

MOX-Report No. 81/2026

Modeling spatio-temporal smoothness using nonlinear differential operators

Clemente, A.; Palummo, A.; Arnone, E.; Aston, J.A.D.; Panzica, 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

To appear in Journal of Multivariate Analysis, 2027, 217, 105728.
Publisher's version available at https://doi.org/10.1016/j.jmva.2026.105728

Modeling spatio-temporal smoothness using nonlinear differential operators

Aldo Clemente[1] Alessandro Palummo[1] Eleonora Arnone[2] John A.D. Aston[3]
Ferruccio Panzica[4] Laura M. Sangalli[1*]

the Alzheimer's Disease Neuroimaging Initiative[†]

[1] MOX - Dipartimento di Matematica, Politecnico di Milano, Milano, Italy
[2] Dipartimento di Management, Università degli Studi di Torino, Torino, Italy
[3] Statistical Laboratory, University of Cambridge, Cambridge, UK
[4] Fondazione IRCCS Istituto Neurologico Carlo Besta, Milano, Italy

Abstract

This work introduces a novel smoothing technique for functional data observed over space and time, based on a regularization term defined through a time-dependent nonlinear Partial Differential Equation (PDE). The proposed approach extends physics-informed statistical models, which have so far primarily relied on linear PDE-based penalties. By incorporating nonlinear PDEs into the regularization term, the method substantially broadens the applicability of smoothing models, while introducing new theoretical and computational challenges. Simulation studies illustrate the benefits of integrating physical knowledge into statistical modeling, with promising implications for real-world applications. The methodology is further demonstrated in a neuroimaging application, where nonlinear PDE regularization is used to model the progression of Alzheimer's disease from longitudinal Positron Emission Tomography data.

Keywords: nonparametric regression; penalized regression; functional data analysis; partial differential equations; optimal control; finite elements.

1 Introduction

In this paper, we introduce a physics-informed smoothing methodology based on a regularization term involving a nonlinear Partial Differential Equation (PDE), aimed at incorporating prior physical knowledge on phenomena governed by nonlinear dynamics. The proposed approach extends classical smoothing frameworks by moving beyond purely smoothness-driven penalties and explicitly embedding nonlinear dynamic information into the estimation procedure. Our work builds on the extensive literature on smoothing techniques based on roughness penalties defined through differential operators. Classical methods in this setting typically rely on simple isotropic linear operators (see, e.g., Cox, 1984; Wahba, 1981, 1990; Ramsay, 2002; Lai and Schumaker, 2007; Wood et al., 2008), where smoothness is most commonly enforced by penalizing second-order derivatives. These approaches are widely used in functional data analysis (Ramsay and Silverman, 2005; Ferraty and Vieu, 2006; Horváth and Kokoszka, 2012; Kokoszka and Reimherr, 2017) and nonparametric regression (Härdle, 1990; Green and Silverman, 1994). However, in this classical formulation, the penalty term primarily quantifies smoothness and does not encode structural or physical information about the underlying process.

More recently, a growing body of literature has focused on incorporating prior knowledge reflecting known physical or structural properties. Differential equations provide a natural and flexible framework for representing such information, as they describe the mechanisms governing the system dynamics. Within penalized estimation frameworks, this has led to the inclusion of PDEs directly in the regularization term. This paradigm has been successfully developed for spatial and functional data over complex domains (Azzimonti et al., 2015; Tomasetto et al., 2024), including two-dimensional manifolds (Ettinger et al., 2016; Lila et al., 2016) and non-convex three-dimensional geometries (Arnone et al., 2023a), and has also been extended to spatio-temporal settings (Arnone et al., 2019). In parallel, Bayesian hierarchical approaches and stochastic PDEs approaches have been developed to embed differential equations within probabilistic models (see, e.g., Wikle, 2003; Wikle and Hooten,

*Corresponding author: laura.sangalli@polimi.it

[†]Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (https://adni.loni.usc.edu). As such, the investigators within ADNI contributed to the design and implementation of ADNI and/or provided data, but did not participate in the analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf.

2010; Stroud et al., 2010; Roques et al., 2016; Lindgren et al., 2011; Carrizo Vergara et al., 2022; Clarotto et al., 2024).

Despite these advances, the existing literature on physics-informed statistical modeling has largely focused on *linear* PDEs. Although this choice simplifies both theoretical analysis and computation, it substantially limits the range of dynamics that can be represented. From a statistical perspective, extending PDE-based regularization to nonlinear operators raises fundamental questions concerning the well-posedness and analytical characterization of the resulting estimators. Nonlinear differential operators are essential for modeling realistic behaviors and arise naturally in numerous biological, chemical, and physical systems. Within penalized estimation frameworks, nonlinear operators have been studied mainly in one-dimensional contexts, considering Ordinary Differential Equations (Ramsay et al., 2007; Campbell et al., 2012; Ramsay and Hooker, 2017). In higher-dimensional settings, only a few Bayesian contributions have addressed nonlinear dynamics, such as the hierarchical modeling of the Fisher-Kolmogorov-Petrovsky-Piskunov (Fisher-KPP) equation in Hooten and Wikle (2008); Abboud et al. (2019). In general, statistical models for data analysis based on nonlinear PDEs in multidimensional domains remain largely unexplored.

These gaps motivate the development of statistical smoothing methods capable of incorporating nonlinear, physically informed spatio-temporal dynamics. In this work, we develop a physics-informed statistical model for the estimation of spatio-temporal fields governed by nonlinear PDEs. Our contribution extends the class of physics-informed statistical models reviewed in Sangalli (2021) by introducing nonlinear differential operators directly into the regularization term. Importantly, the proposed framework naturally reduces to existing linear PDE-based smoothing models when the underlying dynamics are linear, thereby ensuring coherence with established methodologies. This extension substantially broadens the applicability of physics-informed smoothing, while raising new theoretical and computational challenges. From a theoretical perspective, the variational arguments used, for instance, in Azzimonti et al. (2015) and Arnone et al. (2019) are no longer applicable. To establish the existence of a solution, we instead resort to results from optimal control theory (see, e.g., Manzoni et al., 2021), which allow us to establish the existence and provide an analytical characterization of the resulting estimator. From a computational standpoint, the presence of nonlinearity prevents solving the estimation problem as a single linear system. To address this issue, we introduce an iterative space-time discretization scheme, combining finite element methods in space with finite difference schemes in time (see, e.g., Quarteroni, 2009). The advantages of explicitly incorporating physical knowledge about nonlinear dynamics into statistical modeling are demonstrated through simulation studies. Finally, the proposed model accommodates functional data observed over multidimensional domains with complex geometries, such as two-dimensional Riemannian manifolds and non-convex volumes.

The methodology is motivated by a neuroimaging application based on tau Positron Emission Tomography (PET) data, which quantify the accumulation of pathological tau protein, a key biomarker of Alzheimer's disease (see, e.g., Sengupta and Kayed, 2022). This application poses substantial statistical challenges due to the complex geometry of the brain and the presence of nonlinear spatio-temporal dynamics. Standard neuroimaging approaches often either neglect the intrinsic anatomy of the brain, relying on Euclidean representations that are ill-suited to its highly convoluted structure, or fail to explicitly account for the underlying biological dynamics (see, e.g., Frackowiak et al., 2004; Bowman, 2014; Ombao et al., 2016). From a modeling perspective, the Fisher-KPP reaction-diffusion equation provides a well-established mechanistic description of the spread of toxic proteins in the brain (e.g., Weickenmeier et al., 2018; Fornari et al., 2019; Kevrekidis et al., 2020). In the applied mathematics literature, this equation has been widely used within advanced numerical frameworks to simulate disease progression (see, e.g., Weickenmeier et al., 2019; Corti et al., 2023). More recently, it has also been incorporated into Bayesian hierarchical models that represent the brain as a simplified network supporting diffusion-like dynamics (Schäfer et al., 2021, 2022). However, such network-based representations may limit the ability to capture the continuous spatial organization and complex geometry of the brain. Here, we propose a unified, data-driven statistical smoothing framework that integrates longitudinal PET data, complex brain anatomy, and Fisher-KPP-based spatio-temporal dynamics of the biomarker.

The remainder of the paper is organized as follows. Section 2 describes how spatio-temporal dynamics can be modeled using nonlinear PDE. Section 3 introduces the proposed physics-informed smoothing model and derives an analytical characterization of the associated estimator. Section 4 describes the discretization strategy and presents an iterative algorithm for the numerical approximation. Section 5 reports simulation studies assessing the performance of the proposed estimators. Section 6 presents an application to tau PET data from the Alzheimer's Disease Neuroimaging Initiative (ADNI). Section 7 concludes with directions for future research. Proofs of theoretical results are provided in Appendix A. The method is implemented in the R/C++ library `fdaPDE` (Arnone et al., 2026).

2 Modeling complex dynamics using nonlinear partial differential equations

Let $\mathcal{D} \subset \mathbb{R}^d$ be a bounded Lipschitz domain, with $d \leq 3$, and let $[0,T] \subset \mathbb{R}$ be a temporal domain. Let $f : \mathcal{D} \times [0,T] \to \mathbb{R}$ denote an unknown spatio-temporal field to be estimated. We assume that prior physical knowledge about f can be described by the following nonlinear PDE:

$$\begin{cases} \partial_t f + Lf - \Phi(f) = 0, & \text{in } \mathcal{D} \times (0,T), \\ f(\boldsymbol{p}, 0) = f_0, & \text{in } \mathcal{D}, \\ \partial_n f = h, & \text{on } \partial\mathcal{D} \times (0,T), \end{cases} \tag{1}$$

where ∂_t denotes the time derivative, L is a linear second-order elliptic operator, $\Phi(\cdot)$ is a nonlinear reaction term governing the system dynamics, f_0 and h denote the initial and boundary conditions, and ∂_n is the derivative along the outward unit normal on $\partial\mathcal{D}$. In this work, we focus on the Fisher-KPP equation, obtained from (1) by setting $Lf = -\operatorname{div}(K(\boldsymbol{p})\nabla f)$ and $\Phi(f) = rf(1-f)$, where $r > 0$. For all $\boldsymbol{p} \in \mathcal{D}$, $K(\boldsymbol{p}) \in \mathbb{R}^{d\times d}$ is a symmetric positive-definite diffusion tensor accounting for anisotropic and possibly spatially heterogeneous dispersion of f. This equation models population dynamics, chemical reaction-diffusion systems, and combustion phenomena, where spatial spread is governed by diffusion and growth by a nonlinear reaction term. In Section 6, we use it to model the spreading and misfolding of tau protein, a key biomarker of Alzheimer's disease. In this setting, the reaction parameter r represents the conversion rate from the healthy to the misfolded state.

Denote by $L^2(\mathcal{D})$ the space of square-integrable functions on $\mathcal{D}$, by $H^1(\mathcal{D})$ the Sobolev space of functions with first-order derivatives in $L^2(\mathcal{D})$ and homogeneous Neumann boundary conditions, and by $H^1(\mathcal{D})^*$ its dual, i.e., the space of bounded linear operators from $H^1(\mathcal{D})$ to $\mathbb{R}$. Note that the triplet $(H^1(\mathcal{D}), L^2(\mathcal{D}), H^1(\mathcal{D})^*)$ forms a Hilbert triplet (Zeidler, 2012, § 23.4), i.e.,

$$H^1(\mathcal{D}) \hookrightarrow L^2(\mathcal{D}) \hookrightarrow H^1(\mathcal{D})^*,$$

where $\hookrightarrow$ denotes continuous embedding (see, e.g., Brézis, 2011). To study the well-posedness of problem (1) and of the associated estimation problem, we consider the unknown function f as an element of the space

$$\mathcal{V} = H^1(0,T; H^1(\mathcal{D}), H^1(\mathcal{D})^*) = \Big\{ f \in L^2(0,T; H^1(\mathcal{D})) \,:\, \partial_t f \in L^2(0,T; H^1(\mathcal{D})^*) \Big\}.$$

Note that $\mathcal{V} \subset C^0([0,T]; L^2(\mathcal{D}))$; see, e.g., Prop. 23.23 in Zeidler (2012). In the linear case, i.e., when $\Phi(f) = 0$, existence and uniqueness of a solution follow, for instance, from the abstract parabolic theory (see, e.g., Chapter 10 of Salsa (2016)). In the nonlinear setting considered here, the analysis is considerably more challenging, and no general well-posedness result holds for all admissible nonlinear operators Φ (see, e.g., Zeidler, 2013).

3 Statistical Model

For $i \in \{1,\dots,n\}$ and $j \in \{1,\dots,m\}$, let $y_{ij} \in \mathbb{R}$ denote the observed value of a variable of interest at time $t_j = j\Delta t \in (0,T]$, with $\Delta t = T/m$. Observations are collected over a spatial support that may be either pointwise, corresponding to measurements at fixed locations $\boldsymbol{p}_i \in \mathcal{D}$, or areal, corresponding to averages over subregions $\mathcal{D}_i \subset \mathcal{D}$ with positive measure $|\mathcal{D}_i| > 0$. We model the observations using the nonparametric model

$$y_{ij} = C_{ij} f + \varepsilon_{ij},$$

where ε_{ij} are uncorrelated random errors with zero mean and constant variance $\sigma^2 > 0$, and C_{ij} denotes the observation operator defined as

$$C_{ij} f = \begin{cases} f(\boldsymbol{p}_i, t_j) = \displaystyle\int_0^T \int_{\mathcal{D}} f(\boldsymbol{x}, t) \delta_{\boldsymbol{p}_i} \, \delta_{t_j} \, d\boldsymbol{x} \, dt, & \text{pointwise} \\ \dfrac{1}{|\mathcal{D}_i|} \displaystyle\int_{\mathcal{D}_i} f(\boldsymbol{x}, t_j) \, d\boldsymbol{x} = \int_0^T \int_{\mathcal{D}} f(\boldsymbol{x}, t) \, \dfrac{\mathbf{1}_{\mathcal{D}_i}(\boldsymbol{x})}{|\mathcal{D}_i|} \, \delta_{t_j} \, d\boldsymbol{x} \, dt. & \text{areal} \end{cases}$$

Here, $\delta_{\boldsymbol{p}_i}$ denotes the Dirac delta function centered at $\boldsymbol{p}_i \in \mathcal{D}$, δ_{t_j} the one centered at $t_j \in [0,T]$, and $\mathbf{1}_{\mathcal{D}_i}$ the characteristic function of $\mathcal{D}_i \subset \mathcal{D}$.

We estimate the unknown function f by minimizing the penalized functional

$$\frac{1}{nm} \sum_{i=1}^{n} \sum_{j=1}^{m} (y_{ij} - C_{ij} f)^2 + \lambda \int_0^T \int_{\mathcal{D}} (\partial_t f + Lf - \Phi(f))^2 \, d\boldsymbol{x} \, dt. \tag{2}$$

The first term in (2) measures the discrepancy between the estimated field and the observed data, while the second term penalizes deviations from the nonlinear PDE (1). The smoothing parameter $\lambda > 0$ controls the trade-off between data fidelity and model fidelity.

3.1 Estimation problem as a nonlinear optimal control problem

When the regularization term in problem (2) involves a linear second-order partial differential equation, i.e., when $\Phi(f) \equiv 0$, the well-posedness of the corresponding estimation problem has been established in Arnone et al. (2019). Here, we study the existence of a solution to (2) in the more general setting of nonlinear reaction terms. For simplicity, we restrict to the areal spatial sampling design. In this setting, the sampling operator $C_{ij} : \mathcal{V} \to \mathbb{R}$ is continuous due to the embedding $\mathcal{V} \hookrightarrow C([0,T]; L^2(\mathcal{D}))$. In the pointwise spatio-temporal setting, stronger regularity assumptions are required; in particular, H^2-regularity in space is needed to ensure that the pointwise sampling operator is well defined.

Let $\mathcal{U} = L^2(0,T; L^2(\mathcal{D}))$, and consider the physical knowledge in (1) with an unknown function $g \in \mathcal{U}$ as the right-hand side of the first equation, i.e.,

$$\partial_t f + Lf - \Phi(f) = g. \tag{3}$$

The function g measures deviations of the estimated field f from the prescribed physical knowledge (1), and can be interpreted as a distributed control term in an optimal control framework (see, e.g., Manzoni et al., 2021; Tröltzsch, 2010). The key to solving the physics-informed smoothing problem (2) is to reformulate it in terms of the two unknown functions f and g, by expressing (3) in operator form. To this end, we introduce the operator $G : \mathcal{V} \times \mathcal{U} \to L^2(0,T; H^1(\mathcal{D})^*) \times L^2(\mathcal{D})$ and write (3) as $G(f,g) = 0$, the so-called state equation. With this notation, the estimation problem (2) can be equivalently written as

$$\underset{(f,g)\in\mathcal{V}\times\mathcal{U}}{\arg\min}\ J(f,g) = \underset{(f,g)\in\mathcal{V}\times\mathcal{U}}{\arg\min}\ \frac{1}{nm}\sum_{i=1}^{n}\sum_{j=1}^{m}(y_{ij} - C_{ij}f)^2 + \lambda\,\|g\|_{\mathcal{U}}^2, \qquad \text{subject to } G(f,g) = 0. \tag{4}$$

Problem (4) takes the form of a nonlinear optimal control problem (Tröltzsch, 2010). Following standard arguments from optimal control theory, we first analyze the state equation. In particular, Lemma 1 ensures that the *control-to-state map* $g \mapsto f(g)$ is well defined and continuous from $\mathcal{U}$ to $\mathcal{V}$, while Lemma 2 establishes its Fréchet differentiability.

Lemma 1 (Well-posedness of the state equation)**.** *For any $g \in \mathcal{U}$ and $f_0 \in L^2(\mathcal{D})$ such that $0 \le f_0 \le 1$ almost everywhere in $\mathcal{D}$, there exists a unique weak solution $f \in \mathcal{V}$ to the state equation. Moreover, the following bound holds:*

$$\|f\|_{\mathcal{V}}^2 \le C\left(\|f_0\|_{L^2(\mathcal{D})}^2 + \|g\|_{\mathcal{U}}^2\right).$$

Lemma 2. *The control-to-state map $g \mapsto f(g)$ is Fréchet differentiable. The operator $f'(g) : \mathcal{U} \to \mathcal{V}$ maps $v \in \mathcal{U}$ to the weak solution $z = f'(g)v$ of*

$$\begin{cases} \partial_t z + Lz + (1 - 2f(g))z = v, & \text{in } \mathcal{D}\times(0,T), \\ z(\boldsymbol{p}, 0) = 0, & \text{in } \mathcal{D}, \\ \partial_n z = 0, & \text{on } \partial\mathcal{D}\times(0,T). \end{cases} \tag{5}$$

Thanks to Lemmas 1 and 2, the following result establishes the existence of a minimizer $(\hat{f}, \hat{g})$ for problem (4).

Proposition 1. *Let L^* denote the adjoint operator of L. There exists at least one optimal pair $(f,g) \in \mathcal{V}\times\mathcal{U}$ solving problem (4). Moreover, if $(\hat{f}, \hat{g})$ is an optimal solution, then there exists a unique adjoint state $\hat{p} \in \mathcal{V}$ such that the triplet $(\hat{f}, \hat{g}, \hat{p})$ satisfies the first-order optimality conditions:*

$$\begin{cases} \partial_t f + Lf - r\,f(1-f) = g, & \text{in } \mathcal{D}\times(0,T), \\ f(0) = f_0, & \text{in } \mathcal{D}, \\ \partial_n f = 0, & \text{on } \partial\mathcal{D}\times(0,T) \end{cases} \tag{6}$$

$$\begin{cases} -\partial_t p + L^* p - r\,(1-2f)p = \dfrac{1}{nm}\displaystyle\sum_{i,j}(y_{ij} - C_{ij}f)\frac{\mathbf{1}_{\mathcal{D}_i}\delta_{t_j}}{|\mathcal{D}_i|}, & \text{in } \mathcal{D}\times(0,T), \\ p(T) = 0, & \text{in } \mathcal{D}, \\ \partial_n p = 0, & \text{on } \partial\mathcal{D}\times(0,T) \end{cases} \tag{7}$$

$$(\lambda\hat{g} - \hat{p}, u - \hat{g})_{\mathcal{U}} \ge 0, \quad \forall u \in \mathcal{U}. \tag{8}$$

Equation (6) corresponds to the state equation, while (7) defines the associated adjoint equation and (8) the resulting variational inequality. Proposition 1 characterizes the estimator as the solution of a coupled optimality

system, forming the basis for the discretization strategy in Section 4. Unlike the setting considered in Azzimonti et al. (2014) Arnone et al. (2019), where a linear state equation leads, after discretization, to a coupled system of linear PDEs, the presence of nonlinearity in (6) fundamentally alters the problem structure. In this case, the optimality conditions yield a nonlinear, mutually coupled system that cannot be solved in a single step. To address this issue, we propose an iterative scheme that exploits the control-to-state structure of the problem, as detailed in the following section.

4 Discretization of the estimation problem

The estimation problem in Section 3 is infinite-dimensional and does not admit a closed-form solution; it must therefore be addressed numerically. We adopt a three-step iterative procedure exploiting the control-to-state structure of the problem. Specifically, we solve the nonlinear state equation (6) for fixed g, then the adjoint equation (7) for fixed f, and finally update g by enforcing the variational inequality (8), iterating until convergence.

We discretize the state and adjoint equations using finite elements in space and finite differences in time. Let $\mathcal{D}_h$ be a triangulation of the spatial domain $\mathcal{D}$; see, e.g., Fig. 4 for a tetrahedral mesh of the considered template brain anatomy. Define the H^1-conforming finite element space of order k as the space of continuous functions that are polynomials of degree k on each element of the triangulation; see, e.g., Quarteroni (2009) for an introduction to finite element analysis. For simplicity, we consider linear finite elements ($k = 1$) and denote by V_h the corresponding finite-dimensional space. Let $\psi_1, \dots, \psi_{N_h}$ be a basis for V_h, and $\xi_1, \dots, \xi_{N_h}$ the corresponding nodes, which coincide with the vertices of $\mathcal{D}_h$. Each ψ_i is Lagrangian, i.e., $\psi_i(\xi_j) = \delta_{ij}$, with $\delta_{ij} = 1$ if $i = j$ and 0 otherwise. Set $\boldsymbol{\psi} := (\psi_1, \dots, \psi_{N_h})^\top$, and for any $v_h \in V_h$, denote by $\boldsymbol{v}$ the vector of nodal values, $\boldsymbol{v} := (v_h(\xi_1), \dots, v_h(\xi_{N_h}))^\top$. By the Lagrangian property, v_h is uniquely determined by its nodal values and can be written as $v_h = \boldsymbol{v}^\top \boldsymbol{\psi}$. The discrete estimator f for the state equation is thus defined as

$$\hat{f}_h(\boldsymbol{p}, t) = \sum_{i=1}^{N_h} f_i(t)\psi_i(\boldsymbol{p}) = \boldsymbol{f}(t)^\top \boldsymbol{\psi}.$$

The discrete solutions $\hat{p}_h$ and $\hat{g}_h$ to the adjoint and variational equations (7) and (8), respectively, are defined similarly. Note that the resulting discrete estimators belong to V_h for each fixed time $t_j = j\Delta t$, with $j \in \{1, \dots, m\}$. Let $(u, v)_2 = \int_{\mathcal{D}} uv \, d\boldsymbol{x}$ denote the inner product in $L^2(\mathcal{D})$, and introduce the following bilinear and trilinear forms:

$$a(u, v) = \int_{\mathcal{D}} (K \nabla u) \cdot \nabla v \, d\boldsymbol{x} - r \int_{\mathcal{D}} u \, v \, d\boldsymbol{x}, \qquad r_N(u, v, w) = r \int_{\mathcal{D}} u \, v \, w \, d\boldsymbol{x}.$$

We further define the sampling-dependent bilinear and linear forms

$$l(u, v) = \begin{cases} \displaystyle\sum_{i=1}^{n} \left(\int_{\mathcal{D}} \frac{\mathbf{1}_{\mathcal{D}_i}}{|\mathcal{D}_i|} u \, d\boldsymbol{x} \right) \left(\int_{\mathcal{D}} \frac{\mathbf{1}_{\mathcal{D}_i}}{|\mathcal{D}_i|} v \, d\boldsymbol{x} \right), & \text{areal} \\ \displaystyle\sum_{i=1}^{n} u(\boldsymbol{p}_i) v(\boldsymbol{p}_i), & \text{pointwise} \end{cases}$$

$$b_j(v) = \begin{cases} \displaystyle\frac{1}{nm} \sum_{i=1}^{n} y_{ij} \left(\int_{\mathcal{D}} \frac{\mathbf{1}_{\mathcal{D}_i}}{|\mathcal{D}_i|} v \, d\boldsymbol{x} \right), & \text{areal} \\ \displaystyle\frac{1}{nm} \sum_{i=1}^{n} y_{ij} v(\boldsymbol{p}_i). & \text{pointwise} \end{cases}$$

Multiplying (6) and (7) by a test function $v_h \in V_h$ and integrating over $\mathcal{D}$ yields the weak formulation

$$\begin{cases} \big(\partial_t \hat{f}_h(t_j), v_h\big)_2 + a\big(\hat{f}_h(t_j), v_h\big) + r_N\big(\hat{f}_h(t_j), \hat{f}_h(t_j), v_h\big) = \big(g_h(t_j), v_h\big)_2, \\ \hat{f}_h(0) = f_0; \end{cases}$$

$$\begin{cases} -\big(\partial_t \hat{p}_h(t_j), v_h\big)_2 + a\big(\hat{f}_h(t_j), v_h\big) + 2\, r_N\big(p_h(t_j), \hat{f}_h(t_j), v_h\big) + \dfrac{1}{nm} l\big(\hat{f}_h(t_j), v_h\big) = b_j\big(v_h\big), \\ \hat{p}_h(T) = 0. \end{cases}$$

Introduce the $N_h \times N_h$ matrices $[A]_{pq} = a(\psi_q, \psi_p)$, $[M]_{pq} = (\psi_q, \psi_p)_2$, and $R[\boldsymbol{f}] = r_N(f_h, \psi_q, \psi_p)$, together with the sampling-dependent matrices $\Psi \in \mathbb{R}^{n \times N_h}$ and $D \in \mathbb{R}^{n \times n}$, defined as

$$D_{ii} = \begin{cases} |D_i|, & \text{areal} \\ 1, & \text{pointwise} \end{cases}, \qquad \Psi_{iq} = \begin{cases} \int_{\mathcal{D}} \frac{\mathbf{1}_{\mathcal{D}_i}}{|\mathcal{D}_i|} \psi_q \, d\boldsymbol{x}, & \text{areal} \\ \psi_q(\boldsymbol{p}_i). & \text{pointwise} \end{cases}$$

Note that, upon discretization, the only difference between areal and pointwise sampling lies in the definition of the matrices Ψ and D. Following the classical finite element method, it suffices to enforce the weak formulation for all basis functions $\{\psi_p\}_{p=1}^{N_h}$, yielding the algebraic formulation:

$$\begin{aligned} M\dot{\boldsymbol{f}}(t_j) + A\boldsymbol{f}(t_j) + R[\boldsymbol{f}(t_j)]\boldsymbol{f}(t_j) &= \boldsymbol{0}, \\ -M\dot{\boldsymbol{p}}(t_j) + A^\top \boldsymbol{p}(t_j) + 2\,R[\boldsymbol{p}(t_j)]\boldsymbol{p}(t_j) + \frac{1}{nm}\Psi^\top D\Psi\boldsymbol{f}(t_j) &= \frac{1}{nm}\Psi^\top D\boldsymbol{y}_j, \end{aligned} \tag{9}$$

where $\boldsymbol{y}_j \in \mathbb{R}^n$ denotes the vector collecting the n observations at time $t_j = j\Delta t$, for $j \in \{1, \dots, m\}$.

To discretize in time, we consider a uniform grid of $m+1$ time instants $t_0, \dots, t_m$, with $t_0 = 0$, $t_j = j\Delta t$, and $t_m = T$. The time derivative is approximated by a backward finite-difference scheme:

$$\partial_t f_h(t_{j+1}) \approx \frac{f_h(t_{j+1}) - f_h(t_j)}{\Delta t}.$$

Let $\boldsymbol{f}^j$ denote the vector of coefficients of $f_h(t_j)$, i.e., $\boldsymbol{f}^j = \boldsymbol{f}_h(t_j)$. Substituting this approximation into (9) and adopting a semi-implicit Euler scheme, we obtain the following discrete state and adjoint equations:

$$\begin{aligned} \left(\frac{1}{\Delta t}M + A + R[\boldsymbol{f}^j]\right)\boldsymbol{f}^{j+1} &= \frac{1}{\Delta t}M\boldsymbol{f}^j + M\boldsymbol{g}^{j+1}, \\ \left(\frac{1}{\Delta t}M + A^\top + 2R[\boldsymbol{f}^j]\right)\boldsymbol{p}^j &= \frac{1}{\Delta t}M\boldsymbol{p}^{j+1} - \frac{1}{nm}\Psi^\top D\left(\Psi\boldsymbol{f}^j - \boldsymbol{y}_j\right). \end{aligned} \tag{10}$$

This discretization scheme allows an efficient treatment of the nonlinear term, reducing both the state and adjoint problems to sequences of linear systems, one per time step. The two equations in (10) are solved in opposite time directions: the discrete state equation (first line) is advanced forward in time, while the discrete adjoint equation (second line) is solved backward in time. This so-called backward semi-implicit Euler scheme is a first-order time discretization that ensures unconditional stability (see, e.g. Quarteroni, 2009) and can be interpreted within the framework of the θ-method for parabolic PDEs with $\theta = 1$, a well-established approach for handling such problems. While higher-order schemes are available (such as the Crank-Nicolson approach, corresponding to $\theta = 1/2$), their advantages are mainly relevant when accurately solving PDEs in noise-free settings. In the present context, the goal is to estimate a smooth spatio-temporal field that is not an exact solution of the PDE and must be inferred from noisy, discretely observed data; in this setting, a first-order discretization provides an appropriate and robust choice.

As a final step, we update $\boldsymbol{g}$ by leveraging the variational inequality (8), which yields an approximation of the gradient of the objective function J with respect to $\boldsymbol{g}$:

$$\nabla J(\boldsymbol{f}^j, \boldsymbol{g}^j) \approx \lambda M\boldsymbol{p}^j - M\boldsymbol{g}^j. \tag{11}$$

Starting from an initial guess $\boldsymbol{g}_0$, we iteratively update $\boldsymbol{g}$ using (11) as

$$\boldsymbol{g}_{k+1} = \boldsymbol{g}_k + \alpha_k \nabla J_k,$$

where α_k is a, possibly adaptive, step size. An optimization method, such as gradient descent or BFGS (see, e.g., Nocedal and Wright, 1999), is then applied until convergence. The method is implemented in the `fdaPDE` library, which employs limited-memory BFGS with a backtracking line search by default.

Note that in this work, for simplicity, we consider data observed at evenly spaced time instants, a choice driven by the numerical scheme, which relies on a structured temporal grid. However, this requirement is not intrinsic to the model and can be relaxed within a missing-data framework as in Arnone et al. (2023b). Moreover, the mesh is chosen a priori to adequately capture the underlying signal. Ultra-fine meshes (i.e., meshes with a resolution significantly exceeding that of the signal) are unnecessary, as they yield comparable accuracy at a substantially higher computational cost (see, for instance, case (d) of the additional simulation studies included in Section B.2 of the Supplementary Material).

5 Simulation studies

This section presents simulation studies comparing the proposed method (Nonlinear) with its linear counterpart (Linear), introduced in Arnone et al. (2019) and implemented in the `fdaPDE` library (Arnone et al., 2026), and with spatio-temporal Thin-Plate Splines (TPS) from the R package `mgcv` (Wood, 2003); see also Wood (2017). The GitHub repository `https://github.com/fdaPDE/case-studies` provides the scripts to reproduce the simulation results.

We generate the data from the solution of the Fisher-KPP equation:

$$\begin{cases} \partial_t f - \operatorname{div}(K(\boldsymbol{p})\nabla f) - rf(1-f) = 0, & \text{in } \mathcal{D} \times (0,T), \\ \partial_n f = 0, & \text{on } \partial\mathcal{D} \times (0,T), \\ f(\boldsymbol{p},0) = f_0, & \text{in } \mathcal{D}, \end{cases} \tag{12}$$

where $\mathcal{D}$ is the unit square and $T = 1$. The model exhibits nonlinear dynamics for $r \neq 0$ and linear dynamics for $r = 0$. We set $r = 1$ in Simulations 1 and 3 and $r = 0$ in Simulation 2. In all cases, we consider a diagonal diffusion tensor K with $\operatorname{diag}(K) = (1.1, 0.9) \times 10^{-2}$, and an initial condition f_0 equal to the field in Section 5.5 of Arnone et al. (2023a), at $t = 0$. The second column of Fig. 1 shows the true field at three time instants.

Simulations 1 and 2 consider pointwise spatio-temporal data, to enable comparison with TPS, which is not implemented for areal observations. Simulation 3 instead considers areal observations in space, restricting the comparison to the Nonlinear and Linear approaches. Data are generated by adding Gaussian noise to the numerical solution of (12), with 30 repetitions per simulation setting. The Nonlinear and Linear approaches use an isotropic diffusion tensor in their regularization terms, a regular spatial mesh with 961 nodes, and a time grid with $m+1 = 11$ points. For TPS, we use 120 basis functions in space and 10 in time. Both the proposed method and the Linear model are provided with the true initial condition f_0. This choice is motivated by the real-data application, where the data dimensionality allows an accurate estimation of the initial condition. For fairness, TPS is also provided with the true value of the initial condition at the data locations for $t = 0$. The smoothing parameter λ is selected by 10-fold cross-validation for the Nonlinear and Linear approaches, and by generalized cross-validation for TPS. For the Nonlinear approach, the reaction parameter r is also treated as unknown, and jointly estimated with λ, via 10-fold cross-validation over the grid $\{0.0, 0.1, 0.25, 0.5, 0.9, 1.0, 1.1, 1.5, 2.0\}$. The methods are compared using the Root Mean Square Error (RMSE), computed on a regular spatial grid of $N = 3600$ spatial points and $m = 10$ time instants,

$$\text{RMSE} = \sqrt{\frac{1}{N\,m}\sum_{i,j}\big(f(\boldsymbol{p}_i, t_j) - \hat{f}(\boldsymbol{p}_i, t_j)\big)^2},$$

where f denotes the true field and $\hat{f}$ its estimate.

5.1 Simulation 1: pointwise observations in space-time, nonlinear dynamics

We generate data by evaluating the solution of (12) with $r = 1$ at $n = 250$ spatial locations, uniformly sampled at each time instant $t_j = j\Delta t$, $j \in \{1, \ldots, m\}$, and adding independent Gaussian noise with standard deviation 0.05, corresponding to approximately 5% of the data range. The first column of Fig. 1 shows the sampled data at three time instants, while the third to fifth columns display the mean estimated fields from the three competing methods over 30 simulation repetitions. For all repetitions, the nonlinear approach correctly identifies the reaction parameter r via 10-fold cross-validation, highlighting its ability to recover the underlying nonlinear dynamics. A visual comparison shows that the proposed method better captures both the spatial structure and the temporal evolution of the test field, especially at later time instants. Quantitatively, it achieves the lowest RMSE, as reported in the first row of Table 1, consistently outperforming both the Linear and TPS approaches.

Section B.2 of the Supplementary Material presents additional experiments that systematically vary key parameters, including the reaction parameter r, the number of time instants m, the number of spatial locations n, and the number of mesh nodes N_h, while keeping the others fixed, to assess the robustness of the proposed smoothing approach.

5.2 Simulation 2: pointwise observations in space-time, linear dynamics

In Simulation 2, we set $r = 0$ in (12), thus generating data from a linear parabolic PDE. The remaining parameters, noise structure, and data generation process are the same as in Section 5.1.

Across all simulation repetitions, the proposed Nonlinear approach correctly identifies the absence of nonlinear dynamics, selecting $r = 0$ via 10-fold cross-validation. As a result, the method reduces to the Linear smoothing approach of Arnone et al. (2019). For brevity, we do not report the estimated fields. However, the second row of Table 1 reports the RMSE, showing that the Nonlinear and Linear approaches, which coincide in this setting, outperform TPS. Additional comparisons between the Linear and TPS approaches are provided in Arnone et al. (2019).

5.3 Simulation 3: areal observations in space, nonlinear dynamics

Simulation 3 uses the same test field as in Section 5.1, but adopts an areal sampling design. The domain is partitioned into 225 spatial regions, and observations are generated at each time instant by averaging the

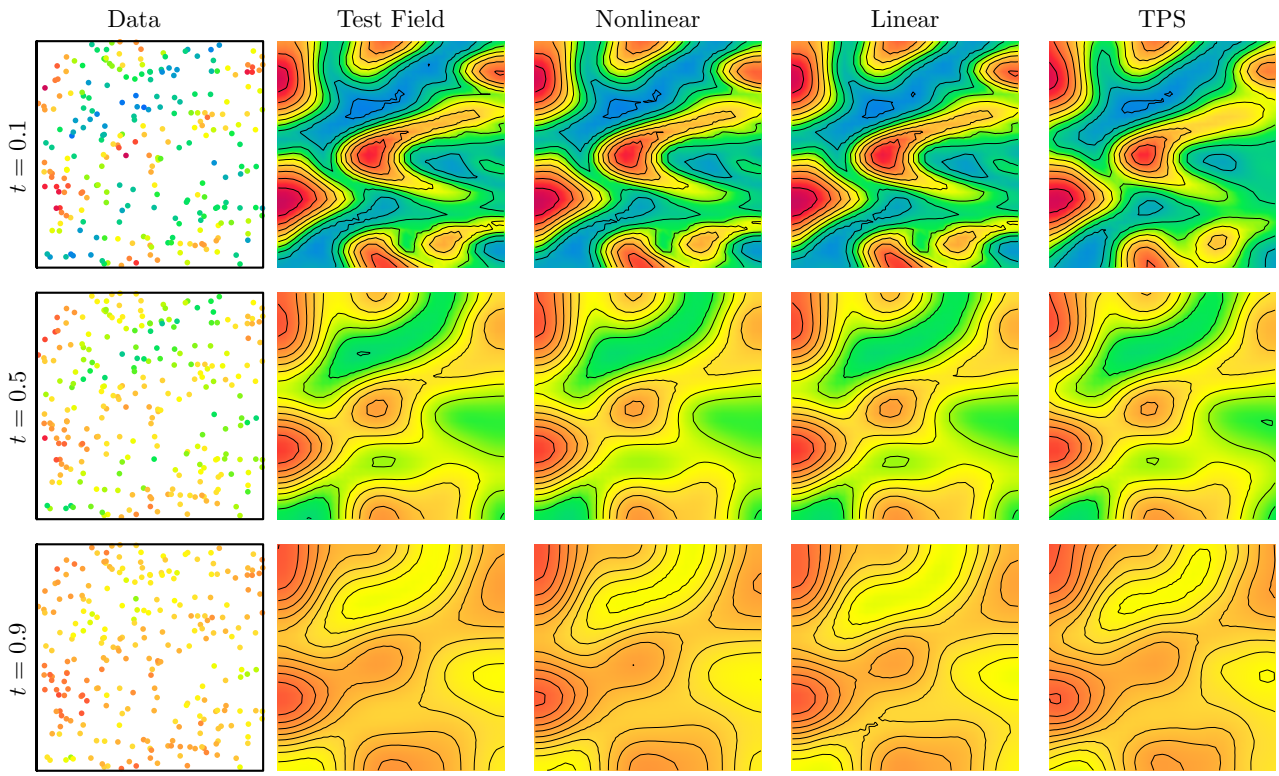


Figure 1: *Simulation 1: Pointwise observations in space and time, nonlinear dynamics.* From left to right: synthetic data, test field, and mean estimates over 30 simulation repetitions obtained with the proposed space-time regression with nonlinear PDE regularization (Nonlinear), its linear counterpart (Linear), and spatio-temporal Thin-Plate Splines (TPS). The Supplementary Material reports the corresponding plots for all time instants.

solution f of Equation (12), setting $r = 1$, over each region. Temporal sampling remains pointwise, with observations collected at the same $m = 10$ time instants as in the previous simulations. Independent Gaussian noise with standard deviation 0.05 is added, and the data generation process is repeated 30 times. The first column of Fig. 2 shows the data for one simulation replicate.

The implementation of Thin-Plate Splines in the R package `mgcv` cannot handle areal spatial observations, and therefore TPS is excluded from this setting. As in the previous simulations, the Nonlinear approach consistently selects the correct value of the reaction parameter ($r = 1$) via 10-fold cross-validation. The third and fourth columns of Fig. 2 show the mean estimated fields from the two competing methods over 30 repetitions. The proposed method again better captures the spatio-temporal structure of the underlying process, as confirmed by the last row of Table 1, which reports the mean (standard deviation) of the RMSE over 30 simulation repetitions.

Finally, Section B.1 of the Supplementary Material presents an additional simulation study with the same setting as the one considered here, but with unevenly spaced areas, obtained by randomly selecting 125 regions from those used here. The results are consistent with those reported above.

Table 1: The table reports the mean (standard deviation) of the RMSE over 30 simulation repetitions for the three simulation studies, for the three competing methods: space-time regression with nonlinear PDE regularization (Nonlinear), its linear counterpart (Linear), and spatio-temporal Thin-Plate Splines (TPS). In Simulation 2 (linear dynamics, $r = 0$), the proposed Nonlinear approach correctly selects a null reaction coefficient in all repetitions and thus coincides with the Linear approach.

	Sampling	Dynamic	Nonlinear	Linear	TPS
1	Pointwise	Nonlinear	**0.0042** (3e-4)	0.0136 (7e-4)	0.0218 (0.0013)
2	Pointwise	Linear	**0.0041** (3e-4)	**0.0041** (3e-4)	0.0217 (0.0013)
3	Areal	Nonlinear	**0.0044** (5e-4)	0.0134 (7e-4)	—

6 Application to longitudinal neuroimaging data

We apply the proposed model to neuroimaging data from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). Alzheimer's disease is characterized by the progressive accumulation and spatial spread of misfolded toxic proteins in the brain, most notably tau proteins, whose concentration is closely associated with disease progression. Longitudinal tau Positron Emission Tomography (PET) imaging provides in vivo measurements of tau burden and is therefore commonly used as a proxy for the spatio-temporal evolution of the disease.

From a mathematical perspective, the interplay between local accumulation and spatial propagation of tau pathology can be described by reaction-diffusion models, in particular the Fisher-KPP equation, which has been widely used to model the spread of neurodegenerative processes in Alzheimer's disease and related disorders (see, e.g., Weickenmeier et al., 2018; Fornari et al., 2019; Kevrekidis et al., 2020; Schäfer et al., 2021, 2022; Corti et al., 2023).

In this application, we consider longitudinal tau PET scans from a subject classified as Mild Cognitive Impairment (MCI) at the initial visit, a clinical condition corresponding to the prodromal stage preceding dementia. The proposed method is well suited to this setting, as it enables direct estimation of the spatio-temporal dynamics of tau protein concentration from noisy longitudinal observations while incorporating a nonlinear dynamical model through the regularization term. Although several numerical approaches have been proposed to simulate such dynamics (see, e.g., Weickenmeier et al., 2019; Corti et al., 2023), they primarily focus on forward simulation and do not directly integrate longitudinal PET observations into the modeling.

The ADNI dataset provides longitudinal raw tau PET scans, which constitute the functional signals of interest, together with corresponding T1-weighted Magnetic Resonance Imaging (MRI) images. The latter offer high anatomical contrast, clearly distinguishing gray matter, white matter, and cerebrospinal fluid, and enable tissue segmentation and the definition of the complex spatial domain where the data are observed, namely the highly non-convex volume of the brain gray matter. Longitudinal PET analyses typically assess temporal changes through a static summary measure, the Standardized Uptake Value Ratio (SUVR), computed at each time instant. SUVR is obtained by normalizing the PET signal by the injected radiotracer dose and the mean signal in a selected reference region. ADNI directly provides SUVR images using cerebellar gray matter as

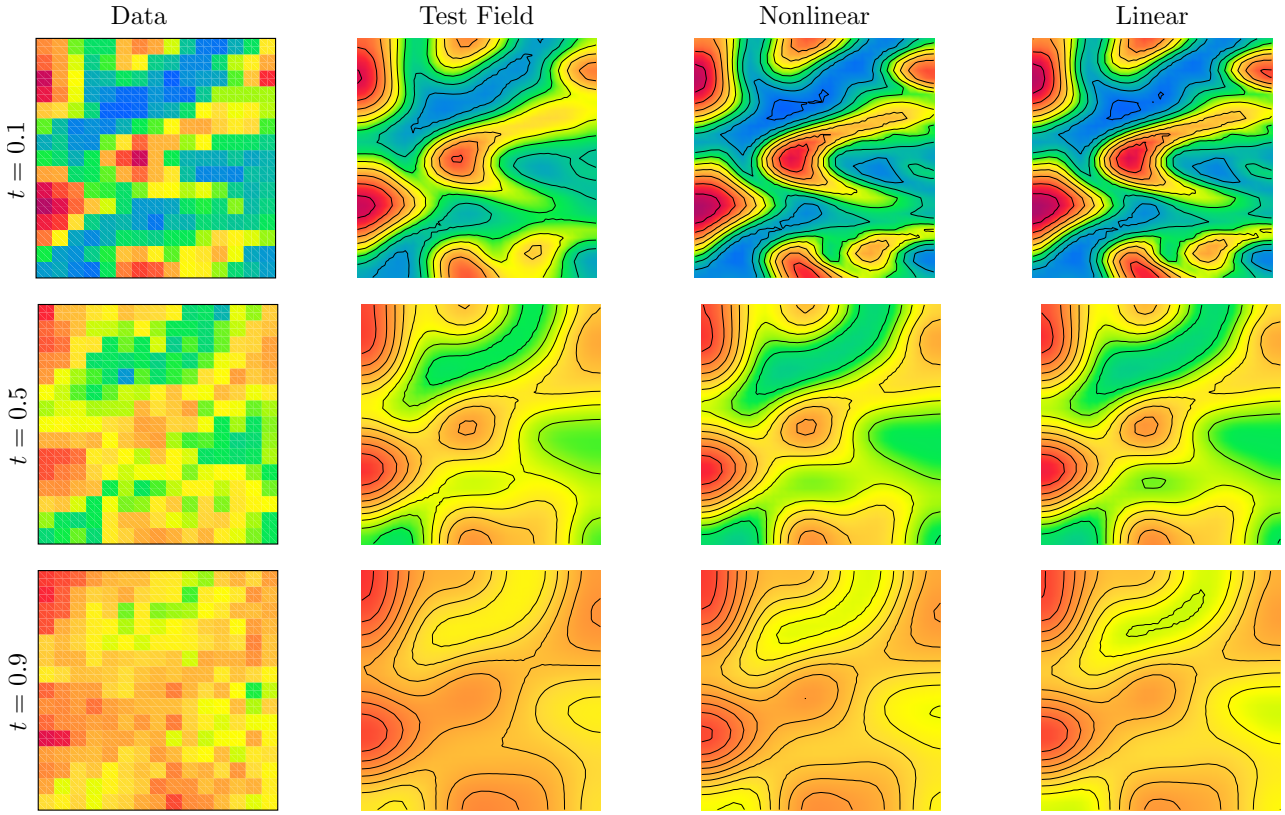


Figure 2: *Simulation 3: Areal observations in space and pointwise observations in time, nonlinear dynamics.* From left to right: synthetic data, test field, and estimates obtained with the proposed space-time regression with nonlinear PDE regularization (Nonlinear) and its linear counterpart (Linear). The Supplementary Material reports the corresponding plots for all time instants.

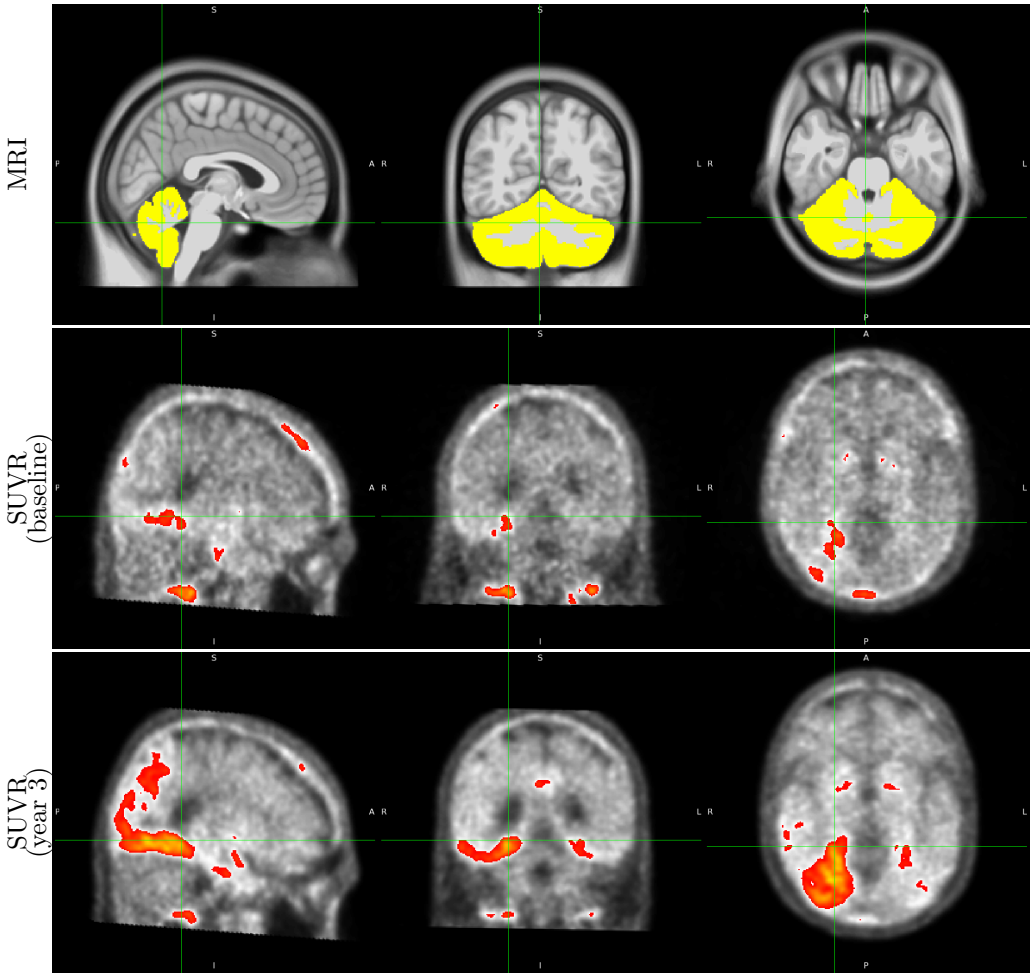


Figure 3: Top: from left to right, sagittal, coronal, and axial slices of the reference T1-weighted Magnetic Resonance Image (MRI) in the MNI152 reference space, with the cerebellar gray matter highlighted in yellow and used as the reference region for Standardized Uptake Value Ratio (SUVR) computation. Center and bottom: corresponding SUVR images, where red hotspots indicate regions of abnormally high tau protein accumulation relative to the reference region at baseline and after three years, respectively.

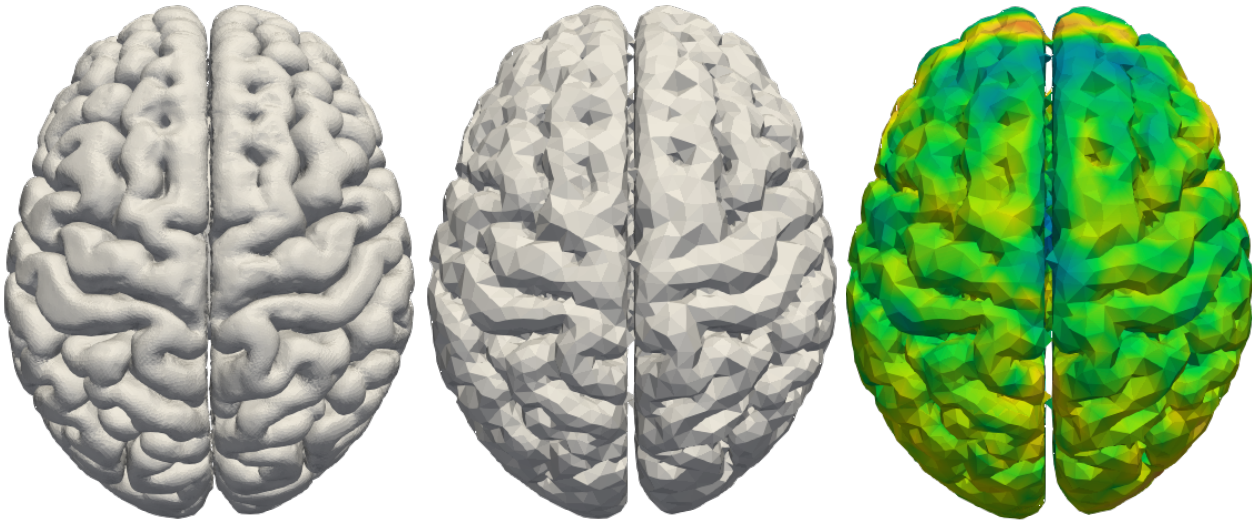


Figure 4: From left to right: refined brain surface derived from the MNI152 template; volumetric brain mesh obtained from this surface using SVMTK; SUVR data mapped onto the mesh.

the reference region. The top row of Fig. 3 shows slices of the reference MRI, highlighting the complex brain anatomy and the cerebellar gray matter in yellow. The central and bottom rows display the corresponding SUVR images at baseline visit and after three years, respectively. Regions shown in red indicate areas of abnormally high tau accumulation relative to the reference region. Prior to analysis, each SUVR image is registered to the subject-specific MRI and mapped to the MNI152 reference space (Fonov et al., 2011, 2009), a high-resolution template brain created by averaging 152 healthy adult brain MRIs that have been nonlinearly registered to a common coordinate system, using tools from the FSL software package (Jenkinson et al., 2012).

To construct the computational domain, the MNI152 template is segmented using FreeSurfer (Fischl, 2012), and the resulting high-resolution brain surface is volumetrically meshed using the SVMTK library (Mardal et al., 2022). The final mesh consists of approximately 10,000 nodes and 150,000 tetrahedral elements, balancing geometric fidelity and computational efficiency. Fig. 4 shows the refined brain surface (left) and the corresponding volumetric mesh (center). Although SUVR data are originally recorded at the voxel level, they are treated here as pointwise observations at the mesh nodes, as the mesh resolution is sufficiently fine to accurately represent voxel-wise values. The right panel of Fig. 4 displays the SUVR data mapped onto the mesh at the initial time instant.

We apply the proposed nonlinear smoothing method, selecting the reaction coefficient via 10-fold cross-validation over the grid $\{0.00, 0.05, 0.25, 0.45, 0.65, 0.85\}$. Fig. 5 reports boxplots of the cross-validation error for the candidate values, identifying $\hat{r} = 0.25$/year as the selected value. Notably, the method selects a nonlinear dynamics regime, consistent with the pathological condition of the subject and with existing findings on the spatio-temporal progression of tau pathology in Alzheimer's disease (see, e.g., Schäfer et al., 2021, and references therein). This result underscores the importance of explicitly accounting for nonlinear dynamics in the regularization term. Fig. 6 shows sagittal, coronal, and horizontal views of the estimated tau concentration field at the initial and final time instants. The horizontal views (third column) clearly illustrate the progressive accumulation of tau protein: regions with elevated concentration at baseline, such as the right posterior hemisphere, not only intensify over time but also spatially expand. The sagittal and coronal views provide complementary perspectives of the same evolving pattern. These regions are consistent with the hotspots observed in the SUVR images in the bottom panels of Fig. 3, highlighting the ability of the proposed method to recover clinically meaningful patterns while providing a refined and continuous characterization of the spatio-temporal progression of the disease.

7 Conclusions

In this work, we extend the class of penalized physics-informed smoothing models reviewed in Sangalli (2021) to accommodate data with nonlinear dynamics. This extension broadens the scope of these models, enhancing their ability to represent complex phenomena governed by nonlinear spatio-temporal processes and enabling their application to challenging real-world problems, as illustrated by the analysis of neuroimaging data. Simulation studies show that the proposed methodology accurately captures nonlinear dynamics, while naturally reducing to the physics-informed linear smoothing framework of Arnone et al. (2019) when the underlying dynamics are linear.

Several extensions of the proposed framework merit further investigation. A natural direction is the incorporation of spatio-temporally varying covariates within a semiparametric setting, following Sangalli et al. (2013). From a theoretical standpoint, it would also be of interest to extend the results for the estimation problem (2)

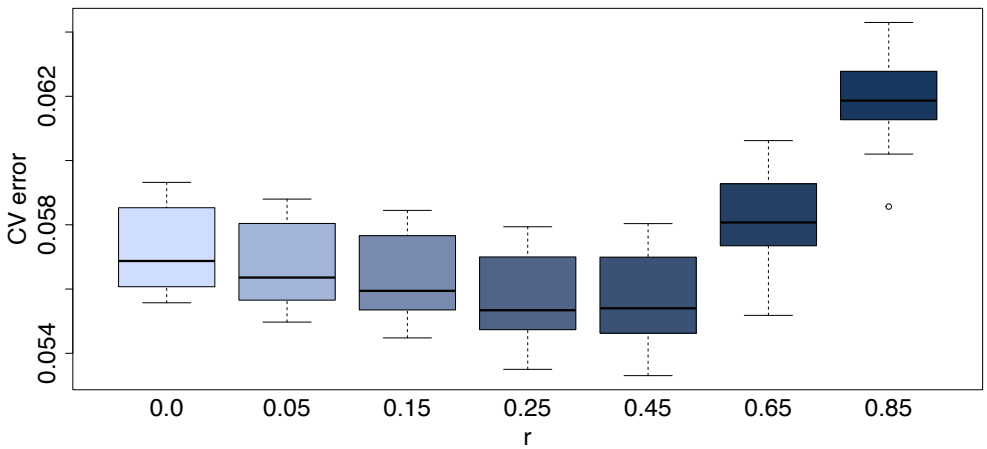


Figure 5: Boxplots of 10-fold cross-validation errors for the selection of the reaction parameter r. The selected value, $\hat{r} = 0.25$, corresponds to a nonlinear dynamics regime.

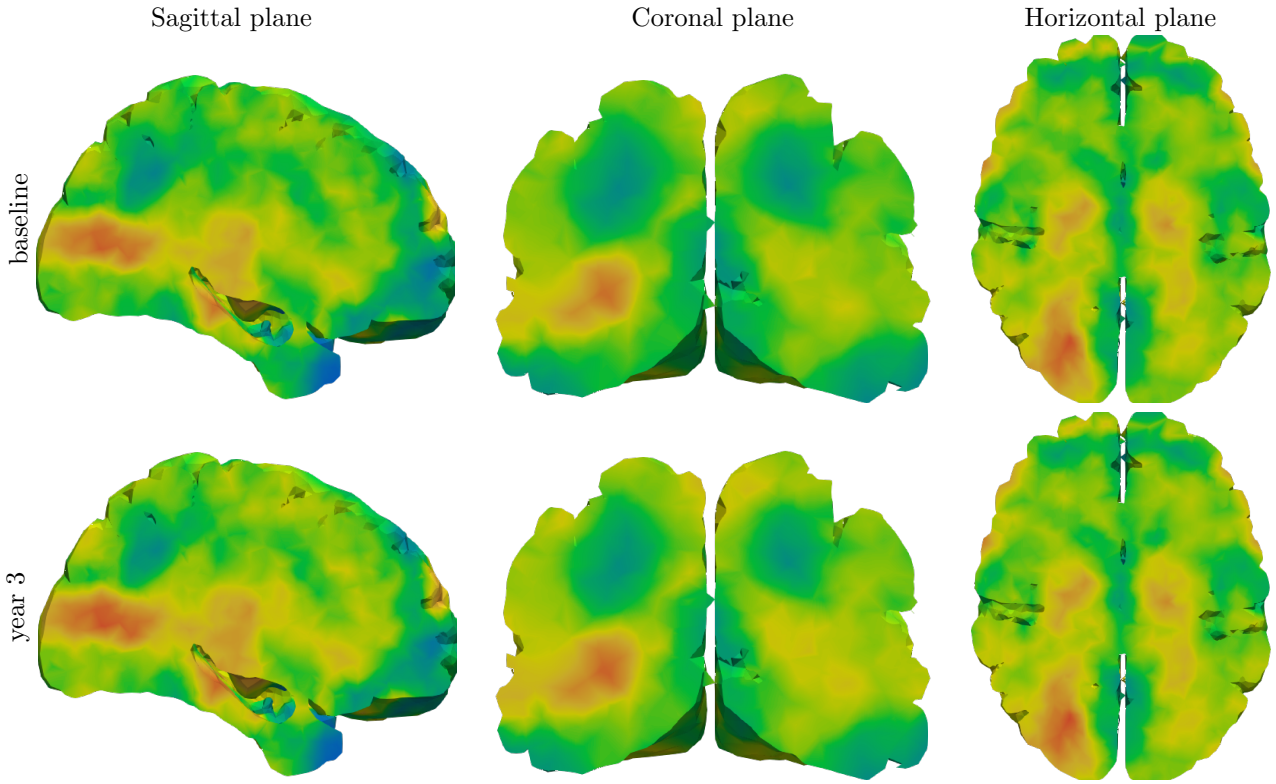


Figure 6: From left to right, sagittal, coronal, and horizontal views of the estimated spatio-temporal tau concentration at two time instants, baseline and year 3. The reddish region in the right-posterior hemisphere highlights an area of elevated tau concentration, which progressively expands and intensifies over time. The three views provide complementary perspectives of the same evolving pattern.

to settings where the quadratic loss is replaced by a nonlinear loss function, such as those arising in generalized linear models (Wilhelm and Sangalli, 2016) or quantile regression (Castiglione et al., 2025; De Sanctis et al., 2025; Di Battista et al., 2026), further broadening the applicability of the methodology. Another important direction concerns the estimation of latent parameters governing the underlying dynamics, such as the reaction coefficient in the Fisher-KPP model (Abboud et al., 2019). A promising approach is the parameter cascading algorithm introduced by Ramsay et al. (2007); Cao and Ramsay (2010); Campbell et al. (2012); Ramsay and Hooker (2017) for ODEs and extended to linear PDEs by Xun et al. (2013). This strategy has been successfully adapted to smoothing problems with linear PDE regularization by Bernardi et al. (2018); Tomasetto et al. (2024). A further challenging direction is the development of smoothing methods based on regularization terms derived from coupled systems of differential equations, including coupled ODEs, coupled PDEs, and mixed ODE-PDE systems. The nonlinear optimal control formulation introduced in Section 3 provides a solid theoretical foundation for addressing these extensions.

Finally, in the neuroimaging context, an interesting extension is the integration of additional imaging modalities. In particular, Diffusion Tensor Imaging (DTI) provides information on the local principal directions and magnitude of water diffusion in the brain, which could be incorporated into the model through a space-varying diffusion tensor in the penalty term. This would allow the methodology to account for nonstationary anisotropic diffusion and further highlight its flexibility in multimodal neuroimaging analyses.

Acknowledgments

We are grateful to the Editor and Reviewers for their insightful comments and suggestions, which have significantly improved the quality and clarity of the manuscript. A. Clemente and A. Palummo acknowledge funding by the European Union - NextGenerationEU programme. This reserach has been partially funded by MUR research project Dipartimento di Eccellenza 2023-2027, Dipartimento di Matematica, Politecnico di Milano, Italy.

The simulations presented in this paper were partially carried out on the HPC cluster of the Department of Mathematics at Politecnico di Milano.

Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award

number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California

Appendix A Proofs

Let V, H and V^* denote the spaces $H^1(\mathcal{D})$, $L^2(\mathcal{D})$, and $(H^1(\mathcal{D}))^*$, respectively. We start by showing that the state equation can be written in operator form.

For almost every $t \in (0,T)$, the following weak formulation holds:

$$\langle \partial_t f(t) + Af(t) + B\left[-\Phi(f(t)) - g(t)\right], v\rangle_* = 0 \quad \forall v \in H^1(\mathcal{D}),$$

where $\langle\cdot,\cdot\rangle_*$ denotes the duality pairing between $H^1(\mathcal{D})$ and $H^1(\mathcal{D})^*$; $A : H^1(\mathcal{D}) \to H^1(\mathcal{D})^*$ is the continuous isomorphism associated with Lf and specifically with the bilinear form $a(f,v) = \langle Af, v\rangle_*$; $\Phi : \mathcal{V} \to \mathcal{U}$ is the operator associated with the nonlinear term (here $\Phi(f) = f(1-f)$); and finally $B : L^2(\mathcal{D}) \to H^1(\mathcal{D})^*$ is the embedding operator defined by $\langle Bu, v\rangle_* = \int_{\mathcal{D}} uv$. Let W denote the space $L^2(0,T;V^*)$ and define $Z = W \times H$.

We prove the results considering the linear operator L to be the Laplacian, i.e., $Lf = -\Delta f$. However, the proofs hold for any second-order linear operator L (Tröltzsch, 2010).

Proof of Lemma 1. By the maximum principle, the solution f satisfies $0 \le f \le 1$ (see, e.g., Section 3.2 in Salsa, 2016). The operator $\Phi : \mathcal{V} \to \mathcal{U}$ associated with the nonlinearity, defined by $\Phi(f) = f(1-f)$, is locally Lipschitz continuous on every bounded subset of $\mathbb{R}$. Therefore, the existence and uniqueness of a weak solution follow from the general semilinear parabolic theory (see, e.g., Theorem 5.5 in Tröltzsch, 2010). □

Proof of Lemma 2. Let $g, v \in \mathcal{U}$, and let $f(g+v)$ and $f(g)$ denote the solutions corresponding to $g+v$ and g, respectively. Define $h = f(g+v) - f(g)$ as their difference. The function h satisfies the following problem:

$$\begin{cases} \partial_t h - \Delta h + (c(g,v) - 1)h = v, & \text{in } \mathcal{D} \times (0,T), \\ \partial_n h = 0, & \text{on } \partial\mathcal{D} \times (0,T), \\ h(\boldsymbol{p}, 0) = 0, & \text{in } \mathcal{D}, \end{cases}$$

where, by Lemma 1, $c(g,v) = f(g+v) + f(g) \in L^\infty(\mathcal{D})$ for almost every $t \in (0,T)$. By the standard abstract parabolic problem (APP) theory (see, e.g., Chapter 10 of Salsa, 2016), there exists a unique weak solution $h \in \mathcal{V}$ to the above problem, satisfying the bound

$$\|h\|_{\mathcal{V}}^2 \le c_2 \|v\|_{\mathcal{U}}^2. \tag{13}$$

Moreover, we have

$$\Phi\big(f(g+v)\big) - \Phi\big(f(g)\big) = (1 - 2f(g))\, h + E(g,v), \quad \text{with } \|E(g,v)\|_{\mathcal{U}} = o(\|h\|_{\mathcal{U}}).$$

Combining this with (13), we get $\|E(g,v)\|_{\mathcal{U}} = o(\|v\|_{\mathcal{U}})$.

We can decompose h as the sum of two terms: $z \in \mathcal{V}$, which solves problem (5), and $e \in \mathcal{V}$, which satisfies

$$\begin{cases} \partial_t e - \Delta e + (1 - 2f(g))e = -E(g,v), & \text{in } \mathcal{D} \times (0,T), \\ \partial_n e = 0, & \text{on } \partial\mathcal{D} \times (0,T), \\ e(\boldsymbol{p}, 0) = 0, & \text{in } \mathcal{D}. \end{cases}$$

By the APP theory, there exists a unique solution $e \in \mathcal{V}$ satisfying

$$\|e\|_{\mathcal{V}}^2 \leq c_3 \|E(g,v)\|_{\mathcal{U}}^2 = o\big(\|v\|_{\mathcal{U}}^2\big)\,.$$

We have thus shown that, for every $g \in \mathcal{U}$, there exists a bounded linear operator $f'(g) : \mathcal{U} \to \mathcal{V}$ such that, for every $v \in \mathcal{U}$,

$$\frac{\|f(g+v) - f(g) - f'(g)v\|_{\mathcal{V}}}{\|v\|_{\mathcal{U}}} \to 0, \quad \|v\|_{\mathcal{U}} \to 0.$$

Hence, the control-to-state map $g \mapsto f(g)$ is Fréchet differentiable. □

Proof of Proposition 1. We divide the proofs in two parts: the first part proves that an optimal pair, solution to problem (4), exists and the second part states the corresponding first-order optimality conditions.

Existence of an optimal pair. We here leverage a well-known result in optimal control theory, that ensures the existence of a solution to problem (4). We report it here for completeness (see, e.g., Theorem 9.4 in Manzoni et al., 2021).

Theorem 1 (Existence of an optimal pair)**.** *Let*

$$\mathcal{O} = \{(f,g) \in \mathcal{V} \times \mathcal{U} \ : \ G(f,g) = 0 \ \text{in } Z\} \subset \mathcal{V} \times \mathcal{U}$$

be the set of feasible points. Under the following assumptions:

1. $\inf\limits_{(f,g)\in\mathcal{V}\times\mathcal{U}} J(f,g) = \mu > -\infty$ *and* $\mathcal{O} \neq \varnothing$;
2. *every minimizing sequence* $\{(f_k, g_k)\}$ *is bounded in* $\mathcal{V} \times \mathcal{U}$;
3. $\mathcal{O}$ *is weakly sequentially closed in* $\mathcal{V} \times \mathcal{U}$;
4. J *is sequentially weakly lower semicontinuous;*

there exists at least one optimal solution $(f,g) \in \mathcal{O}$ *to problem* (4).

Hence, we now need to verify the assumptions of Theorem 1.

1. The first assumption is trivially satisfied by leveraging Lemma 1.

2. Let $\{(f_k, g_k)\}$ be a minimizing sequence in $\mathcal{V} \times \mathcal{U}$. Then there exists a constant $C > 0$ such that $J(f_k, g_k) < C$ for all k. Moreover, since $\lambda > 0$, by construction we have $J(f_k, g_k) \geq \lambda \|g_k\|_{\mathcal{U}}^2$, which implies that $\{g_k\}$ is bounded in $\mathcal{U}$. Lemma 1 then ensures that $\{f_k\}$ is bounded in $\mathcal{V}$, yielding the desired result.

3. Let $\{(f_k, g_k)\} \subset \mathcal{O}$ be such that $f_k \rightharpoonup f$ in $\mathcal{V}$ and $g_k \rightharpoonup g$ in $\mathcal{U}$. We must show that $(f,g) \in \mathcal{O}$. We first prove that, for almost every $s \in (0,T)$, $(\Phi(f_k(s)) - \Phi(f(s)), v)_H \to 0$, $\forall v \in V$. Indeed, the following equation holds:

$$\int_{\mathcal{D}} (\Phi(f_k(s)) - \Phi(f(s)))\, v = \int_{\mathcal{D}} (f_k(s) - f(s))\, v \, d\boldsymbol{x} - \int_{\mathcal{D}} (f_k(s) + f(s))(f_k(s) - f(s))\, v \, d\boldsymbol{x}. \tag{14}$$

The first term in (14) tends to zero because $f_k \rightharpoonup f$ in H. Moreover, since $\{f_k\}$ is bounded in V, the Sobolev embedding theorem ensures that $f_k \to f$ strongly in $L^4(\mathcal{D})$. Using Hölder's inequality with exponents $1/4, 1/4, 1/2$, we obtain

$$\Big| \int_{\mathcal{D}} (f_k(s) + f(s))(f_k(s) - f(s))\, v \, d\boldsymbol{x} \Big| \leq C \, \|f_k(s) - f(s)\|_{L^4(\mathcal{D})} \|v\|_{L^2(\mathcal{D})},$$

which tends to zero as $k \to \infty$. Hence, $(\Phi(f_k(s)) - \Phi(f(s)), v)_H \to 0$, $\forall v \in V$.
Finally, for almost every $s \in (0,T)$,

$$\begin{aligned}\langle G(f_k(s), g_k(s)) - G(f(s), g(s)), v\rangle_* = &\int_{\mathcal{D}} \partial_t (f_k(s) - f(s))\, v \, d\boldsymbol{x} + \int_{\mathcal{D}} \nabla (f_k(s) - f(s)) \cdot \nabla v \, d\boldsymbol{x} \\ &+ \int_{\mathcal{D}} (\Phi(f_k(s)) - \Phi(f(s)))\, v \, d\boldsymbol{x} + \int_{\mathcal{D}} (g_k(s) - g(s))\, v \, d\boldsymbol{x}.\end{aligned}$$

All the terms in the previous expression converge to zero, and therefore, for almost every $s \in (0,T)$,

$$0 = G(f_k(s), g_k(s)) \rightharpoonup G(f(s), g(s)) \quad \text{in } V^*.$$

Integrating over $(0,T)$ yields the desired result, i.e., $(f,g) \in \mathcal{O}$.

4. J is continuous on $\mathcal{V} \times \mathcal{U}$, and hence sequentially weakly lower semicontinuous. □

First order optimality conditions. In this part of the proof, we state the first-order optimality conditions, which are discretized in Section 4 to obtain a numerical approximation of the solution to the estimation problem, leveraging the following theorem (see Manzoni et al., 2021).

Theorem 2 (First-order optimality conditions)**.** *Let $(\hat{f},\hat{g})$ be a solution to problem (4). Under the following assumptions:*

1. *$J : \mathcal{V} \times \mathcal{U} \to \mathbb{R}$ and $G : \mathcal{V} \times \mathcal{U} \to Z$ are Fréchet differentiable;*
2. *for each g, the state equation $G(f,g) = 0$ in Z defines a unique control-to-state map $g \mapsto f(g)$, i.e., $G(f(g),g) = 0$ for any $g \in \mathcal{B}_{\mathcal{U}}$; moreover, this map is Fréchet differentiable;*
3. *the partial derivative $G'_f(f(g),g) \in \mathcal{L}(\mathcal{V},Z)$ is a continuous isomorphism between $\mathcal{V}$ and Z for every $g \in \mathcal{B}_{\mathcal{U}}$. Equivalently, the adjoint operator $G'_f(f(g),g)^* \in \mathcal{L}(Z^*,\mathcal{V}^*)$ is a continuous isomorphism between Z^* and $\mathcal{V}^*$.*

Then, there exists a unique multiplier $\hat{p} \in \mathcal{V}$ such that $\hat{f} = f(\hat{g})$, and the triplet $(\hat{f},\hat{g},\hat{p})$ satisfies the following optimality system:

$$\begin{cases} G(\hat{f},\hat{g}) = 0, & \textit{State equation,} \\ G'_f(\hat{f},\hat{g})^*\hat{p} = -J'_f(\hat{f},\hat{g}), & \textit{Adjoint equation,} \\ J'_g(\hat{f},\hat{g}) + G'_g(\hat{f},\hat{g})^*\hat{p} = 0, & \textit{Variational equation.} \end{cases} \tag{15}$$

We now verify the assumptions of Theorem 2 to prove this part of the Proposition.

1. The mappings $J : \mathcal{V} \times \mathcal{U} \to \mathbb{R}$ and $G : \mathcal{V} \times \mathcal{U} \to Z$ are Fréchet differentiable. Indeed, $\forall \varphi \in \mathcal{V}$ and $\forall u \in \mathcal{U}$ we have

$$\begin{aligned} J'_f(f,g)\varphi &= -\frac{1}{nm}\sum_{i,j}(y_{ij} - C_{ij}f)\,C_{ij}\varphi, & J'_g(f,g)u &= (g,u)_{\mathcal{U}}, \\ G'_f(f,g)\varphi &= \begin{pmatrix} \partial_t\varphi - \Delta\varphi - B[\Phi'(f)\varphi] \\ \varphi(0) \end{pmatrix}, & G'_g(f,g)u &= \begin{pmatrix} -Bu \\ 0 \end{pmatrix}. \end{aligned} \tag{16}$$

2. This point follows directly from Lemmas 1 and 2, which ensure that the control-to-state map is well-defined and $\mathcal{J}$-differentiable.

3. Let $v \in \mathcal{U}$. We must prove that the problem $G'_f(f,g)\varphi = Bv$ in Z admits a unique solution. Note that

$$G'_f(f,g)\varphi = \partial_t\varphi - \Delta\varphi - B[\Phi'(f)\varphi - v] = 0 \quad \text{in } Z,$$

corresponds to problem (5) for φ, written in operator form. Hence, $G'_f(f,g) \in \mathcal{L}(\mathcal{V},Z)$ is a continuous isomorphism, as guaranteed by standard results from the abstract parabolic theory (see, e.g., Salsa, 2016). Finally, following the argument of Chapter 9 in Salsa (2016), we have that a solution $p \in L^2(0,T;V)$ belongs to $\mathcal{V}$, completing the proof. □

References

Abboud, C., Bonnefon, O., Parent, E., Soubeyrand, S., 2019. Dating and localizing an invasion from post-introduction data and a coupled reaction–diffusion–absorption model. Journal of Mathematical Biology 79, 765–789.

Arnone, E., Azzimonti, L., Nobile, F., Sangalli, L.M., 2019. Modeling spatially dependent functional data via regression with differential regularization. Journal of Multivariate Analysis 170, 275–295.

Arnone, E., Clemente, A., Sangalli, L.M., Lila, E., Ramsay, J., Formaggia, L., 2026. fdaPDE: Physics-Informed Spatial and Functional Data Analysis. URL: https://CRAN.R-project.org/package=fdaPDE. r package version 1.1-24.

Arnone, E., Negri, L., Panzica, F., Sangalli, L.M., 2023a. Analyzing data in complicated 3d domains: Smoothing, semiparametric regression, and functional principal component analysis. Biometrics 79, 3510–3521.

Arnone, E., Sangalli, L.M., Vicini, A., 2023b. Smoothing spatio-temporal data with complex missing data patterns. Statistical Modelling 23, 327–356.

Azzimonti, L., Nobile, F., Sangalli, L.M., Secchi, P., 2014. Mixed finite elements for spatial regression with pde penalization. SIAM/ASA Journal on Uncertainty Quantification 2, 305–335.

Azzimonti, L., Sangalli, L.M., Secchi, P., Domanin, M., Nobile, F., 2015. Blood flow velocity field estimation via spatial regression with pde penalization. Journal of the American Statistical Association 110, 1057–1071.

Bernardi, M.S., Carey, M., Ramsay, J.O., Sangalli, L.M., 2018. Modeling spatial anisotropy via regression with partial differential regularization. Journal of Multivariate Analysis 167, 15–30.

Bowman, F.D., 2014. Brain imaging analysis. Annual Review of Statistics and Its Application 1, 61–85.

Brézis, H., 2011. Functional analysis, Sobolev Spaces and Partial Differential Equations. volume 2. Springer, New York.

Campbell, D.A., Hooker, G., McAuley, K.B., 2012. Parameter estimation in differential equation models with constrained states. Journal of Chemometrics 26, 322–332.

Cao, J., Ramsay, J., 2010. Linear mixed-effects modeling by parameter cascading. Journal of the American Statistical Association 105, 365–374.

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

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

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

Corti, M., Bonizzoni, F., Quarteroni, A.M., Antonietti, P.F., et al., 2023. Discontinuous galerkin methods for fisher–kolmogorov equation with application to α-synuclein spreading in parkinson's disease. Computer Methods in Applied Mechanics and Engineering 417, 116450.

Cox, D.D., 1984. Multivariate smoothing spline functions. Siam Journal on Numerical Analysis 21, 789–813.

De Sanctis, M.F., Di Battista, I., Arnone, E., Castiglione, C., Palummo, A., Bernardi, M., Ieva, F., Sangalli, L.M., 2025. Exploring nitrogen dioxide spatial concentration via physics-informed multiple quantile regression. Environmental and Ecological Statistics 32, 855–892.

Di Battista, I., De Sanctis, M.F., Arnone, E., Castiglione, C., Palummo, A., Sangalli, L.M., 2026. A semiparametric space-time quantile regression model. Journal of Nonparametric Statistics 38, 279–310.

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

Ferraty, F., Vieu, P., 2006. Nonparametric Functional Data Analysis: Theory and Practice. Springer, New York.

Fischl, B., 2012. Freesurfer. NeuroImage 62, 774–781. URL: https://www.sciencedirect.com/science/article/pii/S1053811912000389, doi:https://doi.org/10.1016/j.neuroimage.2012.01.021.

Fonov, V., Evans, A., McKinstry, R., Almli, C., Collins, D., 2009. Unbiased nonlinear average age-appropriate brain templates from birth to adulthood. Neuroimage 47, S102. URL: https://www.sciencedirect.com/science/article/pii/S1053811909708845, doi:https://doi.org/10.1016/S1053-8119(09)70884-5. organization for Human Brain Mapping 2009 Annual Meeting.

Fonov, V., Evans, A.C., Botteron, K., Almli, C.R., McKinstry, R.C., Collins, D.L., 2011. Unbiased average age-appropriate atlases for pediatric studies. Neuroimage 54, 313–327. URL: https://www.sciencedirect.com/science/article/pii/S1053811910010062, doi:https://doi.org/10.1016/j.neuroimage.2010.07.033.

Fornari, S., Schäfer, A., Jucker, M., Goriely, A., Kuhl, E., 2019. Prion-like spreading of alzheimer's disease within the brain's connectome. Journal of the Royal Society Interface 16, 20190356.

Frackowiak, R.S., Friston, K.J., Frith, C.D., Dolan, R.J., Mazziotta, J., 2004. Human brain function .

Green, P.J., Silverman, B.W., 1994. Nonparametric Regression and Generalized Linear Models. volume 58. Chapman & Hall, London.

Härdle, W., 1990. Applied Nonparametric Regression. 19, Cambridge University Press, Cambridge, UK.

Hooten, M.B., Wikle, C.K., 2008. A hierarchical bayesian non-linear spatio-temporal model for the spread of invasive species with application to the eurasian collared-dove. Environmental and Ecological Statistics 15, 59–70.

Horváth, L., Kokoszka, P., 2012. Inference for Functional Data with Applications. Springer Series in Statistics, Springer, New York.

Jenkinson, M., Beckmann, C.F., Behrens, T.E., Woolrich, M.W., Smith, S.M., 2012. FSL. Neuroimage 62, 782–790. URL: https://www.sciencedirect.com/science/article/pii/S1053811911010603, doi:https://doi.org/10.1016/j.neuroimage.2011.09.015.

Kevrekidis, P., Thompson, T.B., Goriely, A., 2020. Anisotropic diffusion and traveling waves of toxic proteins in neurodegenerative diseases. Physics Letters a 384, 126935.

Kokoszka, P., Reimherr, M., 2017. Introduction to Functional Data Analysis. Chapman and Hall/CRC, Boca Raton, FL.

Lai, M.J., Schumaker, L.L., 2007. Spline Functions on Triangulations. 110, Cambridge University Press, Cambridge, UK.

Lila, E., Aston, J.A.D., Sangalli, L.M., 2016. Smooth Principal Component Analysis over two-dimensional manifolds with an application to neuroimaging. The Annals of Applied Statistics 10, 1854 – 1879. URL: https://doi.org/10.1214/16-AOAS975, doi:10.1214/16-AOAS975.

Lindgren, F., Rue, H., Lindström, J., 2011. An explicit link between gaussian fields and gaussian markov random fields: the stochastic partial differential equation approach. Journal of the Royal Statistical Society Series B: Statistical Methodology 73, 423–498.

Manzoni, A., Quarteroni, A., Salsa, S., 2021. Optimal Control of Partial Differential Equations. Springer, Cham.

Mardal, K.A., Rognes, M.E., Thompson, T.B., Valnes, L.M., 2022. Mathematical modeling of the human brain: from magnetic resonance images to finite element simulation. Springer Nature, Cham.

Nocedal, J., Wright, S.J., 1999. Numerical Optimization. Springer Series in Operations Research, Springer, New York.

Ombao, H., Lindquist, M., Thompson, W., Aston, J., 2016. Handbook of Neuroimaging Data Analysis. CRC Press, Boca Raton, FL.

Quarteroni, A., 2009. Numerical Models for Differential Problems. volume 2. Springer, Milan.

Ramsay, J., Hooker, G., 2017. Dynamic Data Analysis: Modeling Data with Differential Equations. Springer, New York, NY.

Ramsay, J., Silverman, B., 2005. Functional Data Analysis. Springer Series in Statistics, Springer.

Ramsay, J.O., Hooker, G., Campbell, D., Cao, J., 2007. Parameter estimation for differential equations: a generalized smoothing approach. Journal of the Royal Statistical Society Series B 69, 741–796.

Ramsay, T., 2002. Spline smoothing over difficult regions. Journal of the Royal Statistical Society Series B 64, 307–319.

Roques, L., Walker, E., Franck, P., Soubeyrand, S., Klein, E., 2016. Using genetic data to estimate diffusion rates in heterogeneous landscapes. Journal of Mathematical Biology 73, 397–422.

Salsa, S., 2016. Partial Differential Equations in Action: from Modelling to Theory. volume 99. Springer, Cham.

Sangalli, L.M., 2021. Spatial regression with partial differential equation regularisation. International Statistical Review 89, 505–531.

Sangalli, L.M., Ramsay, J.O., Ramsay, T.O., 2013. Spatial spline regression models. Journal of the Royal Statistical Society Series B 75, 681–703.

Schäfer, A., Chaggar, P., Goriely, A., Kuhl, E., Initiative, A.D.N., 2022. Correlating tau pathology to brain atrophy using a physics-based Bayesian model. Engineering with Computers 38, 3867–3877.

Schäfer, A., Peirlinck, M., Linka, K., Kuhl, E., (ADNI), A.D.N.I., 2021. Bayesian physics-based modeling of tau propagation in alzheimer's disease. Frontiers in Physiology 12, 702975.

Sengupta, U., Kayed, R., 2022. Amyloid β, tau, and α-synuclein aggregates in the pathogenesis, prognosis, and therapeutics for neurodegenerative diseases. Progress in Neurobiology 214, 102270.

Stroud, J.R., Stein, M.L., Lesht, B.M., Schwab, D.J., Beletsky, D., 2010. An ensemble kalman filter and smoother for satellite data assimilation. Journal of the American Statistical Association 105, 978–990.

Tomasetto, M., Arnone, E., Sangalli, L.M., 2024. Modeling anisotropy and non-stationarity through physics-informed spatial regression. Environmetrics , e2889.

Tröltzsch, F., 2010. Optimal Control of Partial Differential Equations: Theory, Methods, and Applications. volume 112. American Mathematical Society, Providence, RI.

Wahba, G., 1981. Spline interpolation and smoothing on the sphere. Siam Journal on Scientific and Statistical Computing 2, 5–16.

Wahba, G., 1990. Spline Models for Observational Data. volume 59. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA.

Weickenmeier, J., Jucker, M., Goriely, A., Kuhl, E., 2019. A physics-based model explains the prion-like features of neurodegeneration in Alzheimer's disease, Parkinson's disease, and amyotrophic lateral sclerosis. Journal of the Mechanics and Physics of Solids 124, 264–281.

Weickenmeier, J., Kuhl, E., Goriely, A., 2018. Multiphysics of prionlike diseases: Progression and atrophy. Physical Review Letters 121, 158101.

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

Wikle, C.K., Hooten, M.B., 2010. A general science-based framework for dynamical spatio-temporal models. Test 19, 417–451.

Wilhelm, M., Sangalli, L.M., 2016. Generalized spatial regression with differential regularization. Journal of Statistical Computation and Simulation 86, 2497–2518.

Wood, S.N., 2003. Thin plate regression splines. Journal of the Royal Statistical Society Series B 65, 95–114.

Wood, S.N., 2017. Generalized Additive Models: an Introduction with R. Chapman and Hall/CRC, Boca Raton, FL.

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

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

Zeidler, E., 2012. Nonlinear Functional Analysis and its Applications: II/A: Linear Monotone Operators. Springer, New York.

Zeidler, E., 2013. Nonlinear Functional Analysis and its Applications: II/B: Nonlinear Monotone Operators. Springer, New York.

Supplementary Material for Modeling spatio-temporal smoothness using nonlinear differential operators

Aldo Clemente[1] Alessandro Palummo[1] Eleonora Arnone[2] John A.D. Aston[3]
Ferruccio Panzica[4] Laura M. Sangalli[1*]

the Alzheimer's Disease Neuroimaging Initiative[†]

[1] MOX - Dipartimento di Matematica, Politecnico di Milano, Milano, Italy
[2] Dipartimento di Management, Università degli Studi di Torino, Torino, Italy
[3] Statistical Laboratory, University of Cambridge, Cambridge, UK
[4] Fondazione IRCCS Istituto Neurologico Carlo Besta, Milano, Italy

B Simulation studies

This section expands the simulation study presented in the main manuscript. We report the complete qualitative results for Simulations 1 and 3, that could not be included in the main paper due to space constraints; see Figures B.1 and B.2, respectively. Additionally, Section B.1 considers the same setting as Simulation 3, but with unevenly spaced areal spatial data. Section B.2 builds on the setting of Simulation 1 and evaluates the proposed method by systematically varying key parameters while keeping the others fixed.

B.1 Simulation 4: unevenly spaced areal observations, nonlinear dynamics

This simulation study uses the same setting as in Section 5.3, but considers unevenly spaced spatial areas, obtained by randomly selecting 125 spatial regions from those used in the main manuscript. We compare the proposed space-time regression with nonlinear PDE regularization (Nonlinear) and its linear counterpart (Linear). The implementation of Thin-Plate Splines in the R package `mgcv` cannot handle areal spatial observations and TPS is therefore excluded from this setting.

The first column of Figure B.3 displays the data for one simulation replicate. As in Simulations 1 and 3 of the main article, the Nonlinear approach consistently selects the correct value of the reaction parameter ($r = 1$) via 10-fold cross-validation. The third and fourth columns of Figure B.3 show the mean estimated fields from the two competing methods over 30 simulation repetitions.

The mean RMSE over 30 simulation repetitions is 0.0045 (sd 0.0005) for the proposed method and 0.0171 (sd 0.0009) for its linear counterpart, highlighting the superiority of the proposed approach in this setting.

B.2 Additional Simulation Experiments

This section reports additional simulation experiments to assess the robustness of the proposed approach under different data settings and smoothing configurations. Specifically, we extend Simulation Study 1 of the main manuscript by systematically varying key parameters while keeping the others fixed. Data are generated according to the same setting as in Simulation 1 of the main manuscript. The following scenarios are considered:

Case (a) *Variation in the reaction parameter r*: We vary r over the range 0.2 to 1.0 with step size 0.2, while keeping n, m, and N_h fixed.

Case (b) *Variation in the number of spatial locations n*: Fixing $r = 1$, we consider $n = 125, 250, 500, 1000, 2000$ spatial locations at each time instant. The number of time instants m and the discretization level N_h are kept fixed.

*Corresponding author: laura.sangalli@polimi.it

†Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (`https://adni.loni.usc.edu`). As such, the investigators within ADNI contributed to the design and implementation of ADNI and/or provided data, but did not participate in the analysis or writing of this report. A complete listing of ADNI investigators can be found at: `http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf`.

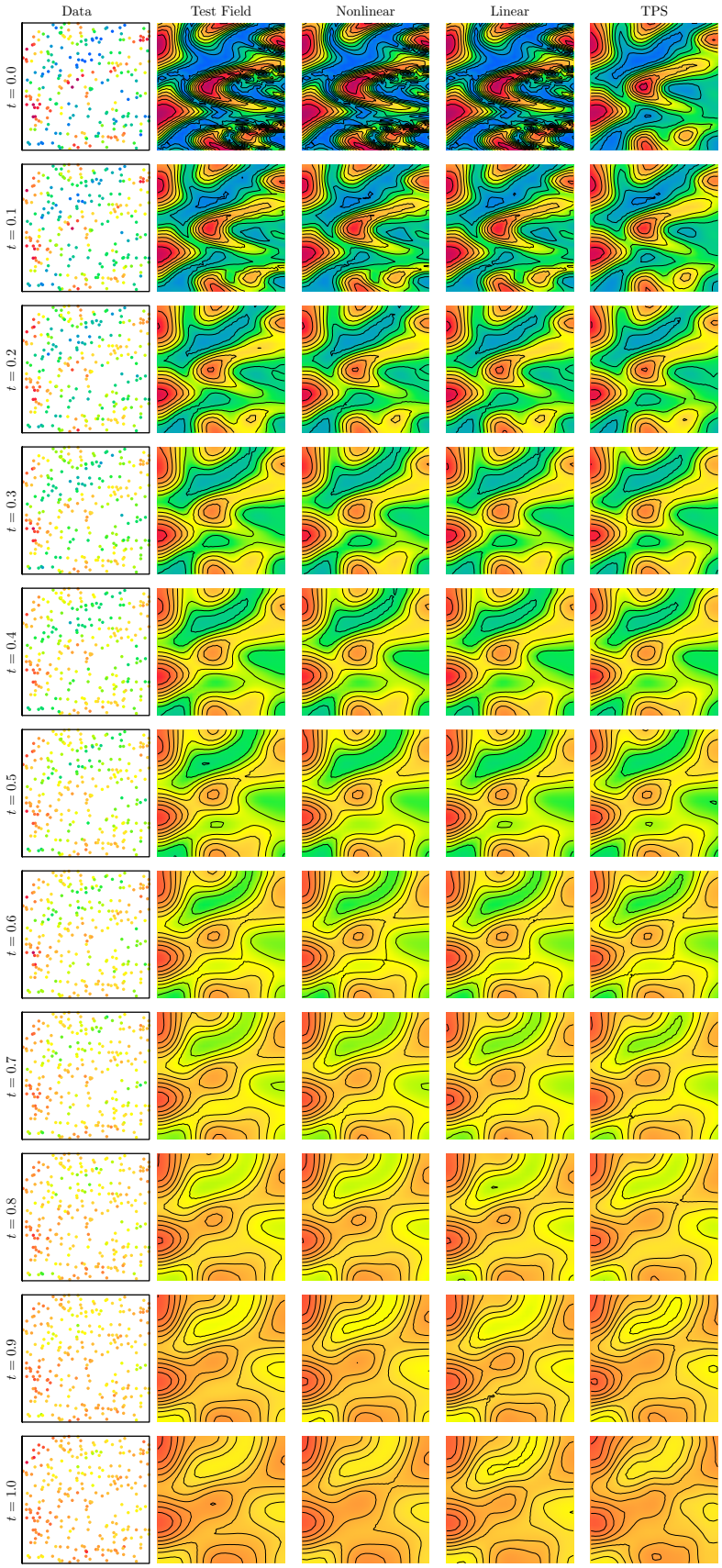


Figure B.1: *Simulation 1: Pointwise observations in space and time, nonlinear dynamics. From left to right: synthetic data, test field, and mean estimates over 30 simulation repetitions obtained with the proposed space-time regression with nonlinear PDE regularization (Nonlinear), its linear counterpart (Linear), and spatio-temporal Thin-Plate Splines (TPS). The figure reproduces Figure 1 from the main article, extended to all time instants.*

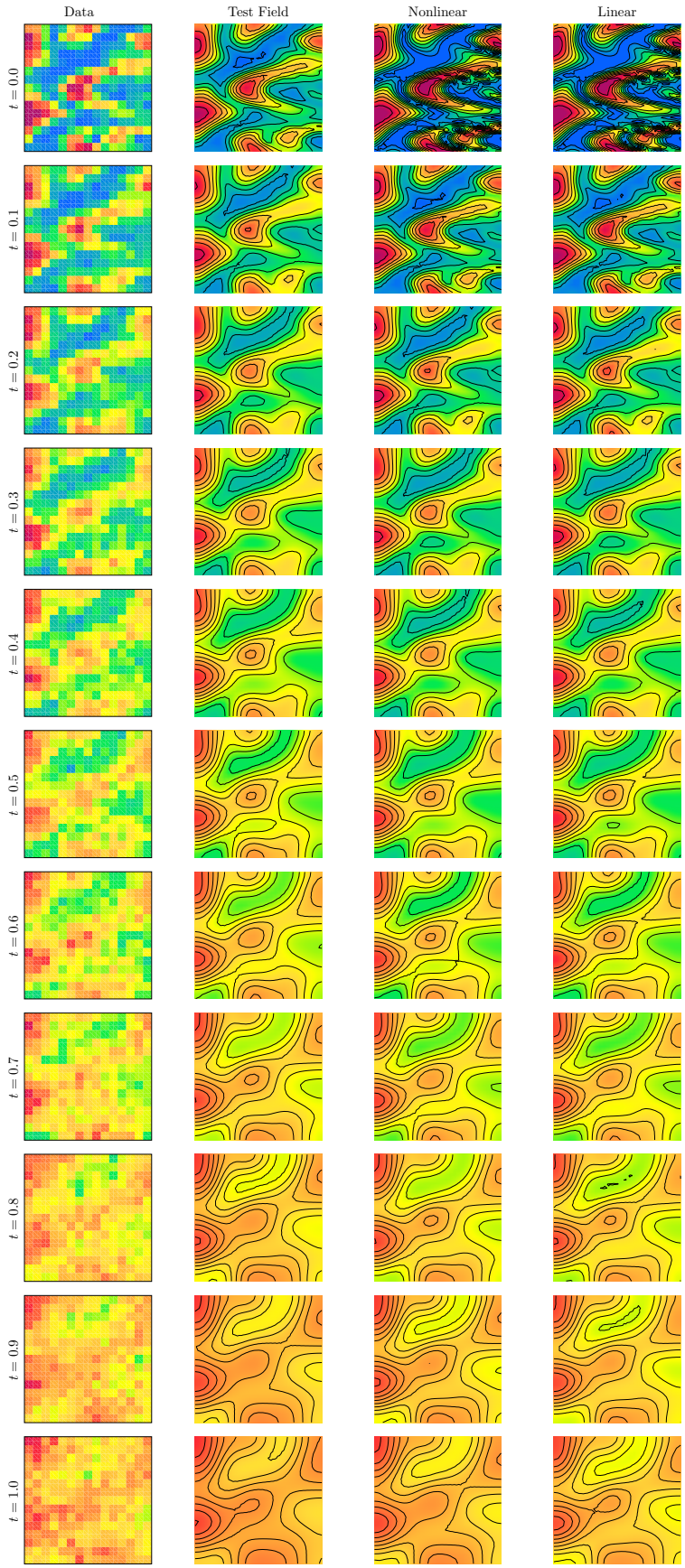


Figure B.2: *Simulation 3: Areal observations in space and pointwise observations in time, nonlinear dynamics. From left to right: synthetic data, test field, and estimates obtained with the proposed space-time regression with nonlinear PDE regularization (Nonlinear) and its linear counterpart (Linear). The figure reproduces Figure 2 from the main article, extended to all time instants.*

Case (c) *Variation in the number of time instants m*: Fixing $r = 1$, we consider $m = 5, 11, 21, 41$ time instants, while keeping $n = 250$ spatial locations per time instant and N_h fixed.

Case (d) *Variation in the number of mesh nodes N_h*: We consider different mesh resolutions corresponding to $N_h = 11^2, 21^2, 31^2, 41^2$ nodes, while keeping the remaining parameters fixed as in Simulation 1.

The performance of the methods is assessed in terms of the Root Mean Squared Error (RMSE), as detailed in Section 5 of the main article, computed with respect to the true underlying signal over 30 independent simulation repetitions, for both the proposed method and its linear counterpart.

Figure B.4 summarizes the results.

Case (a) (upper left panel) shows that the linear method performs progressively worse as the data are generated under increasingly strong nonlinear dynamics (i.e., for larger values of r), as expected. In contrast, the proposed nonlinear approach achieves consistently low RMSE across all values of r.

Cases (b) and (c) (upper right and bottom left panels) demonstrate that, with increasing data availability, both methods exhibit a decreasing trend in RMSE, illustrating the numerical consistency of the methods.

Finally, case (d) (bottom right panel) highlights that ultra-fine meshes, far beyond the signal resolution, are not necessary: once the mesh is sufficiently fine, further refinement does not lead to improvements in RMSE.

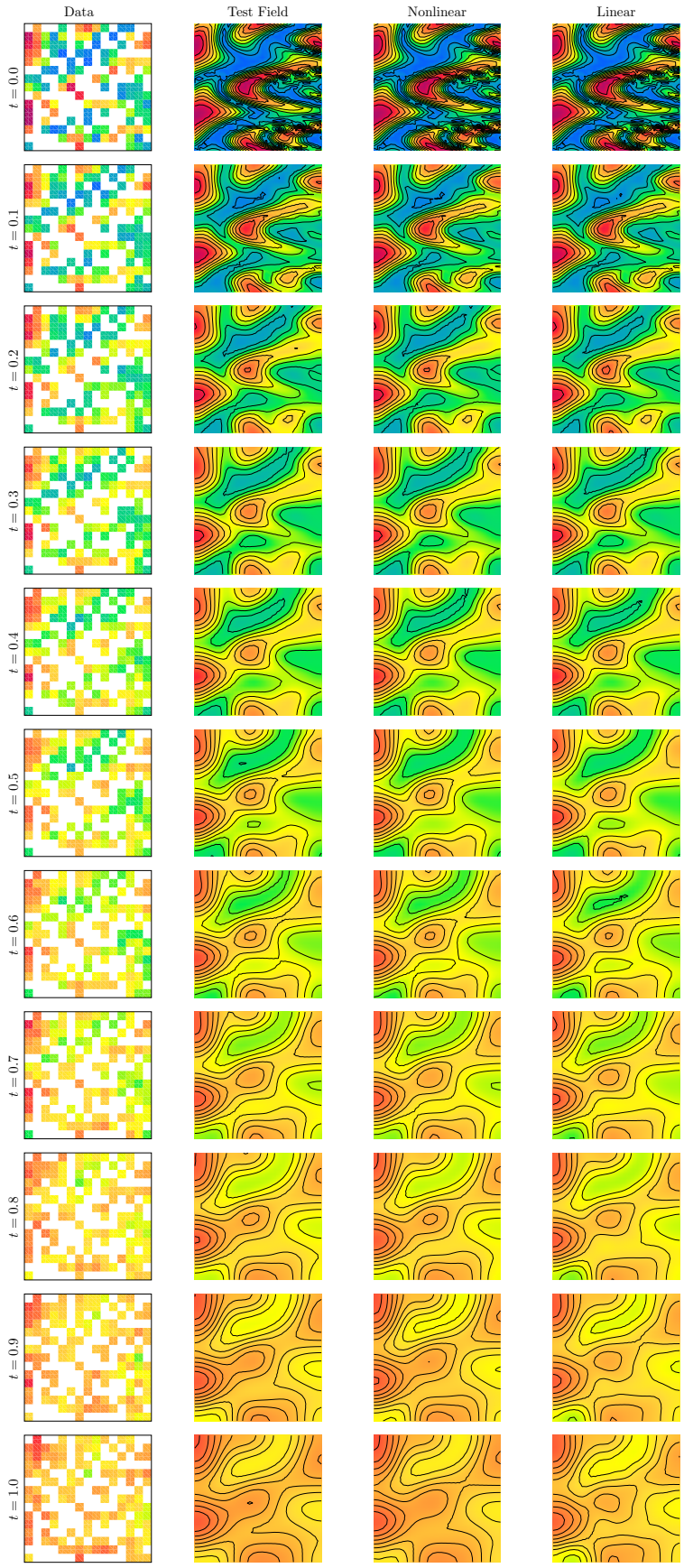


Figure B.3: Simulation 4: Unevenly spaced areal observations in space and pointwise observations in time, nonlinear dynamics. From left to right: synthetic data, test field, and estimates obtained with the proposed space-time regression with nonlinear PDE regularization (Nonlinear) and its linear counterpart (Linear).

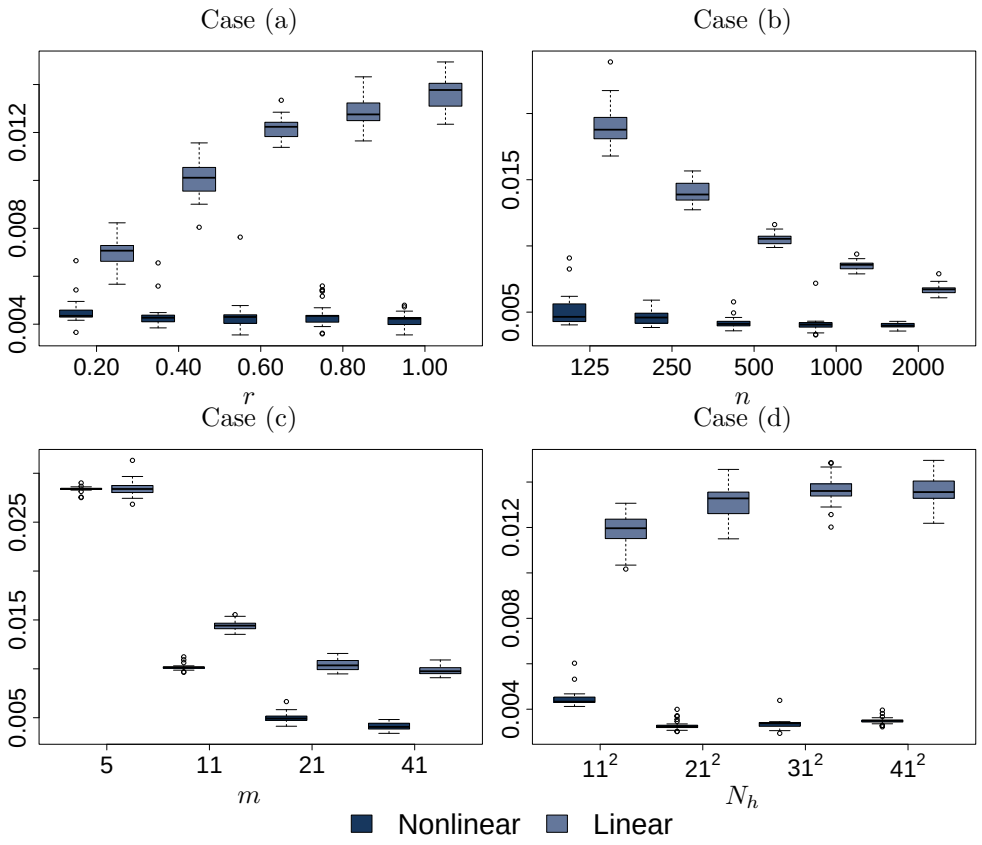


Figure B.4: RMSE of the proposed space-time regression method with nonlinear PDE regularization (Nonlinear) and its linear counterpart (Linear) over 30 simulation replicates, under different scenarios: (a) variation in the reaction parameter r; (b) variation in the number of spatial locations n; (c) variation in the number of time instants m; (d) variation in the number of mesh nodes N_h. In each panel, all other parameters are fixed as in Simulation 1 of the main article.

MOX Technical Reports, last issues

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

80/2026 Botteghi, N; Glas, S; Brune, C
Deep Invertible Autoencoders for Dimensionality Reduction of Dynamical Systems

78/2026 Botteghi, N; Brook, O; Fasel, U; Califano, F
Interconnection and Damping Assignment Passivity-Based Control using Sparse Neural ODEs

79/2026 Vitullo, P.; Macaluso, C.; Gazzoni, S.; Cassanmagnago, A.; Dimola, N.; Possenti, L.; Zunino, P.
A computational model for phenotype-structured tumour dynamics with reduced microvascular oxygen and event-based radiotherapy

77/2026 Renzi, F.; Pagani, S.; Antonietti, P.F.; Vergara, C.
A Neural-Network Approach for the Stability of Loosely-Coupled FSI Simulations in Hemodynamics

75/2026 Botteghi, N; Tomasetto, M; Fasel, U; Braghin, F; Manzoni, A
HypeMARL: Multi-Agent Reinforcement Learning For High-Dimensional, Parametric, and Distributed Systems

76/2026 Pizzamiglio, S.M.; Pagani, S.; Regazzoni, F.
Learning PDE solution operators with variable initial conditions via Latent Dynamics Networks

74/2026 Brivio, S.; Franco, N. R.
Deep Symmetric Autoencoders from the Eckart-Young-Schmidt Perspective

73/2026 Corti, M.
A weighted polygonal discontinuous Galerkin method for hierarchically coupled reaction–diffusion systems with cubic interactions

72/2026 Lorenzon, G.; Regazzoni, F.
Optimal Transport Dropout for Structured Predictive Uncertainty

70/2026 Micheletti, S.
Longitudinal versus quasi-P waves in weakly anisotropic elasticity: operator comparison and long-time dephasing