



MOX-Report No. 33/2025

## **A semiparametric space-time quantile regression model**

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

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

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

<https://mox.polimi.it>

# A semiparametric space-time quantile regression model

Ilenia Di Battista<sup>a‡</sup>, Marco F. De Sanctis<sup>a‡</sup>, Eleonora Arnone<sup>b</sup>, Cristian Castiglione<sup>c</sup>, Alessandro Palumbo<sup>a</sup>, Laura M. Sangalli<sup>a\*</sup>

<sup>a</sup>MOX, Dipartimento di Matematica, Politecnico di Milano, Piazza Leonardo Da Vinci 32, Milano, 20133, Italy; <sup>b</sup> Dipartimento di Management, Università degli Studi di Torino, Corso Unione Sovietica 218 bis, Torino, 10134, Italy; <sup>c</sup> Bocconi Institute for Data Science and Analytics, Bocconi University, Via Röntgen 1, Milano, 20136, Italy

## ABSTRACT

Spatio-temporal data often exhibit non-Gaussian behaviour, heteroscedasticity and skewness. Such data are, for example, highly prevalent in environmental and ecological sciences. In this work, we propose a semiparametric model for space-time quantile regression. The estimation functional incorporates roughness penalties based on differential operators over both the spatial and temporal dimensions. We study the theoretical properties of the model, proving the consistency and asymptotic normality of the associated estimators. To evaluate the effectiveness of the proposed method, we conduct simulation studies, bench-marking it against state-of-the-art techniques. Finally, we apply the model to analyse the space-time evolution of nitrogen dioxide concentration in the Lombardy region (Italy). The analyses of this pollutant are of primary importance for informing policies aimed at improving air quality.

## KEYWORDS

Smoothing with roughness penalties, Spatio-temporal data, Functional data analysis

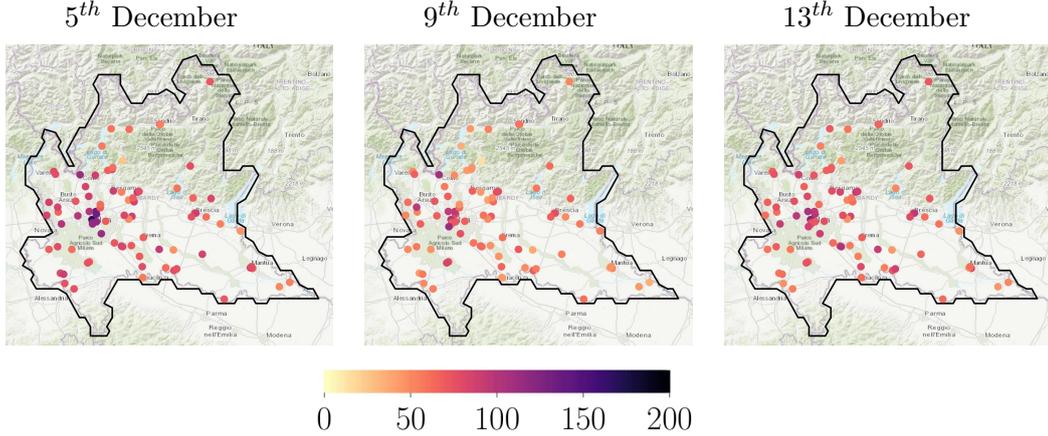
## 1. Introduction

In this work we focus on spatio-temporal data where the main interest is in analysing non-standard tail behaviours, rather than mean behaviours, of the phenomenon under study. Many real-world phenomena, especially in environmental and meteorological sciences, involve data characterised by heteroscedasticity, skewed distributions, and significant local variations over space and time. A key example is air pollutant concentration measurements, where the dispersion is strongly influenced by human factors, like population density and traffic, as well as geographical and climatic features, such as altitude, geographical conformation and air circulation, leading to complex spatial distributions, characterized by strong skewness and heteroscedasticity. Figure 1 presents an example of these data, showing Nitrogen Dioxide ( $\text{NO}_2$ ) concentration data, collected by the Agenzia Regionale per la Protezione dell'Ambiente (ARPA), across the Lombardy region of Italy, over three different days. The highest concentrations are observed in the metropolitan area of Milano, followed by Brescia and its surroundings, while the mountain regions have lower values, reflecting a complex spatial pattern. In the study of pollutant concentrations, experts do not primarily concentrate on the

---

<sup>‡</sup>These authors contributed equally to this work.

<sup>\*</sup>Corresponding author: Laura M. Sangalli. Email: [laura.sangalli@polimi.it](mailto:laura.sangalli@polimi.it)



**Figure 1.** Collected data at the ARPA sensors on the days 5-9-13 December 2018.

average behaviour, but rather on quantile levels, to capture extreme characteristics of the underlying process.

In today’s literature, several approaches have been proposed to investigate quantiles of complex spatio-temporal phenomena. Some of them, like the Bayesian models proposed by Reich (2012) and Das and Ghosal (2017), assume the quantile function to be linear with respect to the time dimension. Moreover, these methodologies employ separable space and time coefficients, without accounting for interaction terms. These simplifying assumptions, however, might be too restrictive for more complicated problems, such as the environmental example discussed above.

Sylvan et al. (2015) addresses scenarios involving complex space-time interactions, but only under the assumption that data are collected on space-time lattices—implying that observations are arranged in regularly spaced blocks, a condition rarely met in real-world applications. Other models, like the hierarchical Bayesian approach introduced by Castillo-Mateo et al. (2023), incorporate the temporal dependence through autoregressive functions, trying to capture the intertwined temporal relationships occurring at the specified quantile level.

Deb et al. (2024) introduces a nonparametric quantile model for space-time dependent data, modeling the spatial dependence by directly specifying the covariance matrix of the data generating process, which could be a limitation in scenarios where a parametric form of the covariance is not available. Furthermore, we point out that all these methods omit a discussion of the missing data problem. In addition to these methods, various space-time linear regression models, such as those proposed by Augustin et al. (2013) and Marra et al. (2012), can be extended to accommodate for the quantile analysis.

In this work, we propose a semiparametric regression model tailored for the quantile analysis of spatio-temporal data. The model belongs to the family of Spatial Regression models with Partial Differential Equation regularisation, reviewed by Sangalli (2021). This family of models permits to analyse complex spatial phenomena, being able to incorporate fundamental information about the spatial domain of interest and the physics of the phenomenon under study; see also, e.g., Sangalli et al. (2013) and Azzimonti, Sangalli, et al. (2015). In order to study the quantiles of space-time data, we extend the nonparametric model explored by Castiglione et al. (2025) for space-only data, introducing two roughness penalties, that promote smoothness of the process in

space and time. Moreover, we include a parametric term, to explore the dependence of the variable of interest on space-time varying covariates. The proposed model can appropriately account for missing data. This is an important modeling feature when dealing with space-time data. Indeed, data recorded by sensors, like meteorological and climate control units and environmental measuring stations, often exhibit missing entries, due to malfunctioning of the device or other specific conditions.

We first study the theoretical properties of the considered model, proving existence, consistency and asymptotic normality of the corresponding spatio-temporal semiparametric quantile estimators. We then illustrate by simulation studies the advantages of the proposed model over state-of-the-art competitors. Finally, we apply the model to the analysis of daily concentrations of NO<sub>2</sub> across the Lombardy region.

The article is organised as follows. Section 2 introduces the proposed semiparametric model for space-time quantile regression. Section 3 presents the estimation procedure. Section 4 examines the asymptotic properties of the detailed estimators. Section 5 reports three simulation studies that illustrate the performances of the presented method. Section 6 comments on the application to the study of the NO<sub>2</sub> concentrations over the Lombardy region. Finally, Section 7 discusses possible future developments.

## 2. Mathematical framework

Let  $\{\mathbf{p}_i\}_{i=1,\dots,n}$  be a set of  $n$  spatial locations over a bounded spatial domain  $\mathcal{D} \subset \mathbb{R}^2$ , and  $\{t_j\}_{j=1,\dots,m}$  be a set of  $m$  time points in a time interval  $[0, T] \subset \mathbb{R}$ . Let  $\{y_{ij}\}_{i=1,\dots,n, j=1,\dots,m}$  be the realisations of a real random variable  $Y_{ij}$ , measured at the space-time location  $(\mathbf{p}_i, t_j)$ . Let  $\mathbf{x}_{ij}$  be a vector of  $q$  space-time varying covariates, also measured at the spatio-temporal location  $(\mathbf{p}_i, t_j)$ . Assume that  $Y_{ij}$  has absolutely continuous distribution with unknown conditional probability density function  $\pi_{Y_{ij}|\mathbf{x}_{ij}, \mathbf{p}_i, t_j}$ , cumulative distribution function  $\Pi_{Y_{ij}|\mathbf{x}_{ij}, \mathbf{p}_i, t_j}$  and quantile function  $Q_{Y_{ij}|\mathbf{x}_{ij}, \mathbf{p}_i, t_j}(\alpha) = \inf\{y \in \mathbb{R} : \Pi_{Y_{ij}|\mathbf{x}_{ij}, \mathbf{p}_i, t_j}(y) \geq \alpha\}$  for any probability level  $\alpha \in (0, 1)$ .

We define the following semiparametric spatio-temporal model for the  $\alpha$ -quantile of  $Y_{ij}$

$$Q_{Y_{ij}|\mathbf{x}_{ij}, \mathbf{p}_i, t_j}(\alpha) = \mathbf{x}_{ij}^\top \boldsymbol{\beta}_\alpha + f_\alpha(\mathbf{p}_i, t_j), \quad i = 1, \dots, n, \quad j = 1, \dots, m, \quad (1)$$

where  $f_\alpha : \mathcal{D} \times [0, T] \rightarrow \mathbb{R}$  is the unknown field describing the non-linear spatio-temporal relationship of the  $\alpha$ -quantile function with respect to the space-time coordinates  $(\mathbf{p}, t)$ , and  $\boldsymbol{\beta}_\alpha \in \mathbb{R}^q$  is the unknown vector of regression coefficients indicating the effect of the covariates on the  $\alpha$ -quantile of the response variable.

Since missing data are highly prevalent in spatio-temporal real-world applications, it is crucial to carefully take them into account when modelling data. To this end, we first define the set of indices of observed data as

$$O = \{(i, j) : y_{ij} \text{ is observed}, \quad i = 1, \dots, n, \quad j = 1, \dots, m \},$$

and we denote by  $|O|$  its cardinality. We then propose to estimate the unknown vector  $\boldsymbol{\beta}_\alpha$  and field  $f_\alpha$  by minimising the following penalised functional

$$J_\alpha(f, \boldsymbol{\beta}) = \frac{1}{|O|} \sum_{(i,j) \in O} \rho_\alpha(y_{ij} - \mathbf{x}_{ij}^\top \boldsymbol{\beta} - f(\mathbf{p}_i, t_j)) + P(f), \quad (2)$$

where the first term involves the so-called pinball loss  $\rho_\alpha(x) := \frac{1}{2}|x| + (\alpha - \frac{1}{2})x$ , a data fidelity criterion tailored for quantile regression problems (see, e.g., Koenker and Bassett [1978]), while the second term is a roughness penalty functional quantifying the smoothness of the spatio-temporal field  $f$ . Here, in particular, in analogy with the space-time linear regression models considered, e.g., by Marra et al. (2012) and Augustin et al. (2013), we consider the penalty term

$$P(f) = \frac{\lambda_D}{2} \int_0^T \int_{\mathcal{D}} (\Delta f)^2 d\mathbf{p} dt + \frac{\lambda_T}{2} \int_0^T \int_{\mathcal{D}} \left( \frac{\partial^2 f}{\partial t^2} \right)^2 d\mathbf{p} dt, \quad (3)$$

where  $\lambda_D > 0$  and  $\lambda_T > 0$  are two positive smoothing parameters, weighting the penalty in space and time respectively, and  $\Delta f$  is the Laplacian operator, defined as  $\Delta f = \partial^2 f / dp_1^2 + \partial^2 f / dp_2^2$ . Note that the employment of the Laplacian operator corresponds to isotropic regularisation in space. However, the model could be extended to penalty terms based on a Partial Differential Equation, encoding problem-specific information about the underlying phenomenon, in analogy to the space-only case explored in Castiglione et al. (2025).

### 2.1. Functional embedding of the estimation problem

We now define the proper functional embedding space in which the problem is formulated, in order to state its properties. Denote by  $H^d(\mathcal{D})$  the Sobolev space of order  $d$ , i.e., the space of functions in  $L^2(\mathcal{D})$  having  $d$  weak derivatives in  $L^2(\mathcal{D})$ , and by  $\|v\|_{H^d(\mathcal{D})}^2$  its norm. Formally,

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

$$\|v\|_{H^d(\mathcal{D})}^2 = \int_{\mathcal{D}} v^2 + \int_{\mathcal{D}} \sum_{1 \leq |k| \leq d} |D^k v|^2 = \int_{\mathcal{D}} \sum_{0 \leq |k| \leq d} |D^k v|^2,$$

where  $D^k f$  is the  $k$ -th weak partial derivative of  $f$ . Let  $L^2(0, T; H^2(\mathcal{D}))$  be the space of square integrable functions from the interval  $[0, T]$  into the Hilbert space  $H^2(\mathcal{D})$ , and let  $L^2(0, T; L^2(\mathcal{D}))$  be the analogous space where the functions take value in  $L^2(\mathcal{D})$ . We define the functional space

$$V = \left\{ f \in L^2(0, T; H^2(\mathcal{D})) : \frac{\partial^2 f}{\partial t^2} \in L^2(0, T; L^2(\mathcal{D})), \nabla f \cdot \nu = 0 \text{ on } \partial\mathcal{D} \times (0, T] \right\},$$

where  $\nu$  is the outward unit normal vector to  $\partial\mathcal{D}$ , and the corresponding norm

$$\|f\|_V^2 = \int_0^T \|f(t)\|_{H^2(\mathcal{D})}^2 dt + \int_0^T \left\| \frac{\partial^2 f(t)}{\partial t^2} \right\|_{L^2(\mathcal{D})}^2 dt.$$

Note that, in the space  $V$  we are considering homogeneous Neumann boundary conditions (i.e.,  $\nabla f \cdot \nu = 0$  on  $\partial\mathcal{D} \times (0, T]$ ), that are the natural boundary conditions for this kind of problems; see, e.g., Arnone, Sangalli, et al. (2023). However, different boundary conditions could also be considered, including Dirichlet, Robin and mixed boundary conditions; we refer to the work of Azzimonti, Sangalli, et al. (2015) for a detailed treatment in the linear regression setting.

## 2.2. Characterization of the estimation problem

We aim to find the semiparametric estimator  $(\hat{f}_\alpha, \hat{\beta}_\alpha) \in V \times \mathbb{R}^q$  such that

$$(\hat{f}_\alpha, \hat{\beta}_\alpha) = \arg \min_{(f, \beta) \in V \times \mathbb{R}^q} J_\alpha(f, \beta) \quad (4)$$

where  $J_\alpha(f, \beta)$  is the functional in equation (2). We denote by  $\mathcal{V}_\alpha$  the space of solutions to the estimation problem (4), i.e.,  $\mathcal{V}_\alpha = \{(\hat{f}_\alpha, \hat{\beta}_\alpha) \in V \times \mathbb{R}^q : J_\alpha(\hat{f}, \hat{\beta}) = \inf_{(f, \beta) \in V \times \mathbb{R}^q} J_\alpha(f, \beta)\}$ . The following proposition states the existence of a solution to this estimation problem.

**Proposition 2.1.** *The space  $\mathcal{V}_\alpha$  is a non-empty, closed, convex set.*

The proof is deferred to the Appendix A.1.

In contrast with the linear and generalised linear models, discussed, e.g., in Sangalli et al. (2013) and Wilhelm and Sangalli (2016), the proposed quantile regression model does not possess a unique solution. See also Castiglione et al. (2025) for a discussion in the space-only case. However, we can state the following characterisation.

**Proposition 2.2.** *Any solution  $(\hat{f}_\alpha, \hat{\beta}_\alpha) \in \mathcal{V}_\alpha$  is a global minimiser of functional (2).*

The proof is deferred to the Appendix A.1. It should be noted that, even if the estimator  $(\hat{f}_\alpha, \hat{\beta}_\alpha)$  exists and is a global minimum, an explicit solution is not available in closed form. We thus need to proceed via iterative numerical optimization, how detailed in the following section.

## 3. Estimation procedure

### 3.1. Functional EM algorithm

Since the estimation problem does not admit a closed-form solution, we approximate it by using of the Expectation-Maximisation (EM) algorithm (see, e.g., Dempster et al. [1977]). The latter is an iterative algorithm characterised by a monotonic convergence property ensuring that, in the limit, it converges to a solution  $(\hat{f}_\alpha, \hat{\beta}_\alpha) \in \mathcal{V}_\alpha$ . Specifically, such algorithm can be used owing to the result proved in Yu and Moyeed (2001), which states that minimising the pinball loss function is equivalent to maximising the likelihood of an Asymmetric Laplace model.

Let  $\mathbf{y}$  denote the vector collecting the observed data,  $X$  the matrix collecting the covariate vectors,  $\mathbf{z}^{(k)} = \mathbf{y} - (1 - 2\alpha)|\mathbf{y} - X\beta_\alpha^{(k)} - \mathbf{f}_{\alpha, n}^{(k)}|$  the vector of pseudo-observations,  $W^{(k)} = \text{diag}(\mathbf{w}^{(k)})$  a working weight matrix with  $1/\mathbf{w}^{(k)} = 2|O||\mathbf{y} - X\beta_\alpha^{(k)} - \mathbf{f}_{\alpha, n}^{(k)}|$ , and finally  $\mathbf{f}_n$  the vector of evaluations of the spatio-temporal field  $f$  at the observed spatio-temporal locations, i.e.  $\mathbf{f}_n = \{f(\mathbf{p}_i, t_j)\}_{(i,j) \in O}$ , as detailed in Section 3.2. Then, as shown in Appendix A.3, exploiting the Gaussian-Exponential stochastic representation of the Asymmetric Laplace likelihood, and following the EM strategy, at each iteration

of the algorithm we minimise a functional  $J_\alpha^{(k)}(f, \boldsymbol{\beta})$ , given by

$$J_\alpha^{(k)}(f, \boldsymbol{\beta}) = (\mathbf{z}^{(k)} - X\boldsymbol{\beta} - \mathbf{f}_n)^\top W^{(k)}(\mathbf{z}^{(k)} - X\boldsymbol{\beta} - \mathbf{f}_n) + P(f). \quad (5)$$

The latter functional  $J_\alpha^{(k)}$  plays the role of a local quadratic approximation of the functional  $J_\alpha(f, \boldsymbol{\beta})$  in a neighborhood of  $(f_\alpha^{(k)}, \boldsymbol{\beta}_\alpha^{(k)})$ , namely the solution at the  $k$ -th iteration of the algorithm.

Owing to the quadratic nature of  $J_\alpha^{(k)}(f, \boldsymbol{\beta})$ , the EM algorithm can be interpreted as a Functional Penalised Iterative Reweighted Least Squares (FPIRLS), as discussed in Wilhelm and Sangalli (2016) in a generalised linear setting. Specifically, at each iteration  $k$ , the minimisation problem can be characterized via a convenient mean variational form, alike in Arnone, Sangalli, et al. (2023), by appropriately considering the weight matrix  $W$  and the vector of pseudo-observations  $\mathbf{z}$ . Moreover, this characterization allows to infer the method's computational cost from the cost of the linear regression model discussed in Arnone, De Falco, et al. (2023), as the latter constitutes a single iteration of our iterative scheme.

It should be noted that the minimisation of functional (5), for  $f \in V$  and  $\boldsymbol{\beta} \in \mathbb{R}^q$ , is an infinite-dimensional problem, that does not admit a closed form solution. In the next section, we introduce a convenient finite-dimensional discretisation of this estimation problem.

### 3.2. From an infinite to a finite-dimensional problem

To discretise the functional (5), we rely on a time-by-space tensorisation based on a convenient space-time basis system. In particular, in the spatial dimension we consider a set of  $N$  piece-wise linear polynomial basis functions defined over a regular triangulation  $\mathcal{D}_\tau$  of the domain  $\mathcal{D}$ , named finite elements, and we indicate it by  $\{\psi_1(\mathbf{p}), \dots, \psi_N(\mathbf{p})\}$ . This choice is particularly convenient when the phenomenon manifests a strongly localised skewed signal, or is observed on a non-convex domain, situations in which classical tensorised basis fail to obtain good approximations. For the temporal dimension, instead, we consider a set  $\{\varphi_1(t), \dots, \varphi_M(t)\}$  of  $M$  cubic B-spline basis functions over the interval  $[0, T]$ . We highlight that the selection of  $N$  and  $M$ , which determine the resolution of the space-time discretisation, involves a trade-off between accuracy and computational efficiency. The mesh must be sufficiently refined to accurately capture local features, while avoiding excessive computational costs, as discussed in Azzimonti, Nobile, et al. (2014).

Let  $\Psi = \{\psi_h(\mathbf{p}_i)\}_{i,h} \in \mathbb{R}^{n \times N}$  be the matrix of evaluations of the  $N$  finite element basis on the  $n$  spatial locations  $\{\mathbf{p}_1, \dots, \mathbf{p}_n\}$  and  $\Phi = \{\phi_k(t_j)\}_{j,k} \in \mathbb{R}^{m \times M}$  the matrix of evaluations of the  $M$  temporal basis on the  $m$  time instants  $\{t_1, \dots, t_m\}$ . Thus,  $f_\alpha$  can be expanded as  $f_\alpha(\mathbf{p}, t) = \sum_{h=1}^N \sum_{k=1}^M f_{hk} \psi_h(\mathbf{p}) \varphi_k(t)$ . Using the Kronecker product  $\Phi \otimes \Psi$ , and the vector of coefficients  $\mathbf{f}_\alpha \in \mathbb{R}^{NM}$ , we can write  $\mathbf{f}_{\alpha,n} = (\Phi \otimes \Psi) \mathbf{f}_\alpha$ . To account for the missingness structure of the data, we define  $A$  as the matrix obtained from  $\Phi \otimes \Psi$  where the  $(i+nj)$ -th row is removed if the datum at  $(\mathbf{p}_i, t_j)$  is not available.

We can thus write the discretised version of the functional in (5) as

$$J_\alpha^{(k)}(\boldsymbol{\beta}_\alpha, f_\alpha) = (\mathbf{z}^{(k)} - X\boldsymbol{\beta}_\alpha - A\mathbf{f}_\alpha)^\top W^{(k)}(\mathbf{z}^{(k)} - X\boldsymbol{\beta}_\alpha - A\mathbf{f}_\alpha) + \mathbf{f}_\alpha^\top P\mathbf{f}_\alpha, \quad (6)$$

where  $P$  is the overall penalty matrix, whose expression is detailed in Appendix A.4. Following analogous derivations as those in Arnone, Sangalli, et al. (2023), set  $H^{(k)} =$

$X(X^T W^{(k)} X)^{-1} X^T W$  and  $Q^{(k)} = W^{(k)}(I_{|O|} - H)$ , where  $I_{|O|}$  is the identity matrix of dimension  $|O|$ . It can be proved that, the estimator  $(\tilde{\mathbf{f}}_\alpha, \tilde{\boldsymbol{\beta}}_\alpha) \in \mathbb{R}^{NM} \times \mathbb{R}^q$  minimising the discretised functional in (6), at each iteration of the FPIRLS algorithm, exists unique and has closed-form solution

$$\begin{aligned}\tilde{\mathbf{f}}_\alpha &= T^{-1} A^\top Q^{(k)} \mathbf{z}^{(k)}, \\ \tilde{\boldsymbol{\beta}}_\alpha &= \left( X^\top W^{(k)} X \right)^{-1} X^\top W^{(k)} (\mathbf{z}^{(k)} - A \tilde{\mathbf{f}}_\alpha),\end{aligned}\tag{7}$$

where  $T = A^\top Q A + P$ .

It is worth to notice that, thanks to the monotonicity property of the EM algorithm (see, e.g., Lange [2016]), for  $k$  large enough, the finite dimensional solution  $(\tilde{\mathbf{f}}_\alpha, \tilde{\boldsymbol{\beta}}_\alpha)$  approximately reaches its steady-state at  $(\hat{\mathbf{f}}_\alpha, \hat{\boldsymbol{\beta}}_\alpha)$ , the true minimum of  $J_\alpha$ .

### 3.3. Modified generalised cross validation criterion

The smoothing parameters  $\lambda_D$  and  $\lambda_T$  weight the amount of regularity we impose on the solution  $\hat{\mathbf{f}}_\alpha$  in the spatial and temporal dimensions, respectively. Their appropriate choice is crucial to obtain accurate estimates and avoid overfitting issues. In the space-only framework explored in Castiglione et al. (2025), the single smoothing parameter  $\lambda$  is selected through the minimisation of an approximated Generalised Cross-Validation (GCV) score, defined via the pinball loss function  $\rho_\alpha$ . Nevertheless, such procedure suffers of instability in more complex scenarios, like the spatio-temporal case, due to the non-differentiability of such function and the complex interaction between space and time. This issue is well-known in literature, indeed many works have proposed modified expressions of the pinball loss, to regularize its non-differentiable point, and obtain more stable results; see, e.g., Nychka et al. (1995), Shin et al. (2022) and Nortier (2021). Specifically, we rely on the smoothed expression of the pinball loss proposed in Fasiolo et al. (2021a), i.e.,

$$\rho_{\alpha,\varepsilon}^*(x) = (\alpha - 1)x + \varepsilon \log(1 + \exp(\frac{x}{\varepsilon})),$$

where the parameter  $\varepsilon > 0$  regulates the amount of approximation enforced on the pinball loss, with higher values imposing a smoother approximation.

The smoothing parameters  $\lambda_D$  and  $\lambda_T$  are thus selected minimising the modified GCV score defined as

$$GCV(\lambda_D, \lambda_T) = \sum_{(i,j) \in O} \frac{\rho_{\alpha,\varepsilon}^*(y_{ij} - \mathbf{x}_{ij}^\top \hat{\boldsymbol{\beta}}_\alpha - \hat{f}_\alpha(\mathbf{p}_i, t_j))}{|O| - df},\tag{8}$$

where  $(\hat{f}_\alpha, \hat{\boldsymbol{\beta}}_\alpha)$  are the estimates obtained from the last step of the EM algorithm,  $df$  are the corresponding effective degrees of freedom of the model, computed as  $df = q + \text{tr}(S)$ , and  $S$  is the smoothing matrix defined as  $S = AT^{-1}A^\top Q$ . We notice that the GCV score in (8) implicitly depends on  $(\lambda_D, \lambda_T)$  through the parameter estimates and the effective degrees of freedom.

#### 4. Asymptotic properties

We now study the asymptotic properties of the estimators presented in (7). We here indicate by  $f_\alpha \in V$  the true  $\alpha$ -quantile field, by  $g_\alpha = \Delta f_\alpha$  the laplacian of  $f_\alpha$ , and by  $\mathbf{f}_\alpha$  and  $\mathbf{g}_\alpha$  the vectors of their evaluations at the knots of the basis expansion tensorised as explained in the previous section. Moreover, we denote by  $\boldsymbol{\beta}_\alpha$  the vector of coefficients expressing the true relation between the set of covariates and the  $\alpha$ -quantile. Furthermore, we define the completed design matrix  $C = [X, A] \in \mathbb{R}^{|O| \times (q+NM)}$ . For simplicity, we denote by  $\mathbf{c}_{ij}$  the row of the matrix  $C$  referring to the spatio-temporal location  $(\mathbf{p}_i, t_j)$ , and by  $\mathbf{a}_{ij}$  the corresponding row of the matrix  $A$ . We define the following  $NM \times NM$  matrices

$$D_{0,|O|} = \alpha(1 - \alpha) \frac{1}{|O|} \sum_{(i,j) \in O} \mathbf{c}_{ij} \mathbf{c}_{ij}^\top, \quad D_{1,|O|} = \frac{1}{|O|} \sum_{(i,j) \in O} \pi_{ij} \mathbf{c}_{ij} \mathbf{c}_{ij}^\top,$$

where the weights  $\pi_{ij} = \pi_{Y_{ij}|\mathbf{x}_{ij}, \mathbf{p}_i, t_j}(Q_{Y_{ij}|\mathbf{x}_{ij}, \mathbf{p}_i, t_j}(\alpha))$  for  $(i, j) \in O$  depend on the true conditional distribution of  $Y_{ij}|\mathbf{x}_{ij}, \mathbf{p}_i, t_j$ . Let  $\tilde{R}_1$  be the  $(q + NM) \times (q + NM)$  matrix having  $\tilde{R}_1$  in the south-east block and all zeros elsewhere. Similarly, we define the  $(q + NM)$  vectors  $\tilde{\mathbf{g}}_\alpha = (\mathbf{0}_q, \mathbf{g}_\alpha)^\top$  and  $\tilde{\mathbf{h}}_\alpha = (\mathbf{0}_q, \mathbf{h}_\alpha)^\top$ , where  $\mathbf{h}_\alpha = (P_t \otimes R_0)\mathbf{f}_\alpha$ . For the explicit definitions of  $\tilde{R}_1$ ,  $\tilde{R}_1$  and  $(P_t \otimes R_0)$ , see the Appendix A.4.

We now present a set of sufficient regularity conditions that ensure the consistency and asymptotic normality of the discretised  $\alpha$ -quantile estimator, under the assumption that the censoring mechanism is independent of the value of the field  $f$  to be estimated.

**Assumption 1.** *There exist two positive constants  $\pi_-$  and  $\pi_+$ , bounded away from 0 and  $\infty$ , such that  $\pi_- < \pi_{ij} < \pi_+$ , for any  $(\mathbf{p}_i, t_j) \in O$ .*

**Assumption 2.** *The discretisation matrix  $A$  and the covariate matrix  $X$  are full-rank matrices. Moreover,  $X$  is such that  $\|X\|_\infty < +\infty$ , where we use  $\|X\|_\infty = \max_{k,l} |x_{kl}|$ .*

**Assumption 3.** *There exist two positive definite matrices  $D_0$  and  $D_1$  such that  $D_{0,|O|} \rightarrow D_0$  and  $D_{1,|O|} \rightarrow D_1$  as  $|O| \rightarrow \infty$ .*

Such assumptions generalise to the semiparametric spatio-temporal setting here tackled those considered by Castiglione et al. (2025) in the simpler nonparametric space-only case, as well as in other discussions of the asymptotic analysis of quantile estimators found in the literature, see, e.g., Koenker (2005). Specifically, Assumption 1 together with Assumption 2, ensures the non-singularity of the matrices  $D_{0,|O|}$  and  $D_{1,|O|}$ . Finally, the existence of a non-singular limit for the matrices  $D_{0,|O|}$  and  $D_{1,|O|}$  is assured by Assumption 3.

Under Assumptions 1, 2 and 3 we can now state the consistency of the finite-dimensional  $\alpha$ -quantile estimator. The proof is deferred to Appendix A.5.

**Proposition 4.1.** *Let  $\hat{\boldsymbol{\theta}}_\alpha = (\hat{\boldsymbol{\beta}}_\alpha, \hat{\mathbf{f}}_\alpha)$  be the  $\alpha$ -quantile estimator. Under Assumptions 1, 2, 3 and assuming that  $\lambda_D \sqrt{|O|} \rightarrow \bar{\lambda}_D$  and  $\lambda_T \sqrt{|O|} \rightarrow \bar{\lambda}_T$ , for some finite values  $\bar{\lambda}_D, \bar{\lambda}_T$ , then  $\hat{\boldsymbol{\theta}}_\alpha$  is a consistent estimator for  $\boldsymbol{\theta}_\alpha = (\boldsymbol{\beta}_\alpha, \mathbf{f}_\alpha)$ , such that  $\|\hat{\boldsymbol{\theta}}_\alpha - \boldsymbol{\theta}_\alpha\| =$*

$O_p(|O|^{-1/2})$ . Moreover, under the same assumptions, the  $\alpha$ -quantile estimator  $\hat{\boldsymbol{\theta}}_\alpha$  has the following asymptotic distribution

$$\sqrt{|O|}(\hat{\boldsymbol{\theta}}_\alpha - \boldsymbol{\theta}_\alpha) \xrightarrow{d} N_{q+NM}(\boldsymbol{\mu}, \boldsymbol{\Sigma}),$$

with  $\boldsymbol{\mu} = \bar{\lambda}_D D_1^{-1} \tilde{R}_1^\top \tilde{\mathbf{g}}_\alpha - \bar{\lambda}_T D_1^{-1} \tilde{\mathbf{h}}_\alpha$  and  $\boldsymbol{\Sigma} = D_1^{-1} D_0 D_1^{-1}$ . Moreover, if  $\lambda_D = o(|O|^{-1/2})$  and  $\lambda_T = o(|O|^{-1/2})$  then the estimator  $\hat{\boldsymbol{\theta}}_\alpha$  is asymptotically unbiased.

## 5. Simulation studies

In this section we discuss three simulation studies showing the performance of the proposed method, named Spatio-Temporal Quantile Regression with Partial Differential Equation regularisation (STQR-PDE). Specifically, Section 5.1 details the data generation process, Section 5.2 presents competing techniques, and Section 5.3 discusses the results obtained.

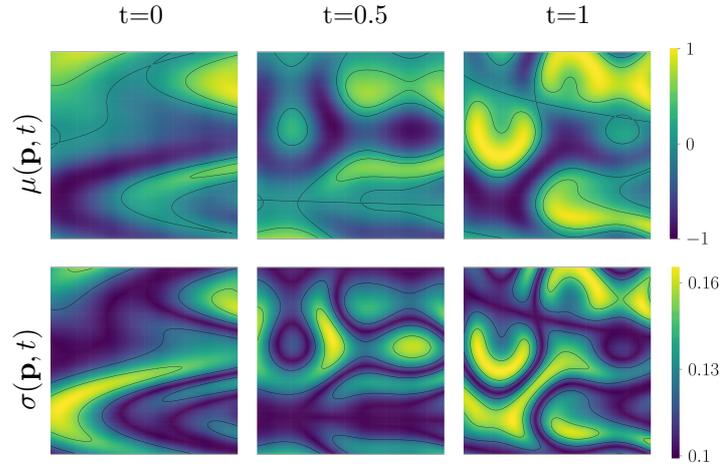
### 5.1. Data generation

We consider heteroscedastic data generated according to the model

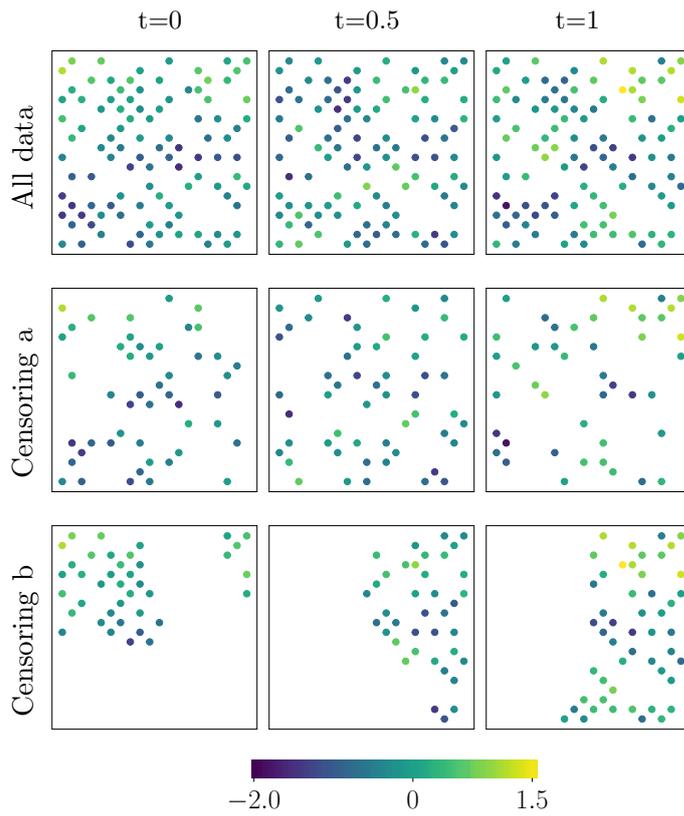
$$y_{ij} = \mathbf{x}_{ij}^\top \boldsymbol{\beta} + \mu(\mathbf{p}_i, t_j) + \varepsilon(\mathbf{p}_i, t_j), \quad (9)$$

where  $\varepsilon(\mathbf{p}_i, t_j) \sim N(0, \sigma^2(\mathbf{p}, t))$ , for  $\mathbf{p}_i \in [0, 1] \times [0, 1]$  and  $t_j \in [0, 1]$ . The function  $\mu(\mathbf{p}, t)$  is shown in the top panel of Figure 2, and its analytical expression is reported in Appendix B, while the spatio-temporal standard deviation surface  $\sigma^2(\mathbf{p}, t)$  is displayed in the bottom panel of Figure 2, and is defined as  $\sigma(\mathbf{p}, t) = 0.1e^{0.5|\mu(\mathbf{p}, t)|}$ . The considered nonparametric term  $\mu(\mathbf{p}, t)$  exhibits strong localised features and rapid variations, both in the space and time dimension, making it very challenging to capture its behaviour. We consider two independent random covariates: the first with normal distribution with mean 0 and variance 0.25; the second one with exponential distribution with mean 20. The corresponding unknown coefficients are  $\beta_1 = 1$  and  $\beta_2 = 2$ . We consider  $n = 100$  spatial locations, randomly sampled from a  $[0, 1] \times [0, 1]$  regular grid, and an equally spaced grid of 21 sampling time points across the time interval  $[0, 1]$ . Building on this setup, we examine three simulation scenarios. The first scenario uses the complete dataset, while the other two incorporate missing data to assess the robustness of our method under incomplete observations. Specifically, we implement two censoring schemes: scheme *a*, which assumes missingness is independent in both time and space, and scheme *b*, where missingness is dependent on both dimensions. The latter is inherently more challenging, as it results in entire regions of the spatio-temporal domain being unobserved. A detailed discussion of these censoring schemes is provided in Arnone, Sangalli, et al. (2023). For both missing-data scenarios, we consider a censoring proportion of  $p = 50\%$ , which is substantial given the relatively small size of the dataset. Figure 3 illustrates the spatial distribution of the data at three selected time instants across the different simulation scenarios. On the other hand, Figure 4 presents the temporal evolution of the data, comparing the different scenarios. These visualisations clearly show as censoring scheme *a* results in missing values that are randomly distributed in both space and time, whereas censoring scheme *b* produces a structured pattern of missingness, leading to entire unobserved regions in

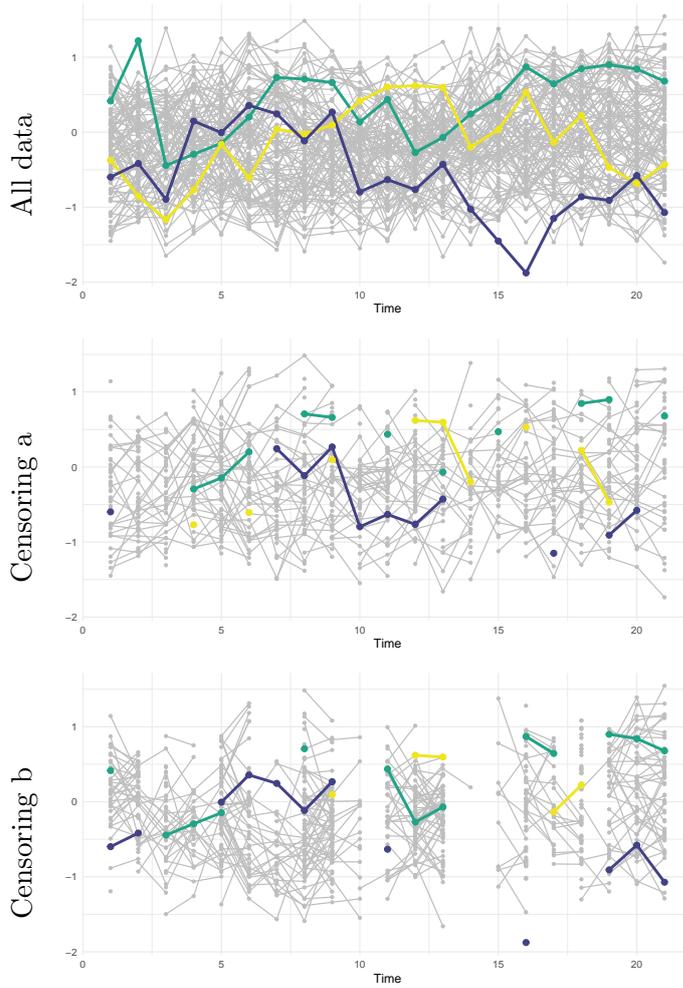
the spatio-temporal domain. Finally, we repeat each data generation 50 times, and we analyse three different levels of  $\alpha$ , namely  $\alpha = 10\%$ ,  $50\%$ ,  $90\%$ .



**Figure 2.** Mean function  $\mu(\mathbf{p}, t)$  (top panel) and standard deviation function  $\sigma(\mathbf{p}, t)$  (bottom panel), displayed at three time instants.



**Figure 3.** Spatial distribution of the simulated data. Each column corresponds to different time instants and each row corresponds to different simulation scenarios.



**Figure 4.** Temporal trend of the simulated data in the three different simulation scenarios. The figure shows the temporal profiles of all spatial locations, where observations at consecutive time points are connected by lines. Three time trends, corresponding to three random locations, are highlighted in colours.

## 5.2. Competing methods

We compare the proposed STQR-PDE model with alternative methods from the literature. Specifically, we consider the quantile version of the spatio-temporal model presented in Augustin et al. (2013) and in Marra et al. (2012). Alike the proposed STQR-PDE model, these techniques adopt a cubic B-spline basis for the discretisation in time. For the space discretisation, instead, following the discussion in Augustin et al. (2013), we rely alternatively on the soap film smoother, proposed by Wood et al. (2008), or on the thin plate spline basis proposed in Wahba (1990). We refer to these two alternative techniques with the acronyms SOAP and TPS, respectively.

We also explored modified versions of these models, incorporating a location-scale calibration term to address the heteroscedasticity of the phenomenon, as described in Fasiolo et al. (2021a). However, these model variants did not show any improvement over the corresponding traditional methods in this setting, and as such, we decided to exclude them from the discussion in Section 5.3. The competing techniques are

implemented through the R function `qgam` of the `qgam` package, developed by Fasiolo et al. (2021b). Instead, STQR-PDE is implemented in the `fdapde` library available in the GitHub repository (Palumbo et al. [2025]). All the methods rely on the GCV procedure to properly select the smoothing parameters. Specifically, STQR-PDE employs the modified GCV criterion discussed in Section 3.3. Unfortunately, the other methods cited in Introduction lack publicly available code and are therefore also omitted from this comparison study.

The competing methods are compared in terms of Root Mean Squared Error (RMSE), computed as

$$RMSE(\hat{\beta}_i) = |\hat{\beta}_i - \beta_i|, \quad RMSE(\hat{f}_\alpha) = \sqrt{\int_0^T \int_{\mathcal{D}} (f_\alpha(\mathbf{p}, t) - \hat{f}_\alpha(\mathbf{p}, t))^2 d\mathbf{p} dt}$$

for the parametric and the nonparametric term, respectively, where  $f_\alpha$  denotes the nonparametric contribution of the  $\alpha$ -quantile, and the corresponding RMSE is computed numerically on a fine grid over the spatio-temporal domain. We also compute the RMSE for the overall semiparametric  $\alpha$ -quantile estimate, as

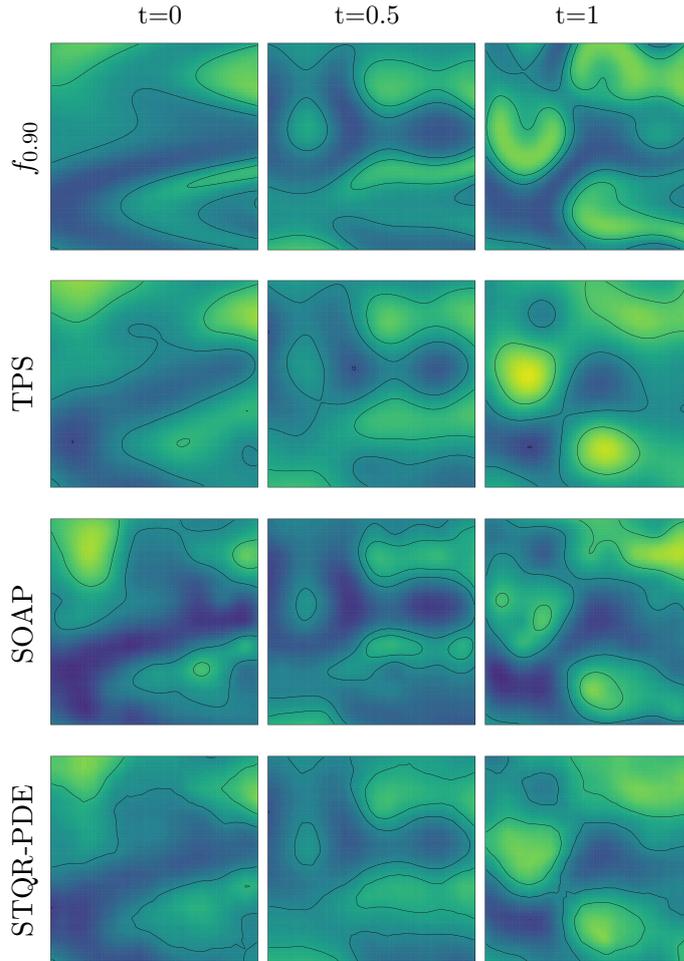
$$RMSE(\hat{Q}_\alpha) = \sqrt{\frac{\sum_{i=1}^n \sum_{j=1}^m (Q_\alpha(\mathbf{p}_i, t_j) - (\hat{Q}_\alpha(\mathbf{p}_i, t_j)))^2}{nm}}$$

over the set of spatio-temporal locations.

### 5.3. Results

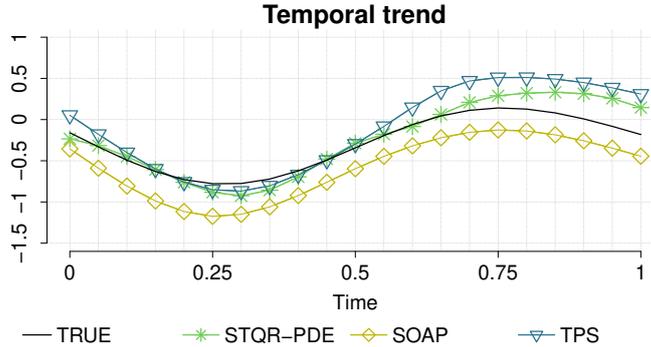
In this section we discuss the results obtained in the three described simulation settings. Figure 5 shows the nonparametric contribution of the 90%-quantile, namely  $f_{0.90}$ , and the corresponding estimates provided by the compared techniques, at three time instants, in the case of complete observed data. It is worth noting that the true  $f_{0.90}$  closely resembles the function  $\mu$  depicted in Figure 2, as the standard deviation field is a function of the mean itself. A visual inspection of the figure does not highlight a clear best estimate. However, it is possible to notice that TPS produces oversmoothed isolines and tends to overshoot the higher picks in the true  $f_{0.90}$ . On the opposite, SOAP tends to underestimate the true  $f_{0.90}$ , as it is clear from the large dark region in the first time instant. Figure 6 shows the temporal evolution of  $f_{0.90}$  in a fixed spatial location in the center of the spatial domain, and the corresponding estimates by the competing methods, again highlighting that TPS tends to produce higher picks, while SOAP underestimate the true  $f_{0.90}$ .

In order to quantify the differences in performances between the methods, Figure 7 reports the RMSE obtained by TPS, SOAP and the proposed STQR-PDE, in the three simulation data settings. The top two rows, report the RMSE of  $\hat{Q}_\alpha$  and  $\hat{f}_\alpha$ . We see that STQR-PDE provides the best estimates in all the simulation scenarios and for all the analysed quantile levels. Indeed, the proposed method always achieves significantly lower values of RMSE, both for the overall estimate and for the estimate of the nonparametric term alone. Moreover, the RMSE ranges across the different simulation scenarios confirm that censoring scheme  $b$  is the most challenging, yielding higher RMSE values for all methods. Nevertheless, performance differences between the methods remain substantial, with results consistently favouring STQR-PDE. The



**Figure 5.** Nonparametric contribution of the 90%-quantile in the setting of complete observed data. Each column corresponds to different time instants. The first row shows the true  $f_{0.90}$  and the subsequent rows the estimates  $\hat{f}_{90}$  obtained by the competing methods: quantile regression based on a Thin-Plate-Spline basis in space (TPS); quantile regression based on a Soap film basis in space (SOAP), and the proposed Spatio-Temporal Quantile Regression with Partial Differential Equation regularisation (STQR-PDE).

bottom two rows of Figure 7 reports the RMSE of  $\hat{\beta}_1$  and  $\hat{\beta}_2$ , indicating mostly comparable estimates for the parametric terms across all the simulation settings (with a slight advantage for the TPS method for some levels, that is though associated with the worst overall RMSE by this technique).



**Figure 6.** Temporal trend of  $f_{0.90}$  in the setting of complete observed data at a fixed spatial point at the center of the spatial domain, and corresponding estimates provided by the competing methods. Same competing methods as in Figure 5.

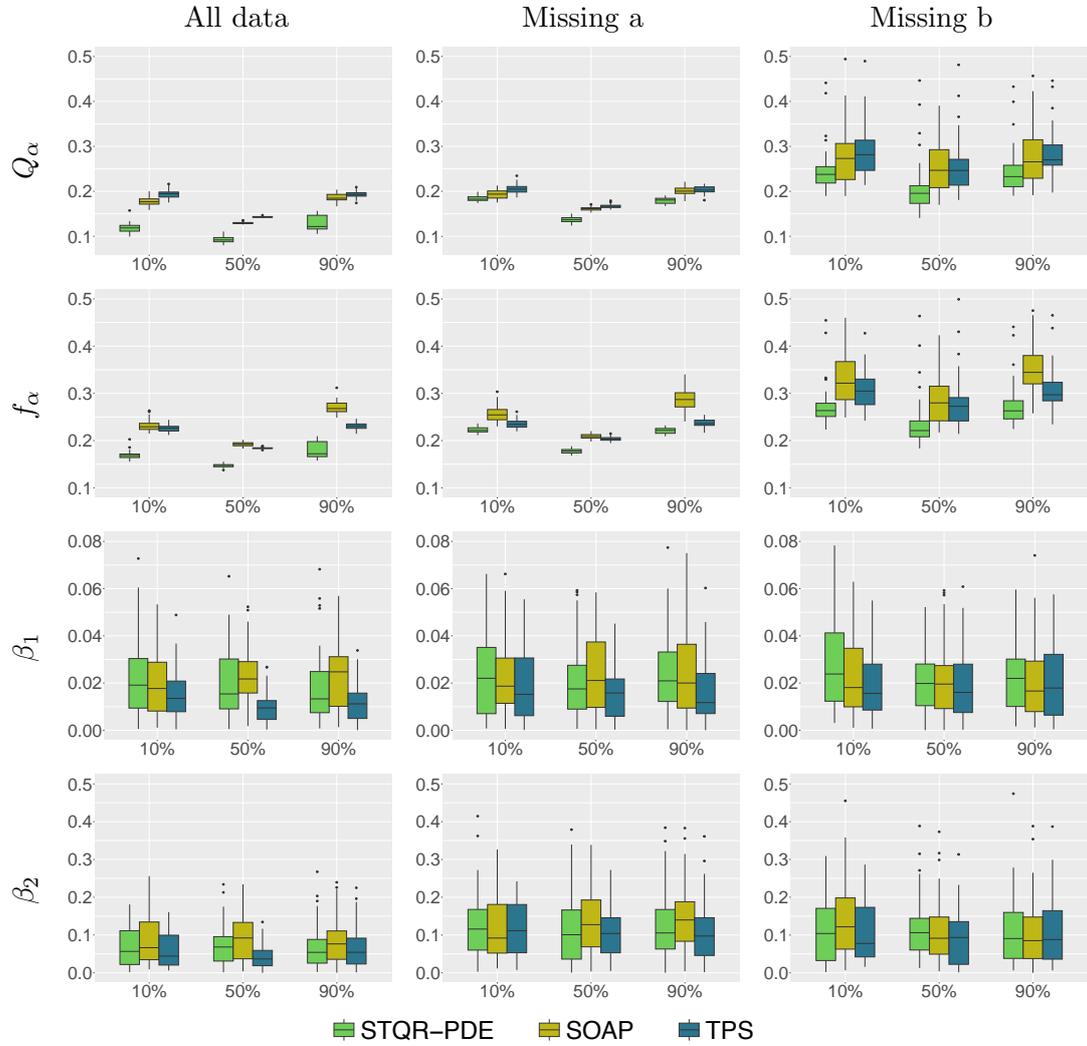
## 6. A semiparametric space-time quantile regression model for air quality data

In this section, we analyse the temporal evolution of the nitrogen dioxide ( $\text{NO}_2$ ) concentration across the Lombardy region, in the north of Italy, utilising the proposed semiparametric STQR-PDE model.  $\text{NO}_2$  is a highly toxic gas, primarily driven by human activities, such as traffic, energy production, domestic heating, and wastewater management; see, e.g., Salama et al. (2022). Additionally,  $\text{NO}_2$  is a short-lived pollutant with rapidly varying concentrations, making the analysis of temporal trends essential to understand how pollutant levels may vary across different days, like weekends and public holidays, when work-related traffic diminishes, and people tend to leave large urban areas. This analysis can reveal critical dynamics, and inform air quality policies aimed at protecting public health, as  $\text{NO}_2$  exposure is linked to severe respiratory issues and increased risk of premature death; see, e.g., California Air Resources Board (2023).

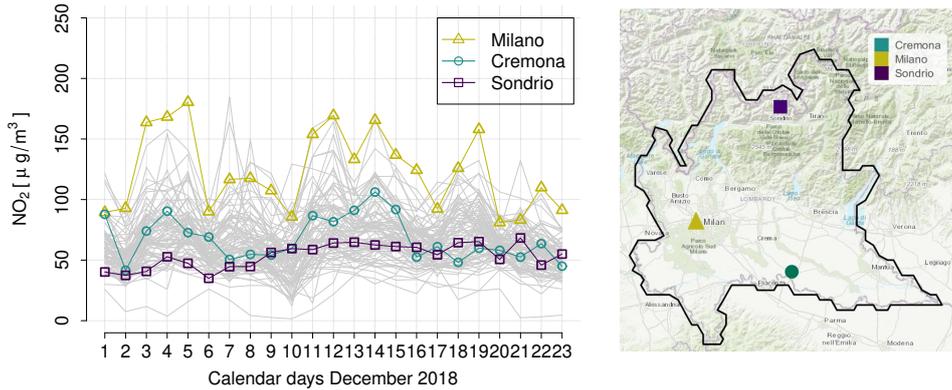
In this study, we specifically focus on  $\text{NO}_2$  levels in Lombardy, one of the most critical regions for air quality in Italy and Europe, due to limited air circulation, dense industrial activity, and extensive urbanisation. For example, the overpopulation in the metropolitan area of Milan leads to poor air quality primarily driven by emissions from traffic and domestic heating. Therefore, monitoring  $\text{NO}_2$  concentrations in Lombardy is crucial to prevent dangerous levels of this toxic substance.

### 6.1. Description

In this study, we analyse  $\text{NO}_2$  data collected during the winter period 1<sup>st</sup>–23<sup>th</sup> December 2018, by 84 monitoring stations managed by Agenzia Regionale per la Protezione dell’Ambiente (ARPA), across the Lombardy region. The data are available in *Regione Lombardia* (Open Data, Regione Lombardia [2024]). To investigate the temporal trend of the pollutant, and to focus on the most concerning conditions, we aggregate the daily data at each spatial location, taking the highest measurement. Figure 1 illustrates these data over three days in the considered period, while Figure 8 displays the



**Figure 7.** RMSEs for  $\alpha = 10\%, 50\%, 90\%$  obtained by the competing methods, across the three simulation settings. Top row: RMSE of  $\hat{Q}_\alpha$ ; second row: RMSE of  $\hat{f}_\alpha$ ; third row: RMSE of  $\hat{\beta}_1$ ; Bottom row: RMSE of  $\hat{\beta}_2$ . Same competing methods as in Figure 5. For visualisation purposes, we removed a few outliers produced by TPS and SOAP under the missing data cenario *b*, specifically for the RMSE of  $\hat{Q}_\alpha$  and  $\hat{f}_\alpha$ .



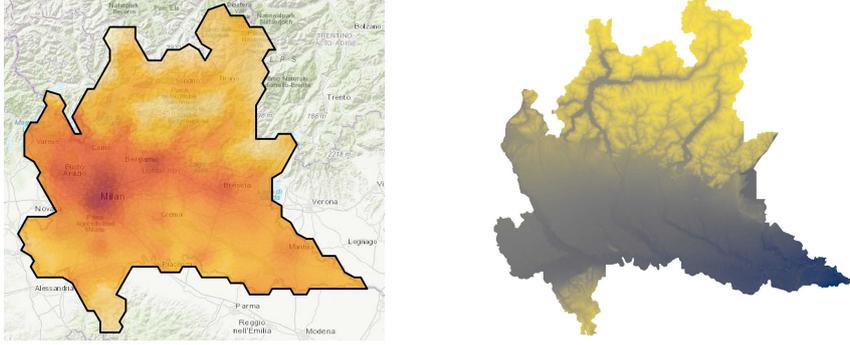
**Figure 8.** Temporal trend of the  $\text{NO}_2$  measurements in the three cities: Milano (dark yellow), Cremona (dark cyan) and Sondrio (dark purple).

temporal trend for each monitoring station. The profiles highlighted in color correspond to three monitoring stations, representative of the heterogeneity of Lombardy’s territory, namely the cities of Milano, Cremona and Sondrio: Milano is the largest urban center, with the highest population density and traffic levels; Cremona is a key industrial hub, hosting numerous food industries; Sondrio is a small mountain town in the Alps, with low population density, except during holidays, and limited industrial activity. These visualisations confirm the region’s heterogeneity, with Milano showing the greatest variability and critical values. We also observe that missing data can occur at different locations depending on the day, though the overall proportion of missing values is minimal for these specific data, under 1%, resulting in a total number of 1914 observed values.

To enhance our analysis, we also consider two spatial covariates, known to affect air pollutant concentrations: population density and altitude. Indeed, densely populated areas, such as major urban centers, often experience higher pollutant levels due to increased emissions. We utilise population density data (in 1000 *inhab/km*<sup>2</sup>) from Istituto Nazionale di Statistica, for all 1506 municipalities in Lombardy, accessible by *Geoportale Lombardia* (Geoportale della Lombardia [2023]). In particular, as shown in Figure 9, we reconstruct a smooth population density field using a mild smoothing obtained through Spatial Regression model with Partial Differential Equation (PDE) regularisation, implemented in *fdapDE* (see, e.g., Sangalli [2021]). On the other hand, altitude plays also an important role in explaining the spatial distribution of  $\text{NO}_2$ . Indeed flat areas, unlike mountainous regions, are more prone to air stagnation, which generally leads to higher concentrations of air pollutants. The altitude map of Lombardy, expressed in *log(meters)*, is provided by the Digital Terrein Model (Geoportale della Lombardia [2023]), and depicted in the right panel of Figure 9.

## 6.2. Exploring the $\text{NO}_2$ concentration via STQR-PDE

We apply a semiparametric STQR-PDE model to analyse the temporal trend of  $\text{NO}_2$  concentration in the period 1<sup>st</sup> – 23<sup>th</sup> December 2018, using the population density and the altitude as covariates. Our analysis focuses on the 99% quantile to examine the right tail of the pollutant distribution, representing the most severe scenario for public health. As expected, the population density gives a positive contribution to



**Figure 9.** Left: population density field (in  $\log(1000\text{inhab}/\text{km}^2)$ ), darker colors correspond to higher density. Right: Digital Terrestrial Model (in  $\log(\text{meters})$ ), darker colors correspond to lower altitude. Both variables are displayed on a logarithmic scale for visualisation purposes.

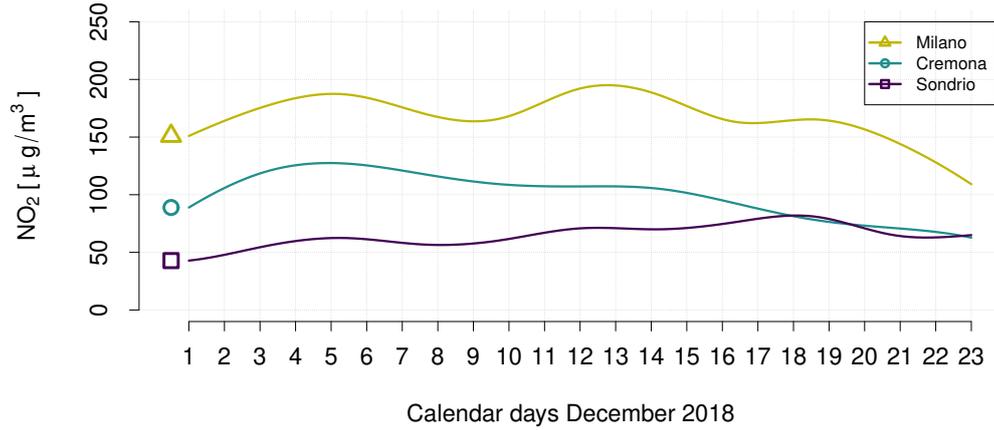
$\text{NO}_2$  concentration, with an estimated coefficient  $\hat{\beta}_{\text{population}} = 0.947$ , while the altitude provides a beneficial effect, with a coefficient  $\hat{\beta}_{\text{altitude}} = -2.018$ .

Figures 10 and 11 display the estimated 99% quantile, over time and space. In particular, Figure 10 illustrates the temporal evolution of the estimated 99% quantile fields, in the three municipalities of Milano, Cremona and Sondrio. The  $\text{NO}_2$  concentration in Milano exhibits the highest values and variability, displaying a clear weekly pattern. Notably, weekends (1-2, 8-9, 15-16, and 22-23 December) show lower concentrations, coherently with the lower work-related traffic and tendency of people to leave the city during weekends. This effect is particularly pronounced during the days 7-8-9 December, with 7<sup>th</sup> December being a bank holiday in Milano. During these days there is a significant decrease in the pollutant levels, that also benefits to subsequent days. Conversely, the maximum value is reached in Milano on the working day 13<sup>th</sup> December, with a  $\text{NO}_2$  concentration nearly reaching  $200\mu\text{g}$ , which corresponds to the hourly maximum threshold set by ARPA. The significant decrease in  $\text{NO}_2$  levels observed both in Milano and in Cremona towards the end of the considered period is likely linked to heavy rainfall occurred across the whole Po Valley during those days. Moreover, we notice as our method effectively captures complex spatio-temporal interactions, as evidenced by the three curves reflecting non-linear trends.

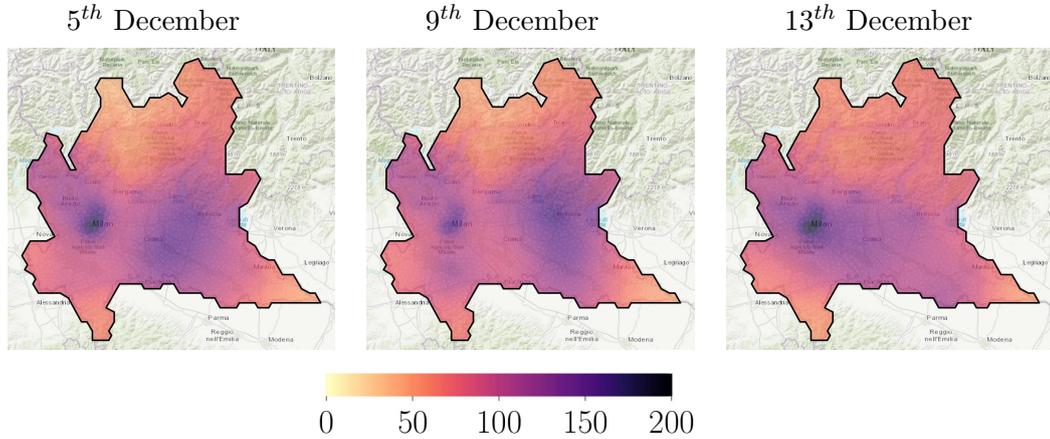
Figure 11 permits to explore the spatial distribution of  $\text{NO}_2$ . This figure shows the 99% quantile estimate in three days: the 5<sup>th</sup>, 9<sup>th</sup> and 13<sup>th</sup> of December. The figure also highlights the concerning situation in Milano, where pollutant concentrations are significantly higher than in the rest of the region. Visual analysis confirms the localised beneficial effect of the bank holiday in Milano, with a reduced peak observed on the 9<sup>th</sup> December, consistent with the previously discussed time trend analysis.

## 7. Discussion

In this work we have proposed an innovative semiparametric model for space-time quantile regression. The STQR-PDE model has proven to be a useful tool for exploring complex space-time phenomena, characterised by non-standard tail distributions and local outliers, with a focus on high-order quantiles rather than the mean. The benefits of this model are demonstrated through its application to  $\text{NO}_2$  concentration analysis in Lombardy, where the data clearly shows a significant right-skewed distribution, which is central to our investigation. The model provides valuable insights



**Figure 10.** Time trend of the estimated 99% quantile in the three cities: Milano (dark yellow), Cremona (dark cyan) and Sondrio (dark purple).



**Figure 11.** Estimated 99% spatial quantile surfaces of  $\text{NO}_2$ , in the days 5-9-13 December (same days as in Figure 1).

into the spatio-temporal dynamics of this issue, revealing crucial correlations between pollutant concentration, morphological factors, and human behaviours, such as people leaving the city during weekends or holidays. These relationships are essential in the environmental sector, forming a foundation for policymakers to develop effective regulations, aimed at addressing this urgent issue and enhancing the health of the local population. Additionally, STQR-PDE exhibits competitive advantages in simulations compared to existing techniques, both with fully observed data and under various missing data structures.

The proposed method offers several directions for future extensions. One potential development goes towards the use of physics-informed penalties, as shown, in a linear regression setting, by Arnone, Azzimonti, et al. (2019) and in the space-only model of Castiglione et al. (2025). Another significant challenge in quantile regression is the simultaneous estimation of multiple quantile levels. When each quantile level is estimated independently, as is commonly done in standard quantile regression settings, there is no guarantee that the resulting quantile curves will respect the monotonicity property, potentially leading to the so-called crossing problem. This issue could be

addressed by developing appropriate strategies to prevent the crossing problem, as shown, for instance, by Bondell et al. (2010) and He et al. (2002) in more classical quantile regression settings.

## Acknowledgements

I. Di Battista and L.M. Sangalli acknowledge the PRIN2022 project CoEnv-Complex Environmental Data and Modeling (CUP2022E3RY23) founded by the European Union – NextGenerationEU program and by the Italian Ministry for University and Research (MUR). M.F. De Sanctis and L.M. Sangalli acknowledge the European Union - NextGenerationEU program, Mission 4, Component 2, in the framework of the GRINS -Growing Resilient, INclusive and Sustainable project (GRINS PE00000018 – CUP D43C22003110001). 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. Finally, I. Di Battista, M.F. De Sanctis and L.M. Sangalli acknowledge MUR research project Dipartimento di Eccellenza 2023 -2027, Dipartimento di Matematica, Politecnico di Milano.

## References

- Arnone, E., Azzimonti, L., Nobile, F., and Sangalli, L. M. (2019). Modeling spatially dependent functional data via regression with differential regularization. In: *Journal of Multivariate Analysis* 170, pp. 275–295. DOI: <https://doi.org/10.1016/j.jmva.2018.09.006>.
- Arnone, E., De Falco, C., Formaggia, L., Meretti, G., and Sangalli, L. M. (2023). Computationally efficient techniques for spatial regression with differential regularization. In: *International Journal of Computer Mathematics* 100.10, pp. 1971–1991. DOI: <https://doi.org/10.1080/00207160.2023.2239944>.
- Arnone, E., Sangalli, L. M., and Vicini, A. (2023). Smoothing spatio-temporal data with complex missing data patterns. In: *Statistical Modelling* 23, pp. 327–356. DOI: <https://doi.org/10.1177/1471082X211057959>.
- Augustin, N., Trenkel, V., Wood, S., and Lorance, P. (2013). Space-time modelling of blue ling for fisheries stock management. In: *Environmetrics* 24, pp. 109–119.
- Azzimonti, L., Nobile, F., Sangalli, L. M., and Secchi, P. (2014). Mixed finite elements for spatial regression with PDE penalization. In: *SIAM/ASA Journal on Uncertainty Quantification* 2.1, pp. 305–335. DOI: <http://dx.doi.org/10.1137/130925426>.
- Azzimonti, L., Sangalli, L. M., Secchi, P., Domanin, M., and Nobile, F. (2015). Blood Flow Velocity Field Estimation Via Spatial Regression With PDE Penalization. In: *Journal of the American Statistical Association* 110, pp. 1057–1071. DOI: <https://doi.org/10.1080/01621459.2014.946036>.
- Bondell, H. D., Reich, B. J., and Wang, H. (2010). Noncrossing quantile regression curve estimation. In: *Biometrika* 97.
- California Air Resources Board (2023). *California Environmental Protection Agency*. Accessed 2023.
- Castiglione, C., Arnone, E., Bernardi, M., Farcomeni, A., and Sangalli, L. (2025). PDE-regularised spatial quantile regression. In: *Journal of Multivariate Analysis* 205. DOI: <https://doi.org/10.1016/j.jmva.2024.105381>.

- Castillo-Mateo, J., Asín, J., Cebrián, A. C., Gelfand, A. E., and Abaurrea, J. (2023). Spatial quantile autoregression for season within year daily maximum temperature data. In: *The Annals of Applied Statistics* 17.
- Das, P. and Ghosal, S. (2017). Analyzing ozone concentration by Bayesian spatio-temporal quantile regression. In: *Environmetrics* 28, e2443.
- Deb, S., Neves, C., and Roy, S. (2024). *Nonparametric quantile regression for spatio-temporal processes*. arXiv: 2405.13783 [stat.ME]. URL: <https://arxiv.org/abs/2405.13783>.
- Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum Likelihood from Incomplete Data Via the EM Algorithm. In: *Journal of the Royal Statistical Society: Series B (Methodological)* 39, pp. 1–22.
- Fasiolo, M., Wood, S. N., Zaffran, M., Nedellec, R., and Goude, Y. (2021a). Fast Calibrated Additive Quantile Regression. In: *Journal of the American Statistical Association* 116, pp. 1402–1412.
- (2021b). *qgam: Bayesian Nonparametric Quantile Regression Modeling in R*. In: *Journal of Statistical Software* 100.
- Geoportale della Lombardia (2023). *Regione Lombardia*. Accessed 2023.
- He, X., Ng, P., and Portnoy, S. (2002). Bivariate Quantile Smoothing Splines. In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 60.3, pp. 537–550.
- Knight, K. (1998). Limiting distributions for  $L_1$  regression estimators under general conditions. In: *The Annals of Statistics* 26, pp. 755–770.
- Koenker, R. (2005). *Quantile Regression*. Econometric Society Monographs. Cambridge: Cambridge University Press.
- Koenker, R. and Bassett, G. (1978). Regression Quantiles. In: *Econometrica* 46, pp. 33–50.
- Kotz, S., Kozubowski, T. J., and Podgórski, K. (2001). *The Laplace Distribution and Generalizations*. Boston, MA: Birkhäuser Boston.
- Lange, K. (2016). *MM Optimization Algorithms*. Philadelphia, PA: Society for Industrial and Applied Mathematics.
- Marra, G., Miller, D. L., and Zanin, L. (2012). Modelling the spatiotemporal distribution of the incidence of resident foreign population. In: *Statistica Neerlandica* 66, pp. 133–160.
- Nortier, B. (2021). “Automated Smoothing Parameter Estimation for Quantile Additive Models”. PhD thesis. University of Bristol.
- Nychka, D., Gray, G., Haaland, P., Martin, D., and O’Connell, M. (1995). A Nonparametric Regression Approach to Syringe Grading for Quality Improvement. In: *Journal of the American Statistical Association* 90.432, pp. 1171–1178.
- Open Data, Regione Lombardia (2024). *Meteo, inquinamento aria e acqua e altri dati da ARPA*. Accessed 2024. URL: <https://www.dati.lombardia.it/stories/s/Meteo-inquinamento-aria-e-acqua-e-altri-dati-da-AR/auv9-c2sj>.
- Palummo, A., Arnone, E., Clemente, A., Sangalli, L. M., Ramsay, J., and Formaggia, L. (2025). *fdaPDE: Physics-Informed Spatial and Functional Data Analysis*. GitHub, <https://github.com/fdaPDE/fdaPDE>.
- Reich, B. J. (2012). Spatiotemporal Quantile Regression for Detecting Distributional Changes in Environmental Processes. In: *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 61, pp. 535–553.
- Salama, D. S., Yousif, M., Gedamy, Y., Ahmed, H. M., Ali, M., and Shoukry, E. M. (2022). Satellite observations for monitoring atmospheric NO<sub>2</sub> in correlation with

- the existing pollution sources under arid environment. In: *Modeling Earth Systems and Environment* 8, pp. 4103–4121.
- Sangalli, L. M. (2021). Spatial Regression With Partial Differential Equation Regularisation. In: *International Statistical Review* 89, pp. 505–531. DOI: <https://doi.org/10.1111/insr.12444>.
- Sangalli, L. M., Ramsay, J. O., and Ramsay, T. O. (2013). Spatial Spline Regression Models. In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 75, pp. 681–703. DOI: <https://doi.org/10.1111/rssb.12009>.
- Shin, W., Kim, M., and Jung, Y. (2022). Efficient information-based criteria for model selection in quantile regression. In: *Journal of the Korean Statistical Society* 51.
- Sylvan, D., Târcolea, C., and Paris, A. (2015). *Space-time quantile surfaces of non-stationary random fields: a comparison study*. Manuscript or unpublished source.
- Wahba, G. (1990). *Spline Models for Observational Data*. Society for Industrial and Applied Mathematics.
- Wilhelm, M. and Sangalli, L. M. (2016). Generalized spatial regression with differential regularization. In: *Journal of Statistical Computation and Simulation* 86, pp. 2497–2518.
- Wood, S. N., Bravington, M. V., and Hedley, S. L. (2008). Soap Film Smoothing. In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 70, pp. 931–955.
- Yu, K. and Moyeed, R. A. (2001). Bayesian quantile regression. In: *Statistics & Probability Letters* 54, pp. 437–447.

## 8. Appendices

### Appendix A. Mathematical framework

#### A.1. Proof of Proposition 2.1

In order to prove the existence of a solution to the estimation problem stated in eq. 4, we first define the space containing the pairs  $(f, \boldsymbol{\beta})$  of our interest and some related properties.

**Definition A.1.** Let  $\mathcal{S}$  be the space of pairs  $(f, \boldsymbol{\beta})$  such that  $f \in V$  and  $\boldsymbol{\beta} \in \mathbb{R}^q$ , namely

$$\mathcal{S} = \{(f, \boldsymbol{\beta}) : f \in V, \boldsymbol{\beta} \in \mathbb{R}^q\}.$$

The addition and multiplication operations within the space  $\mathcal{S}$  are defined as follows

$$\begin{aligned} (f_1, \boldsymbol{\beta}_1) + (f_2, \boldsymbol{\beta}_2) &= (f_1 + f_2, \boldsymbol{\beta}_1 + \boldsymbol{\beta}_2), & \forall f_1, f_2 \in V, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2 \in \mathbb{R}^q, \\ a(f, \boldsymbol{\beta}) &= (af, a\boldsymbol{\beta}), & \forall f \in V, \boldsymbol{\beta} \in \mathbb{R}^q, a \in \mathbb{R}. \end{aligned}$$

Finally, we define the notion of convergence in the space  $\mathcal{S}$  as

$$(f_n, \boldsymbol{\beta}_n) \rightarrow (f, \boldsymbol{\beta}) \text{ in } \mathcal{S} \iff \begin{cases} f_n \rightarrow f & \text{in } V \\ \boldsymbol{\beta}_n \rightarrow \boldsymbol{\beta} & \text{in } \mathbb{R}^q \end{cases}, \quad \forall \{(f_n, \boldsymbol{\beta}_n)\} \subset \mathcal{S}.$$

Given the above definitions, we proceed relying on the characterisation reported in Proposition 6.5.1 in Lange (2016), followed also in Castiglione et al. (2025), which

states that any continuous convex function defined over a closed and convex domain achieves its minimum values within the domain. Therefore,  $\mathcal{V}_\alpha$  is non-empty provided that  $\mathcal{S}$  is closed and convex and  $J_\alpha$  is continuous and convex.

We first prove that  $\mathcal{S}$  is closed and convex. The closedness of  $\mathcal{S}$  follows from the closedness of the spaces  $V$  and  $\mathbb{R}^q$ . Specifically,  $V$  is closed owing to the vector space structure of the space  $V$ , ensured in its turn by the vector structure of the two Sobolev spaces appearing in its definition, plus the continuity of the differential operator  $B$ . On the other hand, the space  $\mathbb{R}^q$  is closed by definition. Moreover,  $\mathcal{S}$  is also convex holding

$$\begin{aligned} a(f_1, \beta_1) + (1-a)(f_2, \beta_2) &= (af_1, a\beta_1) + ((1-a)f_2, (1-a)\beta_2) = \\ &= (af_1 + (1-a)f_2, a\beta_1 + (1-a)\beta_2) \in \mathcal{S}, \end{aligned}$$

for any  $f_1, f_2 \in V, \beta_1, \beta_2 \in \mathbb{R}^q, a \in [0, 1]$ , where we used the convexity and vector space properties of  $\mathbb{R}^q$  and  $V$ , which are ensured by the vector space structure of  $\mathbb{R}^q$  and  $V$  and the linearity of the differential operator  $B$ . Moving to the continuity and convexity of the functional  $J_\alpha$ , we can observe that:  $\rho_\alpha$  is continuous and convex, see, e.g. Koenker and Bassett (1978);  $\int_0^T \int_{\mathcal{D}} (\Delta f)^2 d\mathbf{p}dt, \int_0^T \int_{\mathcal{D}} (\partial^2 f / \partial t^2)^2 d\mathbf{p}dt$  are continuous as proved in Arnone, Azzimonti, et al. (2019). These directly prove our thesis. Thus, both the statements hold meaning that the space  $\mathcal{V}_\alpha$  is not empty.

Finally we are left to prove that  $\mathcal{V}_\alpha$  is closed and convex. To this end, we define the sublevel set as

$$\mathcal{V}_\alpha(\tau) = \{v \in \mathcal{S} : J_\alpha(v) \leq \tau\} \quad \forall \tau : \mathcal{V}_\alpha(\tau) \neq \emptyset.$$

This set is convex since for any  $u, v \in \mathcal{V}_\alpha(\tau)$  and  $a \in [0, 1]$  we have

$$J_\alpha(au + (1-a)v) \leq aJ_\alpha(u) + (1-a)J_\alpha(v) \leq a\tau + (1-a)\tau = \tau \quad \forall u, v \in \mathcal{V}_\alpha(\tau), \forall a \in [0, 1],$$

which implies  $au + (1-a)v \in \mathcal{V}_\alpha(\tau)$ , where we used the convexity of the functional  $J_\alpha$ . Moreover, the sublevel set is also closed thanks to the continuity of the functional  $J_\alpha$ . Thus, the thesis follows by showing that  $\mathcal{V}_\alpha$  can be written as a sublevel set for a proper level. Indeed,

$$\mathcal{V}_\alpha = \{(\hat{f}_\alpha, \hat{\beta}_\alpha) : J_\alpha(\hat{f}_\alpha, \hat{\beta}_\alpha) = \inf_{(f, \beta) \in \mathcal{S}} J_\alpha(f, \beta)\} = \mathcal{V}_\alpha(\tau^*),$$

with  $\tau^* = J_\alpha(f^*, \beta^*)$  for some  $(f^*, \beta^*) \in \mathcal{S}$ . This concludes the proof.

## A.2. Proof of Proposition 2.2

With the same notation of Proof A.1, consider a pair  $(\tilde{f}_\alpha, \tilde{\beta}_\alpha) \in \mathcal{V}_\alpha$ , that is  $(\tilde{f}_\alpha, \tilde{\beta}_\alpha)$  is a solution to Equation 4. The space of solutions can be written as

$$\mathcal{V}_\alpha = \{(\hat{f}_\alpha, \hat{\beta}_\alpha) : J_\alpha(\hat{f}_\alpha, \hat{\beta}_\alpha) = \inf_{(f, \beta) \in \mathcal{S}} J_\alpha(f, \beta)\},$$

namely  $(\tilde{f}_\alpha, \tilde{\beta}_\alpha)$  is infimum for  $J_\alpha$ . This means that  $(\tilde{f}_\alpha, \tilde{\beta}_\alpha)$  is a global minimizer of  $J_\alpha$ , which concludes the proof.

### A.3. Details on the EM algorithm

As we have anticipated in Section 3.1, in this work we rely on a functional version of the EM algorithm in order to approximate the estimation problem stated in eq. 4 into a sequence of simpler problems having a more tractable representation.

The derivation of the proposed EM algorithm builds upon the result proved in Yu and Moyeed (2001), which guarantees that solving the estimation problem is equivalent to maximise the penalised log-likelihood Asymmetric Laplace of the following semiparametric model with  $AL$  error, as shown in Kotz et al. (2001)

$$Y_{ij} = \mathbf{x}_{ij}^\top \boldsymbol{\beta} + f(\mathbf{p}_i, t_j) + \epsilon_{ij} \quad \text{with} \quad \epsilon_{ij} \sim AL(\alpha, 0, 1), \quad i = 1, \dots, n, \quad j = 1, \dots, m, \quad (\text{A1})$$

where  $AL(\alpha, \mu, \sigma)$  is the Asymmetric Laplace law with shape  $\alpha \in \mathbb{R}$ , location  $\mu \in \mathbb{R}$  and scale  $\sigma \in \mathbb{R}^+$  parameters. The error terms  $\{\epsilon_{ij}\}_{i,j}$  are assumed to be independent and identically distributed as in Arnone, Azzimonti, et al. (2019). Then, leveraging the location-scale mixture representation of the  $AL$  law we can write

$$\epsilon_{ij} | \nu_{ij} \sim N(a_1 \nu_{ij}, a_2^2 \nu_{ij}), \quad \nu_{ij} \sim Exp(1),$$

where  $N(\cdot, \cdot)$  is the univariate Gaussian law and while  $Exp(\cdot)$  is the Exponential law. The parameters defining the Gaussian distribution are non-stochastic constants determined by the probability level  $\alpha$  as  $a_1 = \frac{1-2\alpha}{\alpha(1-\alpha)}$  and  $a_2 = \frac{2}{\alpha(1-\alpha)}$ . Thus, the augmented log-likelihood of model (A1) reads as follows

$$l_\alpha(f, \nu_{ij}, y_{ij}) = -\frac{1}{2} \log(2a_2^2 \nu_{ij}) - \frac{(y_{ij} - \mathbf{x}_{ij}^\top \boldsymbol{\beta} - f(\mathbf{p}_i, t_j) - a_1 \nu_{ij})^2}{2a_2^2 \nu_{ij}} - \nu_{ij}.$$

Combining such augmented representation of the Asymmetric Laplace model with the two PDE regularisation terms in (1), we obtain the penalised completed log-likelihood as

$$l_\alpha(f, \boldsymbol{\beta}; \boldsymbol{\nu}, \mathbf{y}) = \sum_{i=1}^n \sum_{j=1}^m l_\alpha(f, \boldsymbol{\beta}; \nu_{ij}, y_{ij}) + \left( -\frac{\lambda_D \nu}{2} \int_0^T \int_{\mathcal{D}} (\Delta f)^2 + \frac{\lambda_T \nu}{2} \int_0^T \int_{\mathcal{D}} \left( \frac{\partial^2 f}{\partial t^2} \right)^2 \right). \quad (\text{A2})$$

At this point, we can maximise the penalised log-likelihood (A2) via the EM algorithm. Specifically, the E-step computes its expected value which can be proved to be equal to  $-1/2 J_\alpha^{(k)}(f, \boldsymbol{\beta})$ , whose definition is given in (5). Finally, the M-step consists in the maximisation of such expected value. For more details on this functional version of the EM algorithm we refer to Castiglione et al. (2025).

### A.4. Discretisation matrices

In this section, we provide the definitions of the discretisation matrices used in Section 3.2. Let  $\psi$  be the vector collecting the  $N$  Finite Element Method (FEM) basis functions, and  $\varphi$  the vector collecting the  $M$  temporal basis. We define the mass, stiffness

and penalty matrices in space as

$$R_0 = \int_{\mathcal{D}} \psi \psi^\top, \quad R_1 = \int_{\mathcal{D}} \nabla \psi \nabla \psi^\top, \quad P_S = R_1^\top R_0^{-1} R_1.$$

For the time discretisation, instead, we set

$$R_t = \int_0^\top \varphi \varphi^\top, \quad P_T = \int_0^\top \frac{\partial^2 \varphi}{\partial t^2} \frac{\partial^2 \varphi^\top}{\partial t^2}$$

as the mass and penalty time matrix, respectively.

Moreover, we indicate by  $\tilde{R}_0$  and  $\tilde{R}_1$  the extensions to the spatio-temporal setting of the mass and stiffness matrices defined as

$$\tilde{R}_0 = R_t \otimes R_0, \quad \tilde{R}_1 = R_t \otimes R_1$$

Finally, combining the space and time dimension we define the overall penalty matrix, following the work in Arnone, Sangalli, et al. (2023), as

$$P = \lambda_D (R_t \otimes P_S) + \lambda_T (P_T \otimes R_0).$$

#### A.5. Proof of Proposition 4.1

In this section we prove the consistency of the  $\alpha$ -quantile estimator presented in Section 3.2. For the sake of simplicity we will denote by  $\nu$  the cardinality of the set of observed data, i.e.  $\nu = |O|$ . Let  $J_\alpha(\boldsymbol{\beta}, \mathbf{f}) = 1/|\nu| \sum_{(i,j) \in O} \rho_\alpha(y_{ij} - \mathbf{x}_{ij}^\top \boldsymbol{\beta} - \mathbf{A}_{ij}^\top \mathbf{f}) + \lambda_D/2 P_S(\mathbf{f}) + \lambda_T/2 P_t(\mathbf{f})$  be the finite element discretisation of the objective functional written in (2). The discretisation of the two penalty terms can be expressed as  $P_S(\mathbf{f}) = \mathbf{f}^\top (R_t \otimes P_S) \mathbf{f}$  and  $P_t(\mathbf{f}) = \mathbf{f}^\top (P_t \otimes R_0) \mathbf{f}$  following the definitions given in Arnone, Sangalli, et al. (2023). Proceeding as done in Castiglione et al. (2025), we introduce the quantile residual  $\varepsilon_{ij} = y_{ij} - \mathbf{c}_{ij}^\top \boldsymbol{\theta}_\alpha$  and the reparametrisation  $\boldsymbol{\delta} = \sqrt{\nu}(\boldsymbol{\theta} - \boldsymbol{\theta}_\alpha)$ , so that we can write the reparametrised scaled objective functional as

$$\begin{aligned} \nu J_\alpha(\boldsymbol{\delta}/\sqrt{\nu} + \boldsymbol{\theta}_\alpha) &= \sum_{(i,j) \in O} \{ \rho_\alpha(\varepsilon_{ij} - \mathbf{c}_{ij}^\top \boldsymbol{\delta}/\sqrt{\nu}) - \rho_\alpha(\varepsilon_{ij}) \} + \\ &+ \frac{\lambda_D \nu}{2} P_S(\boldsymbol{\delta}/\sqrt{\nu} + \boldsymbol{\theta}_\alpha) + \frac{\lambda_T \nu}{2} P_t(\boldsymbol{\delta}/\sqrt{\nu} + \boldsymbol{\theta}_\alpha), \end{aligned}$$

where it is worth noting that the penalty terms truly depend only on  $\mathbf{f}$ . Once we have introduced this reparametrisation, it is easy to show that minimising  $J_\alpha(\cdot)$  with respect to  $\boldsymbol{\theta}$  is equivalent to minimise it with respect to  $\boldsymbol{\delta}$ , with solution equal to  $\hat{\boldsymbol{\delta}}_\alpha = \sqrt{\nu}(\hat{\boldsymbol{\theta}}_\alpha - \boldsymbol{\theta}_\alpha)$ . Thus, we study the asymptotic properties of  $\hat{\boldsymbol{\delta}}_\alpha$  to then discuss the limiting behaviour of  $\hat{\boldsymbol{\theta}}_\alpha$ . We first focus on the data misfit term of the functional, denoted by  $S_\alpha(\boldsymbol{\delta})$ , and we apply the decomposition used in Knight (1998), so that

$S_\alpha(\boldsymbol{\delta}) = S'_\alpha(\boldsymbol{\delta}) + S''_\alpha(\boldsymbol{\delta})$ , with

$$S'_\alpha(\boldsymbol{\delta}) = \frac{1}{2\sqrt{\nu}} \sum_{(i,j) \in O} [(-\mathbf{c}_{ij}^\top \boldsymbol{\delta})(\mathbb{1}_{(\varepsilon_{ij} > 0)} - \mathbb{1}_{(\varepsilon_{ij} < 0)}) + (1 - 2\alpha)(-\mathbf{c}_{ij}^\top \boldsymbol{\delta})],$$

$$S''_\alpha(\boldsymbol{\delta}) = \sum_{(i,j) \in O} \int_0^{\mathbf{c}_{ij}^\top \boldsymbol{\delta} / \sqrt{\nu}} (\mathbb{1}_{(\varepsilon_{ij} \leq t)} - \mathbb{1}_{(\varepsilon_{ij} \leq 0)}) dt.$$

To study the asymptotic behaviour of  $S'_\alpha(\boldsymbol{\delta})$  we can observe that  $\mathbb{1}_{(\varepsilon_{ij} < 0)}$  and  $\mathbb{1}_{(\varepsilon_{ij} > 0)}$  are a sequence of independent Bernoulli random variables with parameter  $\alpha$  and  $1 - \alpha$ , respectively. So, it is easy to verify that

$$\mathbb{E}[S'_\alpha(\boldsymbol{\delta})] = 0, \quad \text{Var}[S'_\alpha(\boldsymbol{\delta})] = \alpha(1 - \alpha) \frac{1}{4\nu} \sum_{(i,j) \in O} (\mathbf{c}_{ij}^\top \boldsymbol{\delta})^2 = \frac{1}{4\nu} \boldsymbol{\delta}^\top D_{0,\nu} \boldsymbol{\delta}.$$

Thus, exploiting the central limit theorem we can write  $S'_\alpha(\boldsymbol{\delta}) = \boldsymbol{\delta}^\top \boldsymbol{\eta}_\nu + o_p(1)$ , with  $\boldsymbol{\eta}_\nu \sim N_{q+NM}(\mathbf{0}, D_{0,\nu})$ . Moving to  $S''_\alpha(\boldsymbol{\delta})$ , we can compute its expected value and approximate it by a second order Taylor expansion around  $\boldsymbol{\delta} = \mathbf{0}$ , obtaining

$$\begin{aligned} \mathbb{E}[S''_\alpha(\boldsymbol{\delta})] &= \sum_{(i,j) \in O} \int_0^{\mathbf{c}_{ij}^\top \boldsymbol{\delta} / \sqrt{\nu}} (\Pi_{\varepsilon_{ij} | \mathbf{p}_{i,t_j}}(t) - \alpha) dt = \\ &= \frac{1}{2\nu} \sum_{(i,j) \in O} \pi_{\varepsilon_{ij} | \mathbf{p}_{i,t_j}}(0) (\mathbf{c}_{ij}^\top \boldsymbol{\delta})^2 + \frac{1}{\nu^{3/2}} \sum_{(i,j) \in O} O(|\mathbf{c}_{ij}^\top \boldsymbol{\delta}|^3) = \\ &= \frac{1}{2} \boldsymbol{\delta}^\top D_{1,\nu} \boldsymbol{\delta} + O(\|\boldsymbol{\delta}\|^3 / \sqrt{\nu}), \end{aligned}$$

where, the residual term  $O(\|\boldsymbol{\delta}\|^3 / \sqrt{\nu})$  converges to 0, and owing to Assumption 1,  $\pi_{ij} = \pi_{\varepsilon_{ij} | \mathbf{p}_{i,t_j}}(0)$  is bounded for any  $(i, j) \in O$ . Moving to the analysis of the variance we exploit the upper bound recalled in Castiglione et al. (2025), so that

$$\text{Var}[S''_\alpha(\boldsymbol{\delta})] \leq \frac{1}{\sqrt{\nu}} \sup_{(i,j) \in O} |\mathbf{c}_{ij}| \|\boldsymbol{\delta}\| \mathbb{E}[S''_\alpha(\boldsymbol{\delta})] = O(\nu^{-1/2}).$$

Therefore,  $S''_\alpha(\boldsymbol{\delta})$  converges in probability to its mean and it can be written as  $S''_\alpha(\boldsymbol{\delta}) = \boldsymbol{\delta}^\top D_{1,\nu} \boldsymbol{\delta} / 2 + o_p(1)$ .

We now move to the analysis of the two non-stochastic penalty terms. Since they depend only on the nonparametric vector  $\mathbf{f}$ , we will refer with  $\boldsymbol{\delta}_f$  to the subvector of  $\boldsymbol{\delta}$  associated to  $\mathbf{f}$ . Moreover, we define  $\tilde{P}_s = R_t \otimes P_s = \tilde{R}_1^\top \tilde{R}_0^{-1} \tilde{R}_1$ ,  $\tilde{P}_t = P_t \otimes R_0$  and we recall that  $\mathbf{g}_\alpha = \tilde{R}_0^{-1} \tilde{R}_1 \mathbf{f}_\alpha$ . Thus, we can write the reparametrised penalty terms as

$$\begin{aligned} \nu P_s(\boldsymbol{\delta}_f / \sqrt{\nu} + \mathbf{f}_\alpha) &= \nu \{(\boldsymbol{\delta}_f / \sqrt{\nu} + \mathbf{f}_\alpha)^\top \tilde{P}_s (\boldsymbol{\delta}_f / \sqrt{\nu} + \mathbf{f}_\alpha)\} = \boldsymbol{\delta}_f^\top \tilde{P}_s \boldsymbol{\delta}_f + 2\sqrt{\nu} \boldsymbol{\delta}_f^\top \tilde{R}_1^\top \mathbf{g}_\alpha + \nu \mathbf{f}_\alpha^\top \tilde{P}_s \mathbf{f}_\alpha, \\ \nu P_t(\boldsymbol{\delta}_f / \sqrt{\nu} + \mathbf{f}_\alpha) &= \nu \{(\boldsymbol{\delta}_f / \sqrt{\nu} + \mathbf{f}_\alpha)^\top \tilde{P}_t (\boldsymbol{\delta}_f / \sqrt{\nu} + \mathbf{f}_\alpha)\} = \boldsymbol{\delta}_f^\top \tilde{P}_t \boldsymbol{\delta}_f + 2\sqrt{\nu} \boldsymbol{\delta}_f^\top \tilde{P}_t \mathbf{f}_\alpha + \nu \mathbf{f}_\alpha^\top \tilde{P}_t \mathbf{f}_\alpha. \end{aligned}$$

In order to write the above expressions in terms of the whole vector  $\boldsymbol{\delta} / \sqrt{\nu} + \boldsymbol{\theta}_\alpha$ , we rely on the completed matrices  $\tilde{\mathbf{P}}_s$ ,  $\tilde{\mathbf{P}}_t$ ,  $\tilde{\mathbf{R}}_1$  and completed vectors  $\tilde{\mathbf{g}}_\alpha$ ,  $\tilde{\mathbf{h}}_\alpha$  as defined

in Section 4. Therefore, we can rewrite the reparametrised penalty terms as

$$\begin{aligned}\nu P_s(\boldsymbol{\delta}/\sqrt{\nu} + \boldsymbol{\theta}_\alpha) &= \boldsymbol{\delta}^\top \tilde{P}_s \boldsymbol{\delta} + 2\sqrt{\nu} \boldsymbol{\delta}^\top \tilde{R}_1^\top \tilde{\mathbf{g}}_\alpha + \nu \boldsymbol{\theta}_\alpha^\top \tilde{P}_s \boldsymbol{\theta}_\alpha, \\ \nu P_t(\boldsymbol{\delta}/\sqrt{\nu} + \boldsymbol{\theta}_\alpha) &= \boldsymbol{\delta}^\top \tilde{P}_t \boldsymbol{\delta} + 2\sqrt{\nu} \boldsymbol{\delta}^\top \tilde{\mathbf{h}}_\alpha + \nu \boldsymbol{\theta}_\alpha^\top \tilde{P}_t \boldsymbol{\theta}_\alpha.\end{aligned}$$

Putting all the terms together and for  $\nu$  sufficiently large we can write the functional as

$$\begin{aligned}\nu J_\alpha(\boldsymbol{\delta}/\sqrt{\nu} + \boldsymbol{\theta}_\alpha) &= \frac{1}{2} \boldsymbol{\delta}^\top D_{1,\nu} \boldsymbol{\delta} - \boldsymbol{\delta}^\top \boldsymbol{\eta}_\nu + \frac{\lambda_D}{2} \boldsymbol{\delta}^\top \tilde{P}_s \boldsymbol{\delta} + \lambda_D \sqrt{\nu} \boldsymbol{\delta}^\top \tilde{R}_1^\top \tilde{\mathbf{g}}_\alpha + \\ &+ \frac{\lambda_T}{2} \boldsymbol{\delta}^\top \tilde{P}_t \boldsymbol{\delta} + \lambda_T \sqrt{\nu} \boldsymbol{\delta}^\top \tilde{\mathbf{h}}_\alpha + o_p(1).\end{aligned}$$

Owing to Assumption 2, the matrix  $D_{1,\nu}$  is positive definite, therefore the functional  $J_\alpha(\boldsymbol{\delta}/\sqrt{\nu} + \boldsymbol{\theta}_\alpha)$  is convex and admits a unique minimiser  $\hat{\boldsymbol{\delta}}_\alpha$  given by

$$\hat{\boldsymbol{\delta}}_\alpha = (D_{1,\nu} + \lambda_D \tilde{P}_s + \lambda_T \tilde{P}_t)^{-1} (\boldsymbol{\eta}_\nu - \lambda_D \sqrt{\nu} \tilde{R}_1^\top \tilde{\mathbf{g}}_\alpha - \lambda_T \sqrt{\nu} \tilde{\mathbf{h}}_\alpha) + o_p(1).$$

Consequently,  $\hat{\boldsymbol{\delta}}_\alpha$  has asymptotic normal distribution with the following mean and variance

$$\begin{aligned}\mathbb{E}[\hat{\boldsymbol{\delta}}_\alpha] &= -(D_{1,\nu} + \lambda_D \tilde{P}_s + \lambda_T \tilde{P}_t)^{-1} (\lambda_D \sqrt{\nu} \tilde{R}_1^\top \tilde{\mathbf{g}}_\alpha + \lambda_T \sqrt{\nu} \tilde{\mathbf{h}}_\alpha) + o(1), \\ \text{Var}[\hat{\boldsymbol{\delta}}_\alpha] &= (D_{1,\nu} + \lambda_D \tilde{P}_s + \lambda_T \tilde{P}_t)^{-1} D_{0,\nu} (D_{1,\nu} + \lambda_D \tilde{P}_s + \lambda_T \tilde{P}_t)^{-1} + o(1).\end{aligned}$$

We can now derive the asymptotic bias and variance of the estimator  $\hat{\boldsymbol{\theta}}_\alpha$  as  $\text{Bias}[\hat{\boldsymbol{\theta}}_\alpha] = \mathbb{E}[\hat{\boldsymbol{\delta}}_\alpha]/\sqrt{\nu}$  and  $\text{Var}[\hat{\boldsymbol{\theta}}_\alpha] = \text{Var}[\hat{\boldsymbol{\delta}}_\alpha]/\nu$ . Proceeding as done in Castiglione et al. (2025), we approximate the bias of the estimator via a second-order Taylor expansion around  $(\lambda_D, \lambda_T) = (0, 0)$ , obtaining

$$\begin{aligned}\text{Bias}[\hat{\boldsymbol{\theta}}_\alpha] &= -\lambda_D D_{1,\nu}^{-1} \tilde{R}_1^\top \tilde{\mathbf{g}}_\alpha - \lambda_T D_{1,\nu}^{-1} \tilde{\mathbf{h}}_\alpha + \lambda_D^2 D_{1,\nu}^{-1} \tilde{P}_s D_{1,\nu}^{-1} \tilde{R}_1^\top \tilde{\mathbf{g}}_\alpha + \\ &+ \lambda_T^2 D_{1,\nu}^{-1} \tilde{P}_t D_{1,\nu}^{-1} \tilde{\mathbf{h}}_\alpha + \lambda_D \lambda_T D_{1,\nu}^{-1} \tilde{P}_t D_{1,\nu}^{-1} \tilde{R}_1^\top \tilde{\mathbf{g}}_\alpha + \\ &+ O(\lambda_D^3) + O(\lambda_T^3) + O(\lambda_D^2 \lambda_T) + O(\lambda_D \lambda_T^2) + o(\nu^{-1/2}),\end{aligned}$$

which is,  $\text{Bias}[\hat{\boldsymbol{\theta}}_\alpha] = O(\lambda_D) + O(\lambda_T) + O(\lambda_D \lambda_T) + o(\nu^{1/2})$ . Applying the same strategy to the variance term we obtain

$$\begin{aligned}\text{Var}[\hat{\boldsymbol{\theta}}_\alpha] &= \frac{1}{\nu} D_{1,\nu}^{-1} D_{0,\nu} D_{1,\nu}^{-1} + \frac{2\lambda_D}{\nu} D_{1,\nu}^{-1} \tilde{P}_s (D_{1,\nu}^{-1} D_{0,\nu} D_{1,\nu}^{-1}) + \frac{2\lambda_T}{\nu} D_{1,\nu}^{-1} \tilde{P}_t (D_{1,\nu}^{-1} D_{0,\nu} D_{1,\nu}^{-1}) + \\ &+ O(\lambda_D^2) + O(\lambda_T^2) + O(\lambda_D \lambda_T) + o(\nu^{-1}),\end{aligned}$$

which is,  $\text{Var}[\hat{\boldsymbol{\theta}}_\alpha] = O(\nu^{-1}) + O(\lambda_D \nu^{-1}) + O(\lambda_T \nu^{-1})$ . Relying on the expressions for the asymptotic bias and variance just computed, and assuming that  $\lambda_D \sqrt{\nu} \rightarrow \bar{\lambda}_D$  and  $\lambda_T \sqrt{\nu} \rightarrow \bar{\lambda}_T$  for some finite values  $\bar{\lambda}_D, \bar{\lambda}_T$ , we can write the asymptotic distribution of the estimator  $\hat{\boldsymbol{\theta}}_\alpha$  as

$$\sqrt{\nu}(\hat{\boldsymbol{\theta}}_\alpha - \boldsymbol{\theta}_\alpha) \xrightarrow{d} N_{q+NM}(-\bar{\lambda}_D D_1^{-1} \tilde{R}_1^\top \tilde{\mathbf{g}}_\alpha - \bar{\lambda}_T D_1^{-1} \tilde{\mathbf{h}}_\alpha, D_1^{-1} D_0 D_1^{-1}).$$

Moreover, assuming that  $\lambda_D = o(\nu^{-1/2})$  and  $\lambda_T = o(\nu^{-1/2})$ , then the estimator  $\hat{\boldsymbol{\theta}}_\alpha$  is asymptotically unbiased. Finally, in order to prove the consistency of the estimator we analyse its Mean Squared Error (MSE), obtaining

$$\begin{aligned} MSE(\hat{\boldsymbol{\theta}}_\alpha) &= \text{Var}[\hat{\boldsymbol{\theta}}_\alpha] + \text{Bias}[\hat{\boldsymbol{\theta}}_\alpha]\text{Bias}[\hat{\boldsymbol{\theta}}_\alpha]^\top = \\ &= O(\nu^{-1}) + O(\lambda_D\nu^{-1}) + O(\lambda_T\nu^{-1}) + \\ &+ O(\lambda_D^2) + O(\lambda_T^2) + O(\lambda_D\lambda_T) + O(\lambda_D^2\lambda_T) + O(\lambda_D\lambda_T^2). \end{aligned}$$

Therefore, the estimator is proved to be consistent for any  $\lambda_D = O(\nu^{-1/2})$  and  $\lambda_T = O(\nu^{-1/2})$ , with a convergent MSE with rate equal to  $O(\nu^{-1/2})$ .

## Appendix B. Simulation studies

In this section we report the analytical expressions of the smooth field  $\mu(\mathbf{p}, t)$  used in the data generation model described in Section 5.1. The mean field is the one proposed in Arnone, Sangalli, et al. (2023) and it is defined as:

$$\begin{aligned} \mu(x, y, t) &= \sin \left( 2\pi \left( \text{coe}(y)x \cos \left( \frac{9}{5t_f}t - 2 \right) - y \sin \left( \frac{9}{5t_f}t - 2 \right) \right) \right) \cdot \\ &\quad \cos \left( 2\pi \left( \text{coe}(y)x \cos \left( \frac{9}{5t_f}t - 2 + \frac{\pi}{2} \right) + \text{coe}(x)y \sin \left( \frac{\pi}{2} \left( \frac{9}{5t_f}t - 2 \right) \right) \right) \right), \end{aligned}$$

where  $\text{coe}(x) = 1/2 \sin(5\pi x)e^{-x^2} + 1$ .

## MOX Technical Reports, last issues

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

- 32/2025** De Sanctis, M.F.; Di Battista, I.; Arnone, E.; Castiglione, C.; Palummo, A.; Bernardi, M.; Ieva, F.; Sangalli, L.M.  
*Exploring nitrogen dioxide spatial concentration via physics-informed multiple quantile regression*
- 31/2025** Botteghi, N.; Fresca, S.; Guo, M.; Manzoni, A.  
*HypeRL: Parameter-Informed Reinforcement Learning for Parametric PDEs*
- 30/2025** Rosafalco, L.; Conti, P.; Manzoni, A.; Mariani, S.; Frangi, A.  
*Online learning in bifurcating dynamic systems via SINDy and Kalman filtering*
- 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*