the pollutant. We thus consider the following penalties

$$P(f_j) = \int_{\mathcal{D}} \big(-\nabla \cdot (I \nabla f_j) + \zeta \, \mathbf{b} \cdot \nabla f_j \big)^2 \quad \forall j = 1, \dots, r;$$

where the unknown hyperparameter ζ controls the relative intensities of the diffusion and transport terms, and is estimated using the parameter cascading technique, as commented in Section 2.

Additionally, to incorporate the morphology of the territory, we include, as a covariate in the model, the logarithm of the altitude at each spatial location. Altitude data, sourced from the Digital Terrein Model (Regione Lombardia, 2024), are illustrated in Figure 12. Anthropogenic activities play a leading role on NO_2 concentration. In this respect, it is worth mentioning that Lombardy serves as Italy's primary industrial and production hub, hosting numerous factories that contribute significantly to emissions, thereby increasing pollutant concentrations in the air. Additionally, the region is characterized by large urban centers, such as the metropolitan city of Milano, where high levels of heating and traffic are significant contributors to atmospheric pollution. To account for these influences, we thus incorporate, as second covariate in the model, the population density, using data from the Istituto Nazionale di Statistica (ISTAT), sourced from Regione Lombardia (2024). Population density is available for all 1506 municipalities in Lombardy, and is included in the model after a linear smoothing of the square root of the data, performed with the `fdaPDE` library, to enable a smooth reconstruction of the field. The spatial pattern of this covariate, displayed in Figure 12, highlights the very high population densities in the cities of Milano, Monza, Bergamo and Brescia.

We then estimate a dense sequence of 27 quantiles, across the whole NO_2 support, concentrating most of them in the right tail of the distribution. This enables us to accurately reconstruct the spatial patterns associated with the highest levels of NO_2, offering valuable insights into the risks faced by residents. The estimated values of the regression coefficients are negative for altitude, and positive for population density, at all quantile levels. This confirms that, across the entire distribution of NO_2 values, altitude has a beneficial effect on air quality, contributing to reduced NO_2 levels, whilst, higher population density is associated with higher NO_2 levels. Figure 13 shows the estimated semiparametric fields for some of the considered quantiles, namely $\alpha = 0.50, 0.90, 0.99$. We can observe that the peak of the NO_2 concentration occurs in the metropolitan area of Milan across all quantile levels. This is expected,

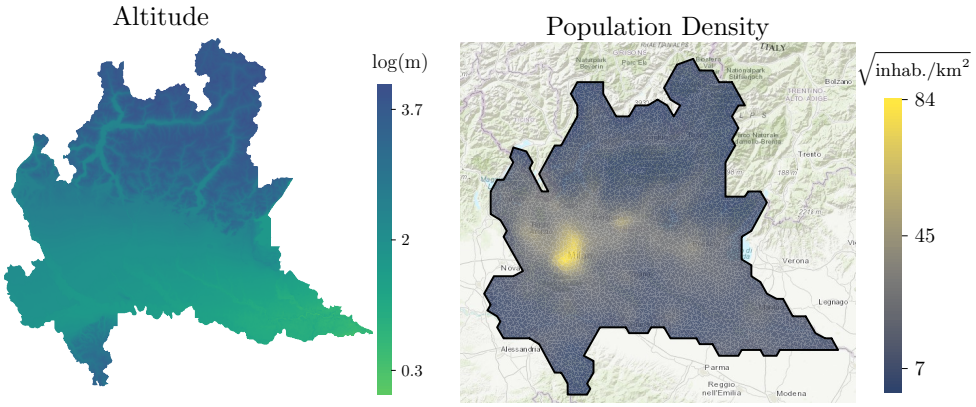


Fig. 12: Maps of altitude (left) and population density (right) over the Lombardy region.

being Milano the largest urban center in the region, with the highest population density and correspondingly highly concentrations of human activities. Analyzing more in detail the extreme quantiles, we note that the 90% and 99% estimated quantiles display more strongly localized features compared with the median, the 50% quantile, showing local outliers in correspondence to the biggest cities of the region, and to the industrial areas in the Po Valley.

Furthermore, to emphasize the importance of incorporating wind stream information as a transport term in the PDE, we compare the purely isotropic diffusion model with the proposed model, which accounts for wind dynamics. We conduct a pairwise Wilcoxon test to evaluate differences in terms of mean cross-validation errors. The test confirms that the proposed model, which includes wind dynamics, achieves a significantly lower CV error (p-value < 0.05). This confirms that the inclusion of prior physical information, given in this application by the wind stream, is particularly beneficial, especially in scenarios with limited sample size, as also shown, e.g., by Tomasetto et al. (2024) and Castiglione et al. (2025).

Starting from the monotone quantile estimates, we can obtain a fully nonparametric reconstruction of the NO_2 probability density function, at any spatial location. The algorithm used to obtain the probability density function, starting from the estimated quantiles, is detailed in Appendix A.5. This reconstruction of the probability density enables us, for example, to compare the distribution function of NO_2 in different cities, exploring how different morphological and urban characteristics may affect the distribution of this pollutant. The left panel of Figure 14, for instance, shows the probability density functions for the cities of Milano, Cremona and Sondrio, on the 11^{th} January 2019. It can be noticed that the density supports and modes shift towards higher values, moving from less polluted areas, as the mountain town of Sondrio, to more populated ones, like the metropolitan city of Milano. Specifically, the density in Milano is marked by a heavy right tail, highlighting the harmful situation faced by the citizens.

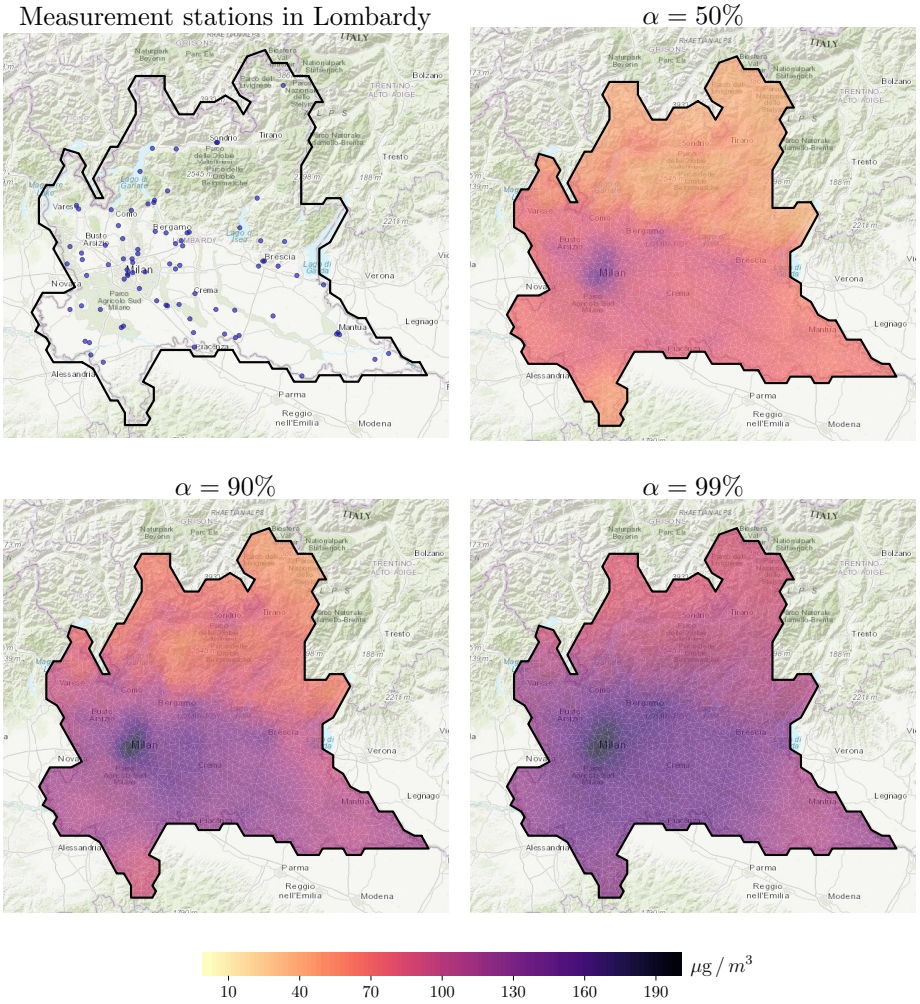


Fig. 13: Estimated quantile maps of the NO_2 concentration over Lombardy region, at quantile levels $\alpha = 0.50, 0.90, 0.99$.

Besides analyzing NO_2 concentrations on 11^{th} January 2019, we also examined data from 27^{th} December 2018, to assess how the pollutant's probability distribution varies between a regular working day and the Christmas holiday period. The right panel of Figure 14, shows the probability density functions for the cities of Milano, Cremona and Sondrio during this bank holiday. The figure highlights a reversed situation with respect that observed in the working day of 11^{th} January. In particular, in the cities of Milano and Cremona, the distribution of the NO_2 levels in the 27^{th} December (right panel) is shifted towards significantly lower values than those of the 11^{th} January (left panel). On the contrary, the mountain town of Sondrio experiences a much worse condition during the Christmas holiday (right panel) than on a standard working

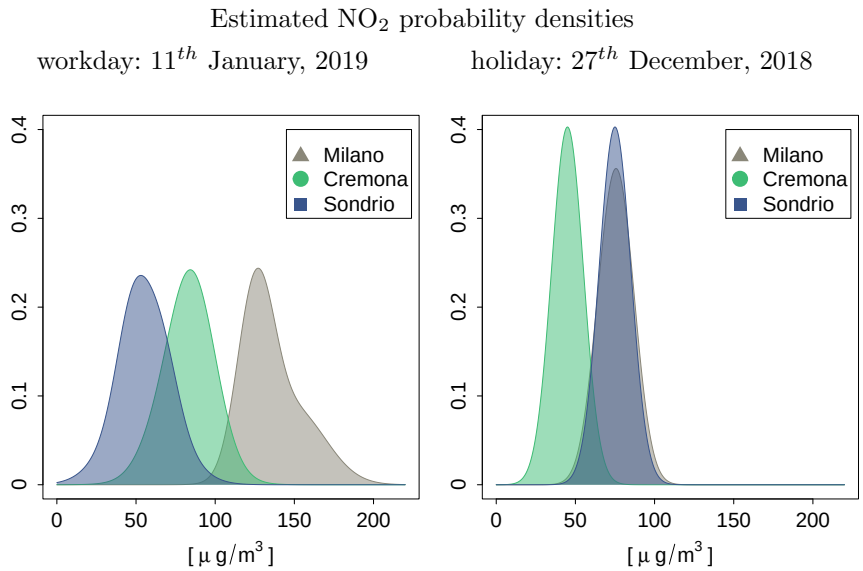


Fig. 14: Estimated NO_2 distributions in Milano (light gray), Cremona (green), and Sondrio (blue), obtained from the monotone quantile fields, on a workday (left panel) and on holiday (right panel).

date (left panel); this deterioration is likely due to the influx of the population to mountain areas during the Christmas period, significantly increasing the mode of NO_2 distribution in this town.

Using monotone quantile estimates, we can produce maps indicating the probability of exceeding certain critical values of the pollutant concentration. In fact, to mitigate the impact of NO_2 on public health, several institutions have established maximum thresholds that should not be exceeded. For instance, ARPA has adopted the European Union (EU) directive limits, setting $200\mu g/m^3$ as the hourly limit. The World Health Organization (WHO) has recommended more stringent guidelines, to prioritize long-term exposure over short-term spikes, proposing $25\mu g/m^3$ for the daily average concentration. Exceedance probability maps provide a clear and probabilistic interpretation of risk, allowing for the identification of areas that are likely to exceed a specified threshold based on statistical distribution characteristics.

The proposed method could also offer valuable data-driven support to evidence-based decision-making at the local level, for policymakers and environmental regulators. In this respect, it may be useful to aggregate measures at administrative levels, such as municipalities or provinces. For instance, Figure 15 shows the aggregated risk of exceeding the threshold of $100\mu g/m^3$ at the provincial level, for the two considered days. The risk map on the workday of 11^{th} January 2019 (left panel) highlights the provinces of Milano, Monza, Cremona and Lodi, as those facing the worst situation. Consistent with the previous observations on probability density functions, the holiday period results in a significant reduction in NO_2 levels in the largest cities. This

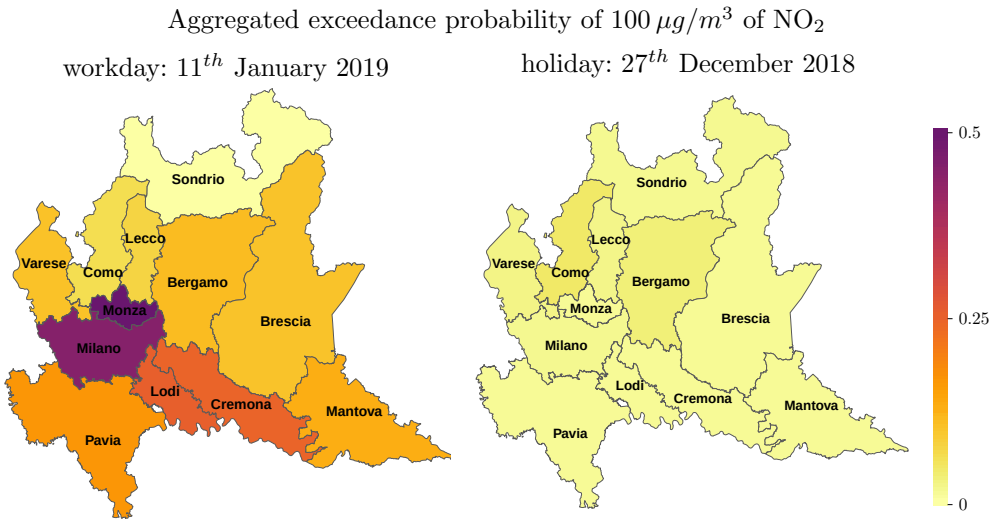


Fig. 15: Aggregated risk of exceeding the probability threshold of $100\mu g/m^3$ NO_2, across the provinces of Lombardy. Left: 11^{th} January 2019 (workday). Right: 27^{th} December 2018 (holiday).

is evidenced by the map on the right of Figure 15, which shows an almost zero probability of exceeding the threshold of $100\mu g/m^3$ across all provinces of Lombardy. In summary, our proposal provides additional analytical tools that can enhance policy-making, offering insights that may support the formulation of targeted interventions, for effective air quality management, grounded in robust, data-driven foundations.

6 Discussion and future work

In this work, we have developed a physics-informed spatial quantile regression model, proposing an innovative method for the joint estimation of several spatial quantiles, that guarantees their monotonicity. The proposed approach enables the direct estimation of multiple monotone quantiles, avoiding any form of post-processing, and encompassing the entire estimation procedure, into a unique and coherent mathematical framework. As shown by the simulation studies in Section 4, MQSR-PDE outperforms the methods currently available in the literature for quantile regression without crossing. Moreover, the methodology proves to be particularly effective in scenarios showing complex spatial dependencies, characterized by strong anisotropies and non-stationary effects, or when physical knowledge is available, such as the presence of winds in the study of air pollutants. In these settings, the ability of the proposed model to include this problem-specific information leads to more accurate estimates, at all the quantile levels. Furthermore, it should be pointed out that the proposed model inherits the ability of the PDE-regularized methods to handle data distributed over domains having complex shapes, including non-convex two-dimensional and three-dimensional domains, as well as curved surfaces; see, e.g., Sangalli (2021); Tomasetto et al. (2024);

Castiglione et al. (2025). The proposed method could thus be used to compute quantiles of variables of interest observed over regions with complicated conformation, such as water bodies with irregular shorelines, or data observed over the globe.

An important direction for future research is the development of appropriate uncertainty quantification tools. In this respect we could relay on asymptotic results, as demonstrated in a Bayesian framework in Fasiolo et al. (2021a). Alternative nonparametric methods, explored by Ferraccioli et al. (2023) and Cavazzutti et al. (2024) in the context of linear regression with PDE penalization, con be extended to the considered quantile regression framework. Another promising direction for future research involves inclusion of the temporal dimension, estimating multiple spatio-temporal quantile maps. Such extension would significantly expand the applicability of the proposed methodology to a wide range of environmental and ecological challenges. For instance, it would enable the creation of multiple quantile spatio-temporal maps that could serve as robust alerting tools. The extension to space-time data could be achieved by appropriately building upon the theory developed by Bernardi et al. (2017) and Arnone et al. (2023) for the linear regression case.

Acknowledgements. M.F. De Sanctis, F. Ieva and L.M. Sangalli acknowledge the project GRINS - Growing Resilient, INclusive and Sustainable (GRINS PE00000018 – CUP D43C22003110001), funded by the European Union - NextGenerationEU programme. The views and opinions expressed are solely those of the authors and do not necessarily reflect those of the European Union, nor can the European Union be held responsible for them. I. Di Battista and L.M. Sangalli acknowledge the PRIN 2022 project CoEnv - Complex Environmental Data and Modeling (CUP 2022E3RY23), funded by the European Union – NextGenerationEU programme, and by the Italian Ministry for University and Research. A. Palummo acknowledges funding by the European Union – NextGenerationEU programme. M. Bernardi acknowledges the PRIN 2022 project SMNNKY (CUP C53D23002580006), funded by the European Union - NextGenerationEU programme, Mission 4 Component 2, and by the Italian Ministry for University and Research. M. Bernardi also acknowledges partial funding by the BERN BIRD2222 01 - BIRD 2022 grant from the University of Padua. Finally, M.F. De Sanctis, I. Di Battista, F. Ieva and L.M. Sangalli also acknowledge the research project Dipartimento di Eccellenza 2023 -2027, Dipartimento di Matematica, Politecnico di Milano, funded by the Italian Ministry for University and Research.

References

Arnone, E., Azzimonti, L., Nobile, F., Sangalli, L.M.: Modeling spatially dependent functional data via regression with differential regularization. J. Multivariate Anal. **170**, 275–295 (2019) https://doi.org/10.1016/j.jmva.2018.09.006

Arnone, E., De Falco, C., Formaggia, L., Meretti, G., Sangalli, L.M.: Computationally efficient techniques for spatial regression with differential regularization. International Journal of Computer Mathematics **100**(10), 1971–1991 (2023) https://doi.org/10.1080/00207160.2023.2239944

Azzimonti, L., Nobile, F., Sangalli, L.M., Secchi, P.: Mixed finite elements for spatial regression with PDE penalization. SIAM/ASA Journal on Uncertainty Quantification **2**(1), 305–335 (2014) https://doi.org/10.1137/130925426

ARPA Lombardia: Il portale ufficiale di ARPA Lombardia — Agenzia Regionale per la Protezione dell'Ambiente. https://www.arpalombardia.it/ (2024)

Azzimonti, L., Sangalli, L.M., Secchi, P., Domanin, M., Nobile, F.: Blood flow velocity field estimation via spatial regression with PDE penalization. J. Amer. Statist. Assoc. **110**(511), 1057–1071 (2015) https://doi.org/10.1080/01621459.2014.946036

Arnone, E., Sangalli, L.M., Vicini, A.: Smoothing spatio-temporal data with complex missing data patterns. Statistical Modelling **23**(4), 327–356 (2023) https://doi.org/10.1177/1471082X211057959

Bernardi, M.S., Carey, M., Ramsay, J.O., Sangalli, L.M.: Modeling spatial anisotropy via regression with partial differential regularization. Journal of Multivariate Analysis **167**, 15–30 (2018) https://doi.org/10.1016/j.jmva.2018.03.014

Brezzi, F., Fortin, M.: Mixed and hybrid finite element method. Springer Series in Computational Mathematics **15**, 350 (1991)

Bassett, G., Koenker, R.: An empirical quantile function for linear models with iid errors. Journal of the American Statistical Association **77**(378), 407–415 (1982)

Bondell, H.D., Reich, B.J., Wang, H.: Noncrossing quantile regression curve estimation. Biometrika **97**(4), 825–838 (2010)

Bernardi, M.S., Sangalli, L.M., Mazza, G., Ramsay, J.O.: A penalized regression model for spatial functional data with application to the analysis of the production of waste in venice province. Stochastic environmental research and risk assessment **31**(1), 23–38 (2017) https://doi.org/10.1007/s00477-016-1237-3

Bakka, H., Vanhatalo, J., Illian, J.B., Simpson, D., Rue, H.: Non-stationary gaussian models with physical barriers. Spatial statistics **29**, 268–288 (2019)

Castiglione, C., Arnone, E., Bernardi, M., Farcomeni, A., Sangalli, L.: Pde-regularised spatial quantile regression. Journal of Multivariate Analysis **205**, 105381 (2025) https://doi.org/10.1016/j.jmva.2024.105381

Cavazzutti, M., Arnone, E., Ferraccioli, F., Galimberti, C., Finos, L., Sangalli, L.M.: Sign-flip inference for spatial regression with differential regularisation. Stat **13**(3), 711 (2024) https://doi.org/10.1002/sta4.711

Cannon, A.J.: Non-crossing nonlinear regression quantiles by monotone composite quantile regression neural network, with application to rainfall extremes. Stochastic Environmental Research and Risk Assessment **32**(11), 3207–3225 (2018)

Clarotto, L., Allard, D., Romary, T., Desassis, N.: The spde approach for spatio-temporal datasets with advection and diffusion. Spatial Statistics **62**, 100847 (2024)

Chernozhukov, V., Fernández-Val, I., Galichon, A.: Quantile and probability curves without crossing. Econometrica **78**(3), 1093–1125 (2010)

Carrizo-Vergara, R., Allard, D., Desassis, N.: A general framework for SPDE-based stationary random fields. Bernoulli **28**(1), 1–32 (2022)

Cressie, N., Wikle, C.K.: Statistics for Spatio-Temporal Data. CourseSmart Series. Wiley, Hoboken, New Jersey (2011)

Das, P., Ghosal, S.: Analyzing ozone concentration by bayesian spatio-temporal quantile regression. Environmetrics **28**(4), 2443 (2017)

Deb, S., Neves, C., Roy, S.: Nonparametric quantile regression for spatio-temporal processes (2024)

European Environment Agency: 98% of EU population exposed to harmful air - EEA report. https://eeb.org/98-of-eu-population-exposed-to-harmful-air-eea-report/ (2018)

Frumento, P., Bottai, M.: Parametric modeling of quantile regression coefficient functions. Biometrics **72**(1), 74–84 (2016)

Frumento, P., Bottai, M., Fernández-Val, I.: Parametric modeling of quantile regression coefficient functions with longitudinal data. Journal of the American Statistical Association **116**(534), 783–797 (2021)

Ferraccioli, F., Sangalli, L.M., Finos, L.: Nonparametric tests for semiparametric regression models. Test **32**(3), 1106–1130 (2023) https://doi.org/10.1007/s11749-023-00868-9

Fasiolo, M., Wood, S.N., Zaffran, M., Nedellec, R., Goude, Y.: Fast Calibrated Additive Quantile Regression. Journal of the American Statistical Association **116**(535), 1402–1412 (2021)

Fasiolo, M., Wood, S.N., Zaffran, M., Nedellec, R., Goude, Y.: `qgam`: Bayesian nonparametric quantile regression modeling in `R`. Journal of Statistical Software **100**(9), 1–31 (2021)

He, X.: Quantile curves without crossing. The American Statistician **51**(2), 186 (1997)

Hefley, T.J., Hooten, M.B., Russell, R.E., Walsh, D.P., Powell, J.A.: When mechanism matters: Bayesian forecasting using models of ecological diffusion. Ecology Letters **20**(5), 640–650 (2017)

Hunter, D.R., Lange, K.: Quantile regression via an MM algorithm. Journal of

Computational and Graphical Statistics **9**(1), 60–77 (2000)

Koenker, R., Bassett, G. Jr.: Regression quantiles. Econometrica **46**(1), 33–50 (1978)

Kuhnert, P.: Physical-statistical modelling. Environmetrics **25**(4), 201–202 (2014)

Lange, K.: MM Optimization Algorithms, p. 223. Society for Industrial and Applied Mathematics, Philadelphia, PA (2016)

Lindgren, F., Bolin, D., Rue, H.: The spde approach for gaussian and non-gaussian fields: 10 years and still running. Spatial Statistics **50**, 100599 (2022)

Li, Y., Liu, Y., Zhu, J.: Quantile regression in reproducing kernel Hilbert spaces. Journal of the American Statistical Association **102**(477), 255–268 (2007)

Lindgren, F., Rue, H., Lindst, J.: An explicit link between gaussian fields and gaussian markov random fields: The spde approach. Journal of the Royal Statistical Society. Series B. Statistical Methodology **73** (2011)

Liu, Y., Wu, Y.: Stepwise multiple quantile regression estimation using non-crossing constraints. Statistics and Its Interface **2**(3), 299–310 (2009)

Liu, Y., Wu, Y.: Simultaneous multiple non-crossing quantile regression estimation using kernel constraints. Journal of Nonparametric Statistics **23**(2), 415–437 (2011)

Lange, K., Zhou, H.: A legacy of em algorithms. International Statistical Review **90**(S1), 52–66 (2022)

Mammen, E.: Nonparametric regression under qualitative smoothness assumptions. The Annals of Statistics **19**(2), 741–759 (1991)

Nychka, D., Gray, G., Haaland, P., Martin, D., O'Connell, M.: A nonparametric regression approach to syringe grading for quality improvement. Journal of the American Statistical Association **90**(432), 1171–1178 (1995)

Neocleous, T., Portnoy, S.: On monotonicity of regression quantile functions. Statistics and Probability Letters **78**(10), 1226–1229 (2008)

Nocedal, J., Wright, S.J.: Numerical Optimization, 2nd edn. Springer Series in Operations Research and Financial Engineering, p. 664. Springer, New York (2006)

Open Data: Regione Lombardia. https://www.dati.lombardia.it/stories/s/Meteo-inquinamento-aria-e-acqua-e-altri-dati-da-AR/auv9-c2sj (2024)

Palummo, A., Arnone, E., Clemente, A., Sangalli, L.M., Ramsay, J., Formaggia, L.: fdaPDE: Physics-Informed Spatial and Functional Data Analysis. GitHub, https://github.com/fdaPDE/fdaPDE (2025)

Quarteroni, A., Sacco, R., Saleri, F., Gervasio, P.: Matematica Numerica. UNITEXT, vol. 77. Springer, Milan (2014)

Quarteroni, A.: Numerical Models for Differential Problems, 3rd edn. MS&A. Modeling, Simulation and Applications, vol. 16, p. 681. Springer, Cham (2017)

Ramsay, T.: Spline smoothing over difficult regions. Journal of the Royal Statistical Society. Series B. Statistical Methodology **64**(2), 307–319 (2002)

Regione Lombardia: Geoportale Lombardia (2024). https://www.geoportale.regione.lombardia.it/

Reich, B.J., Fuentes, M., Dunson, D.B.: Bayesian spatial quantile regression. Journal of the American Statistical Association **106**(493), 6–20 (2011)

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

Richardson, R.A.: Sparsity in nonlinear dynamic spatiotemporal models using implied advection. Environmetrics **28**(6), 2456 (2017)

Sangalli, L.M.: Spatial regression with partial differential equation regularisation. International Statistical Review (2021) https://doi.org/10.1111/insr.12444

Sun, Y., Babu, P., Palomar, D.P.: Majorization-minimization algorithms in signal processing, communications, and machine learning. IEEE Transactions on Signal Processing **65**(3), 794–816 (2017)

Sottile, G., Frumento, P.: Parametric estimation of non-crossing quantile functions. Statistical Modelling **23**(2), 173–195 (2023)

Sangalli, L.M., Ramsay, J.O., Ramsay, T.O.: Spatial spline regression models. Journal of the Royal Statistical Society. Series B. Statistical Methodology **75**(4), 681–703 (2013) https://doi.org/10.1111/rssb.12009

Tomasetto, M., Arnone, E., Sangalli, L.M.: Modeling anisotropy and non-stationarity through physics-informed spatial regression. Environmetrics (2024) https://doi.org/10.1002/env.2889

Takeuchi, I., Le, Q.V., Sears, T.D., Smola, A.J.: Nonparametric quantile estimation. Journal of Machine Learning Research **7**(45), 1231–1264 (2006)

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

Wikle, C.K., Hooten, M.: A general science-based framework for dynamical spatio-temporal models. TEST **19**, 417–451 (2010)

Wikle, C.K.: Hierarchical bayesian models for predicting the spread of ecological processes. Ecology **84**(6), 1382–1394 (2003)

Wu, T., Lange, K.: The MM alternative to EM. Statistical Science **25**(4) (2010)

Woodbury, M.A.: Inverting Modified Matrices. Memorandum Report / Statistical Research Group, Princeton. Department of Statistics, Princeton University (1950)

Wood, S.N.: Thin plate regression splines. Journal of the Royal Statistical Society. Series B. Statistical Methodology **65**(1), 95–114 (2003)

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

Youngman, B.D.: Generalized additive models for exceedances of high thresholds with an application to return level estimation for us wind gusts. Journal of the American Statistical Association (2019)

Yang, Y., Tokdar, S.: Joint estimation of quantile planes over arbitrary predictor spaces. Journal of the American Statistical Association (2015)

Yuan, M.: GACV for quantile smoothing splines. Computational Statistics & Data Analysis **50**(3), 813–829 (2006)

Appendix A Appendix A

A.1 Matrix definitions

In this section, we provide the exact analytical definitions of the matrices appearing in Proposition 1. Let us denote by $\hat{\boldsymbol{\beta}}^{(k-1)}$ and $\hat{\boldsymbol{f}}^{(k-1)}$ the estimates for $\boldsymbol{\beta}$ and $\boldsymbol{f}$, at the $(k-1)$-th step of the MM algorithm. Also, we introduce the difference matrices

$$D_j^{(m)} = (O_m \dots O_m \underbrace{-I_m}_{j} \underbrace{I_m}_{j+1} O_m \dots O_m) \in \mathbb{R}^{m \times rm}, \quad \forall j = 1, .., r-1, \ \forall m \in \mathbb{N},$$

and

$$D^{(m)} = \begin{pmatrix} D_1^{(m)} \\ \vdots \\ D_{r-1}^{(m)} \end{pmatrix}.$$

Here, O_m is the zero square matrix of dimension m. Then, we define the following quantities

$$w_{ij} = |y_i - \mathbf{x}_i^\top \hat{\boldsymbol{\beta}}_j^{(k-1)} - \hat{f}_j^{(k-1)}(\mathbf{p}_i)|, \quad \forall i = 1, \dots, n, \quad \forall j = 1, \dots, r$$
$$\delta_{ij} = |\varepsilon - \mathbf{x}_i^\top D_j^{(q)} \hat{\boldsymbol{\beta}}^{(k-1)} - (D_j^{(n)} \hat{\boldsymbol{f}}_n^{(k-1)})_i|, \quad \forall i = 1, \dots, n, \quad \forall j = 1, \dots, r-1,$$

The former represents the quantile residuals, the latter represents the constraint violation residuals. Given those definitions, we can construct the correspondent diagonal and block-diagonal matrices as

$$W_j = \begin{bmatrix} \frac{1}{2nw_{1j}} & & \\ & \ddots & \\ & & \frac{1}{2nw_{nj}} \end{bmatrix} \quad \forall j = 1, \dots, r, \quad W_f = \begin{bmatrix} W_1 & & \\ & \ddots & \\ & & W_r \end{bmatrix}, \quad W_\beta = \tilde{X}^\top W_f \tilde{X},$$
$$\Delta_j = \begin{bmatrix} \frac{1}{2\delta_{1j}} & & \\ & \ddots & \\ & & \frac{1}{2\delta_{nj}} \end{bmatrix} \quad \forall j = 1, \dots, r-1, \quad \Delta_f = \begin{bmatrix} \Delta_1 & & \\ & \ddots & \\ & & \Delta_{r-1} \end{bmatrix}, \quad \Delta_\beta = \tilde{X}^\top \Delta_f \tilde{X},$$

where $\tilde{X} = I_r \otimes X$. Denoting by $\mathbf{z}_j = \mathbf{y} - (\frac{1}{2} - \alpha_j) W_j^{-1} \mathbf{e}^{(n)} \in \mathbb{R}^n$ the vector of pseudo-observations for the j-th quantile level, we define

$$\boldsymbol{\xi}_\beta = (\mathbf{z}_1^\top W_1 X, \cdots, \mathbf{z}_r^\top W_r X)^\top \in \mathbb{R}^{rq}, \qquad \boldsymbol{\xi}_f = (\mathbf{z}_1^\top W_1, \cdots, \mathbf{z}_r^\top W_r)^\top \in \mathbb{R}^{rn}.$$

Finally, we set

$$\begin{aligned} T_f &= {D^{(n)}}^\top \Delta D^{(n)} \\ T_\beta &= {D^{(q)}}^\top (I_{r-1} \otimes X^\top) \Delta (I_{r-1} \otimes X) D^{(q)} \\ C &= W_f \tilde{X} + \gamma D^{(n)} \Delta \tilde{X} D^{(q)} \\ \mathbf{v}_f &= \varepsilon {D^{(n)}}^\top \Delta \mathbf{e}^{((r-1)n)} + \frac{1}{2} \boldsymbol{l}_{rn} \\ \mathbf{v}_\beta &= \tilde{X} \mathbf{v}_f, \end{aligned}$$

where $\boldsymbol{l}_{rn} = (-\mathbf{e}^{(n)\top}, 0, \cdots, 0, \mathbf{e}^{(n)\top})^\top \in \mathbb{R}^{rn}$.

Notice that the completed matrices W and T and the vectors $\boldsymbol{\xi}$ and $\mathbf{v}$ can be decomposed as

$$W = \begin{bmatrix} W_\beta & C^\top \\ C & W_f \end{bmatrix}, \quad T = \begin{bmatrix} T_\beta & 0 \\ 0 & T_f \end{bmatrix}, \quad \boldsymbol{\xi} = (\boldsymbol{\xi}_\beta, \boldsymbol{\xi}_f), \quad \mathbf{v} = (\mathbf{v}_\beta, \mathbf{v}_f),$$

where we emphasize the individual contributions of the parametric and nonparametric terms of the model, and the corresponding interaction term, encoded by the matrix C.

A.2 Model estimation

Proof of Proposition 1.

With the same notation of Section A.1, we indicate as $\hat{\boldsymbol{\beta}}^{(k-1)}$ and $\hat{\boldsymbol{f}}^{(k-1)}$ the estimates for $\boldsymbol{\beta}$ and $\boldsymbol{f}$, at the $(k-1)$-th step of the MM algorithm. Moreover, with the notation $(\cdot)_i$ we refer to the i-th component of a vector. Finally, let us refer to the data misfit and the crossing penalty terms appearing in (4) as follows:

$$\frac{1}{n}\sum_{i,j=1}^{n,r}\rho_{\alpha_j}(y_i-\mathbf{x}_i^\top\boldsymbol{\beta}_j-f_j(\mathbf{p}_i))=(\star)$$

$$\sum_{i,j=1}^{n,r-1}\max\{0,\varepsilon-(\mathbf{x}_i^\top\boldsymbol{\beta}_{j+1}+f_{j+1}(\mathbf{p}_i)+\mathbf{x}_i^\top\boldsymbol{\beta}_j+f_j(\mathbf{p}_i))\}=(\star\star).$$

To find a functional majorizer for the k-th step of the iterative algorithm, it is convenient to represent $\rho_{\alpha_j}(\cdot)$ and $\max(\cdot)$ as $\rho_{\alpha_j}(t)=\frac{1}{2}|t|-(\frac{1}{2}-\alpha_j)t$ and $\max(t)=\frac{1}{2}(|t|+t)$. On the other hand, we can write $|t|=\sqrt{t^2}$, and use the first-order Taylor expansion for $\sqrt{\cdot}$ to find a quadratic upper bound of $|\cdot|$. Therefore, applying this idea to the terms in (4), with the center of the Taylor expansion in $(\hat{\beta}_j^{(k-1)},\hat{f}_j^{(k-1)}(\mathbf{p}_i))$, we have that

$$(\star)\leq\frac{1}{n}\sum_{i,j=1}^{n,r}\left(\frac{(y_i-\mathbf{x}_i^\top\boldsymbol{\beta}_j-f_j(\mathbf{p}_i))^2}{4w_{ij}}+\frac{w_{ij}}{4}-\big(\frac{1}{2}-\alpha_j\big)(y_i-\mathbf{x}_i^\top\boldsymbol{\beta}_{j+1}-f_j(\mathbf{p}_i))\right)$$

$$(\star\star)\leq\sum_{i,j=1}^{n,r-1}\left(\frac{(\varepsilon-\mathbf{x}_i^\top D_j^{(q)}\boldsymbol{\beta}-(D_j^{(n)}\boldsymbol{f}_n)_i)^2}{4\delta_{ij}}+\frac{\delta_{ij}}{4}+\frac{1}{2}(\varepsilon-\mathbf{x}_i^\top D_j^{(q)}\boldsymbol{\beta}-(D_j^{(n)}\boldsymbol{f}_n)_i)\right).$$

First, let us consider inequality $(\star)$. Putting all the terms which do not depend on $(\boldsymbol{\beta},\boldsymbol{f})$ in a unique constant c, we have that

$$\begin{aligned}(\star)&\leq\sum_{j=1}^{r}\Big(\frac{1}{2}\boldsymbol{f}_{n,j}^\top W_j\boldsymbol{f}_{n,j}+\frac{1}{2}\boldsymbol{\beta}_j^\top X^\top W_jX\boldsymbol{\beta}_j-\boldsymbol{f}_{n,j}^\top W_j\mathbf{y}\\&\quad+(\tfrac{1}{2}-\alpha_j)\boldsymbol{f}_{n,j}^\top\mathbf{e}^{(n)}-\boldsymbol{\beta}_j^\top X^\top W_j\mathbf{y}+(\tfrac{1}{2}-\alpha_j)\boldsymbol{\beta}_j^\top X^\top\mathbf{e}^{(n)}+c\Big)\\&=\frac{1}{2}(\boldsymbol{f}_n^\top W_f\boldsymbol{f}_n+\boldsymbol{\beta}^\top W_\beta\boldsymbol{\beta})-\boldsymbol{f}_n^\top\boldsymbol{\xi}_f-\boldsymbol{\beta}^\top\boldsymbol{\xi}_\beta+\boldsymbol{f}_n^\top W_f\tilde{X}\boldsymbol{\beta}.\end{aligned}$$

Let us now consider $(\star\star)$. Following the same idea,

$$
\begin{aligned}
(\star\star) \leq \frac{1}{2}\sum_{j=1}^{r-1}(\boldsymbol{f}_n^\top {D_j^{(n)}}^\top \Delta_j D_j^{(n)} \boldsymbol{f}_n + \boldsymbol{\beta}^\top {D_j^{(q)}}^\top X^\top \Delta_j X D_j^{(q)} \boldsymbol{\beta} \\
- 2(\boldsymbol{f}_n^\top {D_j^{(n)}}^\top + \boldsymbol{\beta}^\top {D_j^{(q)}}^\top X^\top)(\Delta_j \boldsymbol{\varepsilon}_n + \frac{1}{2}\mathbf{e}^{(n)})) + 2\boldsymbol{f}_n^\top {D_j^{(n)}}^\top \Delta_j X D_j^{(q)} \boldsymbol{\beta} + c \\
= \frac{1}{2}(\boldsymbol{f}_n^\top T_f \boldsymbol{f}_n + \boldsymbol{\beta}^\top T_\beta \boldsymbol{\beta} - 2\boldsymbol{f}_n^\top \mathbf{v}_f - 2\boldsymbol{\beta}^\top \mathbf{v}_\beta) + \boldsymbol{f}^\top {D^{(n)}}^\top \Delta \tilde{X} D^{(q)} \boldsymbol{\beta} + c.
\end{aligned}
$$

Combining the two terms, we have

$$
\begin{aligned}
(\star) + \gamma(\star\star) + P(\boldsymbol{f}) \leq \frac{1}{2}(\boldsymbol{f}_n^\top W_f \boldsymbol{f}_n + \boldsymbol{\beta}^\top W_\beta \boldsymbol{\beta}) - \boldsymbol{f}_n^\top \boldsymbol{\xi}_f - \boldsymbol{\beta}^\top \boldsymbol{\xi}_\beta + \boldsymbol{f}_n^\top W_f \tilde{X} \boldsymbol{\beta} + \\
+ \frac{\gamma}{2}(\boldsymbol{f}_n^\top T_f \boldsymbol{f}_n + \boldsymbol{\beta}^\top T_\beta \boldsymbol{\beta} - 2\boldsymbol{f}_n^\top \mathbf{v}_f - 2\boldsymbol{\beta}^\top \mathbf{v}_\beta) + \\
+ \gamma \boldsymbol{f}^\top {D^{(n)}}^\top \Delta \tilde{X} D^{(q)} \boldsymbol{\beta} + P(\boldsymbol{f}) + c.
\end{aligned}
$$

Finally, setting $C = W_f \tilde{X} + \gamma D^{(n)} \Delta \tilde{X} D^{(q)}$ and using the definitions of $\boldsymbol{\eta}, W, T, \boldsymbol{\xi}$ and $\mathbf{v}$ stated in Section 3.1, we have established that there exists a functional $\mathcal{G}$ such that

$$
J_P(\boldsymbol{\eta}) \leq \frac{1}{2}\boldsymbol{\eta}^\top W \boldsymbol{\eta} - \boldsymbol{\eta}^\top(\boldsymbol{\xi} + \gamma \mathbf{v}) + \frac{\gamma}{2}(\boldsymbol{\eta}^\top T \boldsymbol{\eta}) + \sum_{j=1}^{r} \lambda_j P(f_j) + c =: \mathcal{G}(\boldsymbol{\eta}) + c \quad \forall \boldsymbol{\eta},
$$

that is the dominance condition stated in (5). Finally, thanks to the employment of the local quadratic approximation of $|\cdot|$, we are guaranteed that the tangency condition is satisfied, meaning that $\mathcal{G}$ is a surrogate functional for J_P.

□

Proof of Proposition 2.

Let $(\hat{\boldsymbol{\beta}}^{(k)}, \hat{\boldsymbol{f}}^{(k)})$ be the minimum of $\mathcal{G}(\boldsymbol{\beta}, \boldsymbol{f})$, at the generic k-th step of the algorithm. Let $\boldsymbol{f}_n$ be the evaluation of $\boldsymbol{f}$ at the spatial locations $\mathbf{p}_1, \ldots, \mathbf{p}_n$, as detailed in Section 3.1. Then, for all $\psi \in \mathcal{F}_0$, the minimizer $(\hat{\boldsymbol{\beta}}^{(k)}, \hat{\boldsymbol{f}}^{(k)})$ must satisfy

$$
\frac{\partial}{\partial \boldsymbol{\beta}} \mathcal{G}(\hat{\boldsymbol{\beta}}^{(k)}, \hat{\boldsymbol{f}}^{(k)}) = 0, \qquad \frac{\partial}{\partial t} \mathcal{G}(\hat{\boldsymbol{\beta}}^{(k)}, \hat{\boldsymbol{f}}^{(k)} + t\boldsymbol{\phi})\Big|_{t=0} = 0.
$$

Working on the first equation, we have

$$
\frac{\partial}{\partial \boldsymbol{\beta}} \mathcal{G}(\hat{\boldsymbol{\beta}}^{(k)}, \hat{\boldsymbol{f}}^{(k)}) = 0 \implies W_\beta \hat{\boldsymbol{\beta}}^{(k)} - \boldsymbol{\xi}_\beta + \gamma(T_\beta \hat{\boldsymbol{\beta}}^{(k)} - \mathbf{v}_\beta) + C^\top \hat{\boldsymbol{f}}_n^{(k)} = 0.
$$

Following the discussion in Azzimonti et al. (2014), the second equation takes the form

$$\begin{aligned}\left.\frac{\partial \mathcal{G}(\hat{\boldsymbol{\beta}}^{(k)}, \hat{\boldsymbol{f}}^{(k)} + t\boldsymbol{\psi})}{\partial t}\right|_{t=0} = 0 \implies &\Big(\boldsymbol{\psi}_{rn}^\top (W_f + \gamma T_f)(\hat{\boldsymbol{f}}_n^{(k)} + t\boldsymbol{\psi}_{rn}) - \boldsymbol{\psi}_{rn}^\top (\boldsymbol{\xi}_f + \gamma \mathbf{v}_f) \\ &+ \boldsymbol{\psi}_{rn} C\boldsymbol{\beta} + \sum_{j=1}^{r} \lambda_j \int_{\mathcal{D}} (L\psi(\mathbf{p})) L\hat{f}_j^{(k)}(\mathbf{p}) d\mathbf{p}\Big)\Big|_{t=0} = 0 \\ \implies &\boldsymbol{\psi}_{rn}^\top (W_f + \gamma T_f)\hat{\boldsymbol{f}}_n^{(k)} - \boldsymbol{\psi}_{rn}^\top (\boldsymbol{\xi}_f + \gamma \mathbf{v}_f) \\ &+ \boldsymbol{\psi}_{rn} C\boldsymbol{\beta} + \sum_{j=1}^{r} \lambda_j \int_{\mathcal{D}} (L\psi(\mathbf{p})) L\hat{f}_j^{(k)}(\mathbf{p}) d\mathbf{p} = 0.\end{aligned}$$

Combining the two results, we obtain

$$\begin{cases} W_\beta \hat{\boldsymbol{\beta}}^{(k)} - \boldsymbol{\xi}_\beta + \gamma(T_\beta \hat{\boldsymbol{\beta}}^{(k)} - \mathbf{v}_\beta) + C^\top \hat{\boldsymbol{f}}_n^{(k)} = 0, \\ \boldsymbol{\psi}_{rn}^\top W_f \hat{\boldsymbol{f}}_n^{(k)} - \boldsymbol{\psi}_{rn}^\top \boldsymbol{\xi}_f + \gamma(\boldsymbol{\psi}_{rn}^\top T_f \hat{\boldsymbol{f}}_n^{(k)} - \boldsymbol{\psi}_{rn}^\top \mathbf{v}_f) + \boldsymbol{\psi}_{rn}^\top C\hat{\boldsymbol{\beta}}^{(k)} + \sum_{j=1}^{r} \lambda_j \int_{\mathcal{D}} (L\psi(\mathbf{p})) \hat{g}_j^{(k)}(\mathbf{p}) d\mathbf{p} = 0, \\ \int_{\mathcal{D}} (L\hat{f}_j^{(k)})\phi - \int_{\mathcal{D}} \hat{g}_j^{(k)} \phi = 0 \qquad \forall \phi \in \mathcal{F}_0,\ \forall j = 1, \dots, r. \end{cases}$$

Finally, exploiting the definitions of the bilinear forms given in (9), we have

$$\begin{cases} W_\beta \hat{\boldsymbol{\beta}}^{(k)} - \boldsymbol{\xi}_\beta + \gamma(T_\beta \hat{\boldsymbol{\beta}}^{(k)} - \mathbf{v}_\beta) + C^\top \hat{\boldsymbol{f}}_n^{(k)} = 0, \\ \boldsymbol{\psi}_{rn}^\top W_f \hat{\boldsymbol{f}}_n^{(k)} - \boldsymbol{\psi}_{rn}^\top \boldsymbol{\xi}_f + \gamma(\boldsymbol{\psi}_{rn}^\top T_f \hat{\boldsymbol{f}}_n^{(k)} - \boldsymbol{\psi}_{rn}^\top \mathbf{v}_f) + \boldsymbol{\psi}_{rn}^\top C\hat{\boldsymbol{\beta}}^{(k)} + \sum_{j=1}^{r} \lambda_j R_1(\psi, \hat{g}_j^{(k)}) = 0, \\ R_1(\phi, \hat{f}_j^{(k)}) - R_0(\phi, \hat{g}_j^{(k)}) = 0 \qquad \forall \phi \in \mathcal{F}_0,\ \forall j = 1, \dots, r. \end{cases}$$

This concludes the proof.

□

A.3 Discretized linear system

In this section, we provide further reasoning on the linear system (10).

Figure A1 illustrates the block structure of the matrix in the left hand side of the system. This matrix consists of nine main blocks. The top row and the first column refers to the parametric term. Specifically, the two rectangular blocks, colored in yellow, account for the interaction between the parametric and the nonparametric part of the model. The central block of the system matrix, instead, refers to the nonparametric term $\mathbf{f}_h$ and, together with the previously mentioned blocks, contains crossing penalty terms between consecutive quantiles. The enforcement of the crossing penalties gives to those matrices a tridiagonal block structure, with the off-diagonal terms representing the monotonicity constraints. Finally, in the south-west part of the main matrix, we have $\tilde{R}_0$ and $\tilde{R}_1$, correspondent to the mass and stiffness discretizations of the bilinear forms (9). The blank spaces in the figure indicate zero entries, highlighting the sparsity structure of the matrix. This sparsity property ensures computational efficiency in the resolution of the linear system (10), at each iteration of the MM algorithm.

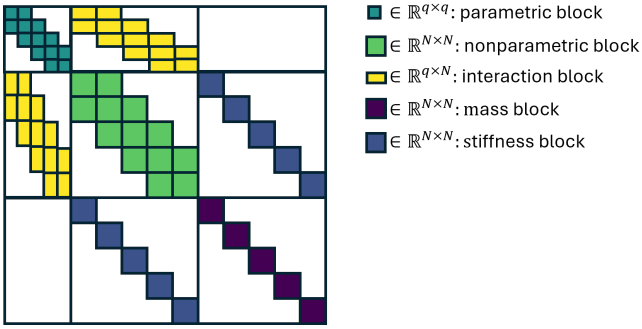


Fig. A1: Block structure of the matrix system (10). The color legend illustrates the role of each block in the estimation problem, with empty spaces representing zero entries. Here, q is the number of covariates, and N denotes the dimension of the basis used for the spatial discretization. The first 2×2 block of the system matrix exhibits a tridiagonal block structure, within each sub-block, resulting from the non-crossing penalties imposed between consecutive quantiles.

Starting from the linear system (10), we can manipulate its blocks to write the system in the same form as presented in Sangalli (2021). This allows us to benefit from the computational properties stated in Arnone et al. (2023) for the semiparametric regression case, relying on the Sherman-Morrison-Woodbury decomposition. With this purpose, we set

$$\tilde{W} := W_f + \gamma T_f = W_f + \gamma D^{(n)\top} \Delta_f D^{(n)}.$$

Then, by noticing that $D^{(q)\top}(I_{r-1} \otimes X^\top)\Delta(I_{r-1} \otimes X)D^{(q)} = \tilde{X}^\top D^{(n)\top} \Delta D^{(n)} \tilde{X}$, we get

$$\begin{aligned} W_\beta + \gamma T_\beta &= \tilde{X}^\top W_f \tilde{X} + \gamma D^{(q)\top}(I_{r-1} \otimes X^\top)\Delta(I_{r-1} \otimes X)D^{(q)} \\ &= \tilde{X}^\top (W_f + \gamma D^{(q)\top} \Delta D^{(n)})\tilde{X} = \tilde{X}^\top \tilde{W} \tilde{X}. \end{aligned}$$

Moreover, with the same reasoning as above, the interaction matrix C can be written as

$$C = W_f \tilde{X} + \gamma D^{(n)} \Delta \tilde{X} D^{(q)} = (W_f + \gamma D^{(n)} \Delta D^{(n)})\tilde{X} = \tilde{W}\tilde{X}.$$

We can now manipulate the right-hand-side of the matrix system (10). In particular, setting $\mathbf{z} = (\mathbf{z}_1, \dots, \mathbf{z}_r)$, and $\tilde{\mathbf{z}} := W_f \mathbf{z} + \gamma \mathbf{v}_f$ we have

$$\begin{aligned} \boldsymbol{\xi}_\beta + \gamma \mathbf{v}_\beta &= (\mathbf{z}_1^\top W_1 X, \cdots, \mathbf{z}_r^\top W_r X)^\top + \gamma \tilde{X} \varepsilon D^{(q)\top} \Delta \mathbf{e}^{((r-1)n)} + \frac{1}{2} \boldsymbol{l}_{rn} \\ &= \tilde{X}^\top W_f \mathbf{z} + \gamma \tilde{X}^\top \mathbf{v}_f = \tilde{X}^\top \tilde{\mathbf{z}} \\ \boldsymbol{\xi}_f + \gamma \mathbf{v}_f &= W_f \mathbf{z} + \gamma \mathbf{v}_f = \tilde{\mathbf{z}}. \end{aligned}$$

Therefore, the linear system (10) takes the form

$$\begin{bmatrix} \tilde{X}^\top \tilde{W} \tilde{X} & \tilde{X}^\top \tilde{W} \tilde{\Psi} & 0 \\ \tilde{\Psi}^\top \tilde{W} \tilde{X} & \tilde{\Psi}^\top \tilde{W} \tilde{\Psi} & \tilde{R}_1^\top \\ 0 & \tilde{R}_1 & -\tilde{R}_0 \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}_h \\ \mathbf{f}_h \\ \mathbf{g}_h \end{bmatrix} = \begin{bmatrix} \tilde{X}^\top \tilde{\mathbf{z}} \\ \tilde{\Psi}^\top \tilde{\mathbf{z}} \\ 0 \end{bmatrix}.$$

Using such reformulation permits a significant computational saving, as shown by Arnone et al. (2023), in a simpler linear regression setting.

A.4 Simulation 2: heteroscedastic Skew-T process

Figure A2 presents the mean estimated quantile fields for all competing methods in the simulation study discussed in Section 4.2, at quantile levels $\alpha = 50\%, 95\%, 99\%$. Among the approaches evaluated, MQSR-PDE provides the most accurate reconstruction of the true quantile field at all levels, with particularly pronounced improvements at the 99% level. Furthermore, TPS-R and SOAP-R exhibit difficulties in accurately reconstructing the field, especially in the northwest region of the domain, where they systematically underestimate the values.

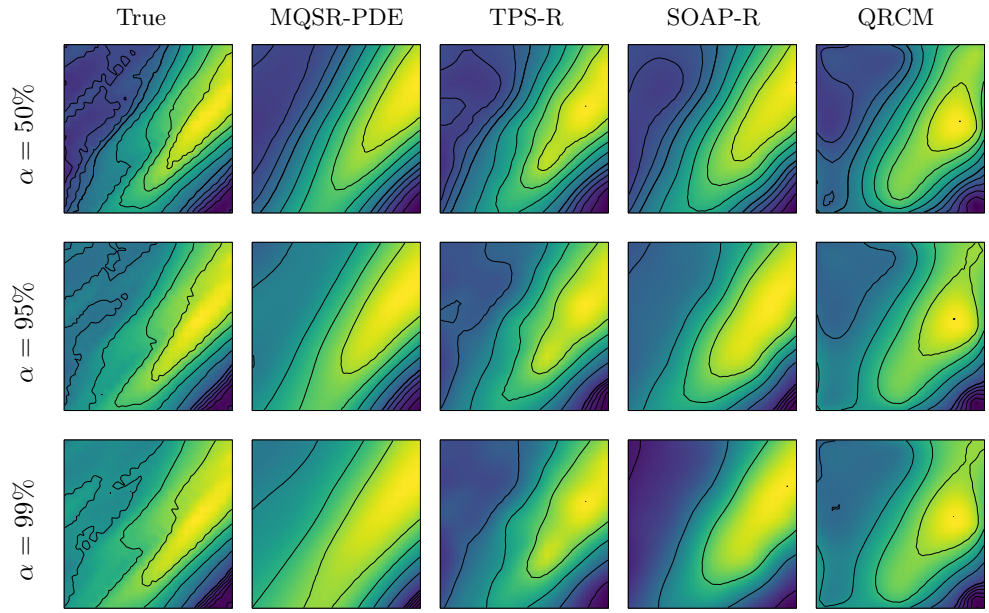


Fig. A2: Simulation 2 (Section 4.2): heteroscedastic Skew-T process. Estimated quantile maps for some of the considered quantile levels. The first column presents the true quantile field and the subsequent columns display the mean estimates for each of the competing methods: the proposed multiple quantile spatial regression model (MQSR-PDE); the quantile regression based on a thin-plate-spline with rearrangement (TPS-R); the quantile regression based on a Soap film smoothing with rearrangement (SOAP-R); the Quantile Regression Coefficient Modeling (QRCM).

Figure A3 shows a comparison, in terms of RMSE, between the proposed multiple estimation method, MQSR-PDE, its single-quantile counterpart, QSR-PDE, and its rearranged version, QSR-PDE-R. As expected, in the central region of the distribution, all three methods yield comparable results since the estimated quantiles

are sufficiently spaced apart and thus less susceptible to the crossing problem. However, in the right tail of the distribution, MQSR-PDE consistently outperforms both QSR-PDE and QSR-PDE-R, with particularly notable improvements at higher quantile levels. Moreover, the difference in errors between MQSR-PDE and QSR-PDE-R underscores the advantages of simultaneous estimation over a simple post hoc rearrangement. This conclusion is further supported by the results of univariate pairwise Wilcoxon tests, conducted for each quantile in the right tail, as shown in Figure A3.

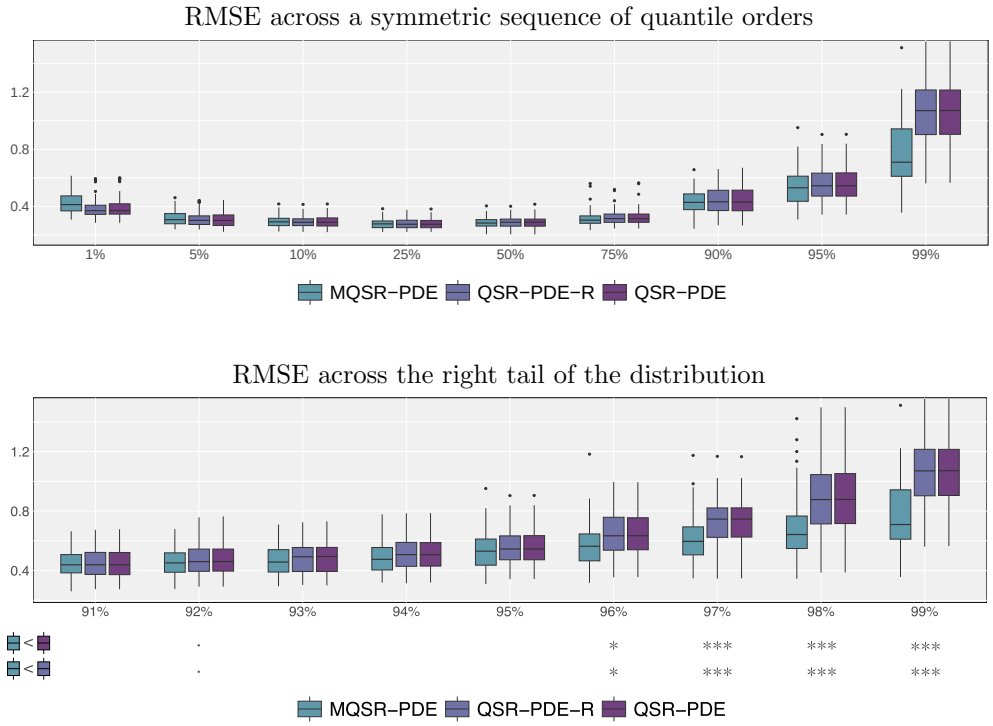


Fig. A3: Simulation 2 (Section 4.2): heteroscedastic Skew-T process. Boxplots of RMSE obtained by different PDE-regularized quantile regression methods: the proposed model for simultaneous quantile estimation (MQSR-PDE); the single-quantile model presented by Castiglione et al. (2025) (QSR-PDE); the latter model combined with the post-processing procedure discussed by Chernozhukov et al. (2010) to ensure non-crossing of the quantiles (QSR-PDE-R). Top panel: RMSE for a symmetric sequence of quantile levels. Bottom panel: RMSE for quantile levels from 91% to 99%. The bottom panel also reports, for each quantile level of the right tail of the distribution, the results of pairwise Wilcoxon tests verifying that MQSR-PDE estimates have significant lower RMSE than QSR-PDE-R estimates, and of QSR-PDE estimates. Legend: . : $0.05 <$ p-value < 0.1; * : $0.01 <$ p-value < 0.05; ** : $0.001 <$ p-value < 0.01; *** : p-value < 0.001.

A.5 Probability density reconstruction

In this section, we describe the algorithm used to reconstruct the probability density functions at fixed spatial locations, as shown in Figure 14. Starting from the fitted quantiles at the spatial location of interest, we invert them to obtain the corresponding Cumulative Distribution Functions (CDF). We then solve a constrained optimization problem to ensure that the parameters define a valid density function. Finally, we differentiate these curves to obtain the desired density. We note that, in principle, any interpolation method could be used for the quantile inversion, but we specifically rely on Gaussian CDF. This choice is particularly convenient due to their monotonic behavior and the availability of an analytically known derivative.

MOX Technical Reports, last issues

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

30/2025 Rosafalco, L.; Conti, P.; Manzoni, A.; Mariani, S.; Frangi, A.
Online learning in bifurcating dynamic systems via SINDy and Kalman filtering

31/2025 Botteghi, N.; Fresca, S.; Guo, M.; Manzoni, A.
HypeRL: Parameter-Informed Reinforcement Learning for Parametric PDEs

29/2025 Centofanti, E.; Ziarelli, G.; Parolini, N.; Scacchi, S.; Verani, M. ; Pavarino, L. F.
Learning cardiac activation and repolarization times with operator learning

28/2025 Ciaramella, G.; Gander, M.J.; Mazzieri, I.
Discontinuous Galerkin time integration for second-order differential problems: formulations, analysis, and analogies

27/2025 Antonietti P.F.; Artoni, A.; Ciaramella, G.; Mazzieri, I.
A review of discontinuous Galerkin time-stepping methods for wave propagation problems

24/2025 Bartsch, J.; Borzi, A.; Ciaramella, G.; Reichle, J.
Adjoint-based optimal control of jump-diffusion processes

22/2025 Leimer Saglio, C. B.; Pagani, S.; Antonietti P. F.
A p-adaptive polytopal discontinuous Galerkin method for high-order approximation of brain electrophysiology

23/2025 Antonietti, P. F.; Caldana, M.; Mazzieri, I.; Re Fraschini, A.
MAGNET: an open-source library for mesh agglomeration by Graph Neural Networks

21/2025 Caldera, L., Masci, C., Cappozzo, A., Forlani, M., Antonelli, B., Leoni, O., Ieva, F.
Uncovering mortality patterns and hospital effects in COVID-19 heart failure patients: a novel Multilevel logistic cluster-weighted modeling approach

20/2025 Botti, M.; Prada, D.; Scotti, A.; Visinoni, M.
Fully-Mixed Virtual Element Method for the Biot Problem