To appear in Statistical Modelling (2022), DOI:10.1177/1471082X211057959
Publisher's version available at https://doi.org/10.1177/1471082X211057959

Smoothing spatio-temporal data with complex missing data patterns

Eleonora Arnone [1], **Laura M. Sangalli** [1], **and Andrea Vicini** [1]

[1] MOX - Dipartimento di Matematica, Politecnico di Milano, Italy

Address for correspondence: Laura M. Sangalli, MOX - Dipartimento di Matematica, Politecnico di Milano, Piazza Leonardo da Vinci 32, 20133 Milano, Italy.
E-mail: laura.sangalli@polimi.it.
Phone: (+39) 02 2399 4554.
Fax: (+39) 02 2399 4568.

Abstract: We consider spatio-temporal data and functional data with spatial dependence, characterized by complicated missing data patterns. We propose a new method capable to efficiently handle these data structures, including the case where data are missing over large portions of the spatio-temporal domain. The method is based on regression with partial differential equation regularization. The proposed model can accurately deal with data scattered over domains with irregular shapes and can accurately estimate fields exhibiting complicated local features. We demonstrate the consistency and asymptotic normality of the estimators. Moreover, we illustrate the good performances of the method in simulations studies, considering different missing data scenarios, from sparse data to more challenging scenarios where the data are missing over large portions of the spatial and temporal domains and the missing data are clustered in space and/or in time. The proposed method is compared to competing techniques, considering predictive accuracy and uncertainty quantification measures. Finally, we show an application to the analysis of lake surface water temperature data, that further illustrates the ability of the method to handle data featuring complicated patterns of missingness and highlights its potentiality for environmental studies.

Key words: functional data with spatial dependence; incomplete and partially observed functional data; nonparametric regression with partial differential equation regularization; smoothing with roughness penalties

1 Introduction

In this work we consider space-time data and spatially dependent functional data affected by complicated missing data patterns. These data have recently attracted an increasing interest, and various approaches have been proposed to handle them, some born within the classical framework of spatio-temporal data analysis (Cressie and Wikle, 2015), while others

more influenced by the literature on functional data analysis (Ramsay and Silverman, 2005; Ferraty and Vieu, 2006; Horváth and Kokoszka, 2012; Kokoszka and Reimherr, 2017) and in particular on functional data with spatial dependence (see, e.g., the reviews and collections in Delicado et al., 2010; Mateu and Romano, 2017; Menafoglio and Secchi, 2017).

Sparsely observed space-time data and incomplete or partially observed functional data with spatial dependence are in fact frequently encountered in applications, especially in geo- and environmental sciences. Gromenko et al. (2017) propose a functional regression approach that can handle spatially indexed incomplete curves with large gaps in temporal coverage, appropriately weighting data at neighbouring locations. Partially observed functional data with spatial dependence are also addressed by Liu et al. (2017), who develop functional principal component analysis for spatially correlated data. Gong et al. (2018) and Gong et al. (2021) investigate two hierarchical dynamic spatio-temporal models constructed under the state-space framework (Cressie and Wikle, 2015). The authors introduce an approach similar to the Kalman filter for sparse data. Another possible approach to fill in unobserved values in this kind of data is the so-called Data INterpolating Empirical Orthogonal Functions (DINEOF). This method, proposed in Alvera-Azcárate et al. (2005) and Beckers and Rixen (2003), is also studied, for example, by Ping et al. (2016), who propose a version of DINEOF that improves the method from the point of view of the reconstruction accuracy and of the computational time.

In this work we propose a spatio-temporal regression model with differential regularization, capable to efficiently handle sparse space-time data and partially observed functional data with spatial dependence. The regularizing term in the model involves a Partial Differential Equation (PDE), enabling the inclusion of the available problem-specific information. We recall that the exploration of possible spatio-temporal methods based on PDEs or stochastic PDEs has attracted a strong interest since the pioneering work of Whittle (1954). Cameletti et al. (2013) propose for instance a sPDE approach based on Lindgren et al. (2011) and extensions (see, e.g., Fuglstad et al., 2015; Bakka et al., 2019). Cressie and Wikle (2015), Wikle and Hooten (2010) and Kuhnert (2014) present overviews on how to incorporate physical information in dynamical spatio-temporal models. The framework is the one of hierarchical modeling. From the seminal work of Berliner (2003), who coined the term physical statistical models, many works have contributed to this research area. Wikle (2003) presents a Bayesian hierarchical model where the process model is based on a PDE with space-varying coefficients, estimated with MCMC, while Stroud et al. (2010) propose an ensemble Kalman filter algorithm where the parameters of the PDE can also be time-dependent. Malmberg et al. (2008) propose a three dimensional spatial model plus temporal dynamics. The high computational cost is a major drawback for these methods, and a lot of attention has been dedicated to techniques to reduce the computational burden. Xu and Wikle (2007) explore EM and general EM algorithms for simple dynamical models; Sigrist et al. (2015) present a discretization based on Fourier expansions, able to reduce the computational cost when the shape of the spatial domain is simple and the PDE is stationary; Hefley et al. (2017) apply the homogenisation of PDEs to reduce the dimension of the problem; Richardson (2017) proposes a dynamical spatio-temporal model where the inverse evolution matrix is sparse, in order to avoid the inversion of dense matrices.

Our proposal instead originates from a different literature, that is about nonparametric and semiparametric regression with roughness penalties. Some instances in this line of research include among the others: Ugarte et al. (2009) that consider a semiparametric model combining a small area random effect and a non-parametric trend based on P-splines (Eilers and Marx, 1996, 2021); Ugarte et al. (2010) and Aguilera-Morillo et al. (2017) that consider a space-time penalized regression model, based on tensor product of univariate P-splines; Marra et al. (2012) and Augustin et al. (2013) that extend the soap film smoothing, introduced by Wood et al. (2008), to space-time data. Our current proposal in particular constitutes an addition to the class of semiparametric and nonparametric regression models with PDE regularization, reviewed in Sangalli (2021); see also Sangalli et al. (2013) and Azzimonti et al. (2015). We discretize the considered estimation functional using B-splines over the temporal domain and finite elements over a triangulation of the spatial domain. Thanks to the unstructured basis used in space, the method is able to deal with data observed over spatial domains with irregular shapes; moreover, it can accurately capture complicated fields with localized features. The methods are implemented in the R package `fdaPDE` (Lila et al., 2020).

This work focuses in particular on the properties of this method when dealing with sparsely observed space-time data and incomplete or partially observed functional data with spatial dependence. In particular, we study the asymptotic properties of the considered estimator in the context of missing data, demonstrating its consistency and asymptotic normality. The obtained results cover also the models for fully observed spatio-temporal data in Bernardi et al. (2017), whose asymptotic properties have never been investigated. We moreover assess the performance of the proposed method under different patterns of observability, from sparse data to more challenging scenarios where the data are missing over large portions of the spatial and temporal domains and the missing data are clustered in space and/or in time. Simulations studies show the comparative advantages of the proposed method with respect to available alternatives, including spatio-temporal kriging, other space-time methods with roughness penalties based on thin-plate splines and soap film smoothing, DINEOF, and the space-time stochastic PDE approach in Cameletti et al. (2013). We also apply the proposed model to study monthly averages of surface water temperature of Lake Victoria. These data, shown in Figures 1 and 2, are affected by complex missingness patterns, with several months where data are missing over a large surface of the lake, and even an interval of some consecutive months were data are completely missing. The same data are also considered by Gong et al. (2018), where the analyses are though restricted to a small rectangular portion of the lake. Here instead we consider the data observed over the whole lake, appropriately complying with its irregular shorelines.

The paper is organized as follows. In Section 2 we present the proposed methodology. In Section 3 we detail the implementation of the proposed method. In Section 4 we study the asymptotic properties. Simulation studies are in Section 5 and the application to surface temperature data in Section 6. Finally, in Section 7, we draw some conclusions and discuss possible future developments.

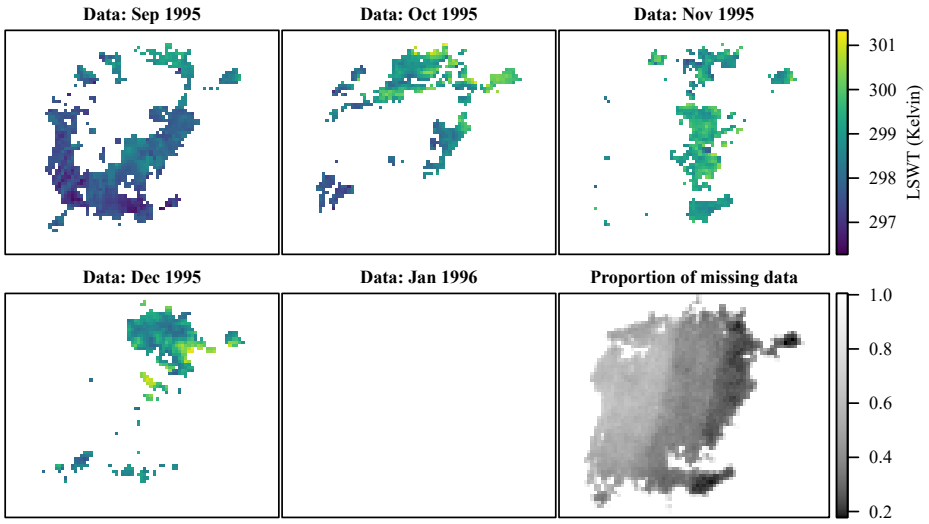

Figure 1: Spatial frames of Lake Victoria surface water temperatures (monthly averages), for five months. For most months, data are missing over large portions of the lake; for some months, no data is available. The bottom right panel displays the proportion of missing data over the considered time period; fewer observations are available on the west part of the lake with respect to the east part.

2 Mathematical framework

In this section we present Spatio-Temporal regression with Partial Differential Equation regularization (ST-PDE) for partially observed data.

2.1 Missing data setting

Let $\{\mathbf{p}_i = (x_i, y_i)\,; i = 1, \ldots, n\}$ be a set of n points on a spatial domain $\Omega \subset \mathbb{R}^2$ with boundary $\partial\Omega \in \mathcal{C}^2$, and let $\{t_j; j = 1, \ldots, m\}$ be a set of m instants in a time interval $T \subset \mathbb{R}$. Denote with z_{ij} the value of a real-valued variable observed at point $\mathbf{p}_i$ and time t_j. We are particularly interested in the case where the values of z_{ij} are missing over possibly large portions of the spatial and temporal domain. For example, for any fixed spatial location $\mathbf{p}_i$, the temporal profile may be observed in a fragmented way and display long gaps in time; analogously, for any fixed temporal instant t_j, the spatial field may be unobserved over large regions of Ω. This is illustrated in Figures 1 and 2 that display the surface temperature data analyzed in Section 6. The two figures highlight the complicated missingness pattern affecting the data, with missing data clustered in space and time. For notation convenience, let O be the set of indices of observed data:

$$O = \{(i, j) : z_{ij} \text{ is observed}, i = 1, \ldots n, j = 1, \ldots m\}.$$

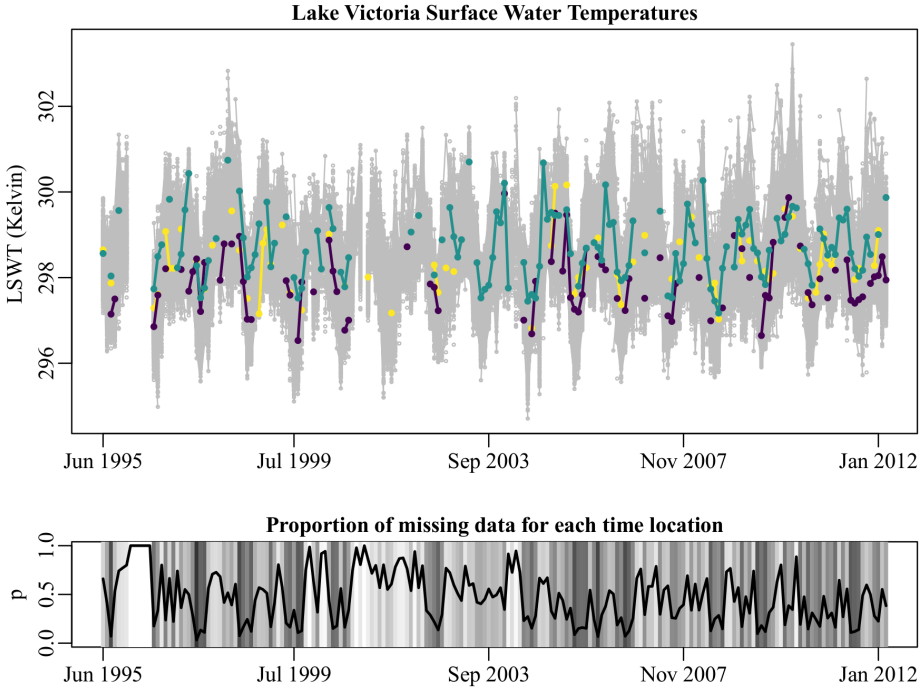

Figure 2: Temporal profiles of Lake Victoria surface water temperatures (monthly averages) observed over a fine uniform grid covering the lake (the spatial grid of observation is shown in the bottom right panel of Figure 1). Values observed at consecutive time instants are linked by a line. Highlighted in color are the temporal profiles measured at three random locations over the lake. The bottom panels shows the proportion of missing data along time, pointing out the high proportion of missing data at various months, including an interval of some consecutive months where no datum is recorded.

2.2 ST-PDE for partially observed data

We model $\{z_{ij} : (i,j) \in O\}$ as noisy observations of an underlying spatio-temporal smooth function $f(\mathbf{p}, t)$ on $\Omega \times T$:

$$z_{ij} = f(\mathbf{p}_i, t_j) + \varepsilon_{ij} \qquad (i,j) \in O \tag{2.1}$$

where $\{\varepsilon_{ij} : (i,j) \in O\}$ are independently distributed residuals with zero mean and constant variance σ^2.

To estimate f we minimize the following regularized sum of square error functional:

$$\begin{aligned} J(f) = & \sum_{(i,j)\in O} (z_{ij} - f(\mathbf{p}_i, t_j))^2 + \\ & + \lambda_S \int_T \int_\Omega (Lf - u)^2 d\mathbf{p}dt + \lambda_T \int_\Omega \int_T \left(\frac{\partial^2 f}{\partial t^2}\right)^2 dtd\mathbf{p} \end{aligned} \tag{2.2}$$

where $\lambda_S > 0$ and $\lambda_T > 0$ are two positive smoothing parameters and $Lf = u$ is a known PDE that formalises the available problem-specific information about the unknown spatial field. The use of a PDE in the regularizing term allows for a rich modeling of the space variation. The differential operator L can for instance be a diffusion-transport-reaction operator of the form:

$$Lf = -\operatorname{div}(K\nabla f) + \mathbf{b} \cdot \nabla f + cf$$

where $K \in \mathbb{R}^2$ is the symmetric and positive definite diffusion tensor, $\mathbf{b} \in \mathbb{R}^2$ the transport vector and $c \in \mathbb{R}^+$ the reaction term. The second order term $-\operatorname{div}(K\nabla f)$ induces a smoothing in all the directions, that can be a isotropic smoothing (when K is a multiple of the identity) or a anisotropic smoothing along a preferential direction and with a certain intensity of anisotropy. Unidirectional effects in space can be modeled with a transport term $\mathbf{b} \cdot \nabla f$. The reaction term cf instead has a shrinkage effect because of the penalization of the $L^2(\Omega)$ norm of f. Moreover, the parameters K, $\mathbf{b}$ and c, as well as the forcing term u, can be space-varying on Ω, further increasing the flexibility in the modeling of space variation, to model various forms of non-stationarity. For simplicity of the presentation, to avoid more involved formulae, we consider here the case in which the forcing term u in the PDE is null. Non-null forcing terms can be handled analogously to what done by Azzimonti et al. (2015) in a model involving only spatial data. The regularization in the temporal dimension, instead of the second derivative in time, could also involve a more general differential operator, encoding the available information about the temporal evolution of the phenomenon. For simplicity, we do not prosecute this direction in this work.

In order to appropriately account for the shape of the domain, we impose conditions on f at the boundary $\partial\Omega$ of the spatial domain. These conditions, that depend on the specific problem being considered, can for instance involve the values of f on $\partial\Omega$ (Dirichlet conditions), the value of the normal derivative of f on $\partial\Omega$ (Neumann conditions), and combinations of the previous conditions, with different conditions on different portions of $\partial\Omega$ (mixed conditions); see, e.g., Arnone et al. (2019). For sake of simplicity, in this work we consider homogeneous Neumann conditions, i.e., null normal derivative of f on $\partial\Omega$. These are natural conditions for the considered estimation problems (and are the default option in the function `smooth.FEM.time` of the R package `fdaPDE`), and should be preferred unless problem-specific information suggests the use of Dirichlet or mixed conditions.

In the special case where the data are observed without missingness, the differential operator L coincides with the Laplace operator ($L = \Delta$, isotropic and stationary case), the forcing term is null ($u = 0$), and homogeneous Neumann boundary conditions are imposed, the method reduces to the one considered in Bernardi et al. (2017).

Alternatively to the use of two separate penalty terms for space and time regularization, it is also possible to consider a unique regularizing term, jointly modeling the regularity of the field in space and time. This choice is particularly convenient when a strong information on the spatio-temporal dynamic of the phenomenon is available and can be formalized in terms of a time-dependent PDE $\frac{\partial f}{\partial t} + Lf = u$. In this case, the functional (2.2) is replaced

by

$$\sum_{(i,j)\in O} \left(z_{ij} - f\left(\mathbf{p}_i, t_j\right)\right)^2 + \lambda \int_T \int_\Omega \left(\frac{\partial f}{\partial t} + Lf - u\right)^2 d\mathbf{p}dt. \tag{2.3}$$

Arnone et al. (2019) studies analogous models under various fully observed samplings designs and demonstrated the well-posedness of the associated estimation problems (i.e., the existence of one and only one solution to the estimation problem).

It should noticed that, in the estimation problems (2.2) and (2.3), we do not assume that the f satisfies the PDE in the regularizing term: the PDE carries only partial information about the field. Because of this, we use the PDE to regularize the estimate, rather than constraining $\hat{f}$ to be a solution of the PDE. The smoothing parameters λ_S and λ_T in (2.2) and λ in (2.3) balance the weight given to the data, in the least-square term, and to the differential model, in the regularizing terms. The values of these parameters can be automatically selected, e.g., by Generalized Cross Validation (GCV); see Section 3.3. This data-driven criterion adjusts the value of the smoothing parameters selecting relatively large values when the considered PDE provides a good description of the phenomenon at hand, and relatively low value when instead the data are far from solutions to the PDE.

In our experience, unless problem-specific information is available on the spatio-temporal behavior of the phenomenon (Arnone et al., 2019), the estimation problem (2.2) is more flexible, enabling to separately model the regularity of the phenomenon in space and in time. There might indeed be situations where we lack information of the overall space-time behavior, but we have some knowledge on either the temporal evolution or on the spatial structure of the phenomenon, which can efficiently be separately included in the two regularizing terms in (2.2). Moreover, (2.2) permits to more easily select the relative weights of the regularity in space and in time.

It should also be noted that the proposed methodology can on the other hand be also applied in absence of problem-specific information or any appreciable anisotropy in the data. In such case, it is natural to assume $L = \Delta$, thus implying isotropic and stationary regularizations. This is the setting considered for the simulations and application presented in this work and is also the default option in the function `smooth.FEM.time` of the R package `fdaPDE` (Lila et al., 2020), that implements the methods corresponding to the general estimation functionals (2.2) and (2.3).

3 From an infinite dimensional to a finite dimensional problem

The minimization of functional (2.2) does not have a closed form solution. In this section we describe the discretization of this problem. Specifically, we look for the estimate $\hat{f}$ in the space of functions that can be expressed as an expansion on an appropriate space-time basis system.

3.1 Basis systems

Let $\{\varphi_k(t) : k = 1, \ldots, M\}$ be a set of M basis functions defined on T. In this work, in particular, we use cubic B-spline bases. Moreover, let $\{\psi_l(\mathbf{p}) : l = 1, \ldots, N\}$ be a set of N basis functions defined on a triangulation $\Omega_{\mathcal{T}}$ of Ω. As an example, Figure 3 shows the domain Ω for the application in Section 6: the left panel of the figure displays the detailed coastline of the Lake Victoria and the right panel gives the triangulation $\Omega_{\mathcal{T}}$ used for the finite dimensional approximation. In this work, we use a linear finite element basis on $\Omega_{\mathcal{T}}$: we have a basis function ψ_l for each vertex of the triangulation, and ψ_l is a continuous piecewise linear function on $\Omega_{\mathcal{T}}$, which takes value one at the corresponding vertex and zero at all other vertices of the triangulation. This choice leads to an efficient discretization of the functional and allows to deal with complex geometries of the spatial domain, which may influence the phenomenon behaviour. Other choices for the basis systems are also possible, for example Wilhelm et al. (2016) explores the use of Non-Uniform rational B-splines for the discretization of the spatial penalty.

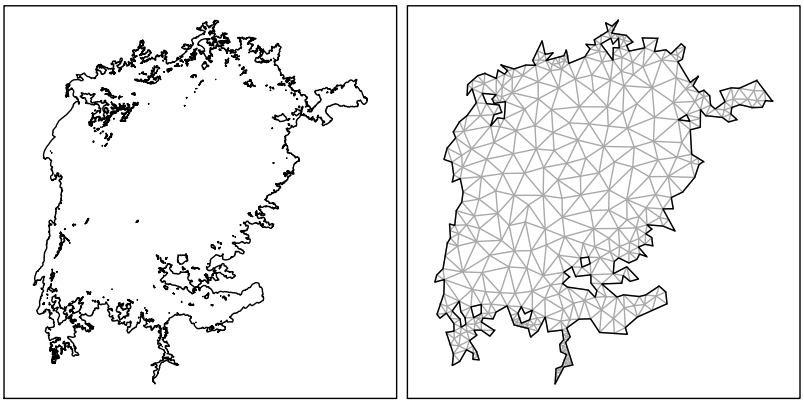

Figure 3: Left: detailed coastlines of Lake Victoria. Right: simplified lake domain and triangulation.

The spatio-temporal field $f(\mathbf{p}, t)$ over $\Omega_{\mathcal{T}} \times T$ is then expressed as:

$$f(\mathbf{p}, t) = \sum_{l=1}^{N} \sum_{k=1}^{M} f_{lk} \psi_l(\mathbf{p}) \varphi_k(t) \tag{3.1}$$

where $\{f_{lk}; l = 1, \ldots, N; k = 1, \ldots, M\}$ are the coefficients of the expansion on the separable spatio-temporal basis.

3.2 Discretization of key elements

Let $\mathbf{z}$ be the vector of length $\nu = |O|$ obtained appropriately stacking the observed values z_{ij}. Here we use a time-by-space tensorization, meaning that we insert in $\mathbf{z}$ all the available observations at the first time instant, then all those at the second time instant, and so on. We define $\mathbf{f}_O$ to be the vector of evaluations of f in the locations corresponding to

the observed data z_{ij}. For example, in case of completely observed data, $\mathbf{z}$ and $\mathbf{f}_O$ assume respectively the form

$$\begin{aligned}\mathbf{z} &= [z_{11}, z_{21}, \dots, z_{n1}, \dots, z_{1m}, z_{2m}, \dots, z_{mn}]^\top;\\ \mathbf{f}_O &= [f(\mathbf{p}_1, t_1), f(\mathbf{p}_2, t_1), \dots, f(\mathbf{p}_n, t_1), \dots, (\mathbf{p}_1, t_m), f(\mathbf{p}_2, t_m), \dots, f(\mathbf{p}_n, t_m)]^\top.\end{aligned}$$

Let $\mathbf{f}$ be the vector of length NM of coefficients of the basis expansion (3.1), i.e.,

$$\mathbf{f} = [f_{11}, f_{21}, \dots, f_{N1}, \dots, f_{1M}, f_{2M}, \dots, f_{MN}]^\top.$$

We set $\boldsymbol{\psi} = [\psi_1, \dots, \psi_N]^\top$, $\boldsymbol{\psi}_x = [\partial\psi_1/\partial x, \dots, \partial\psi_N/\partial x]^\top$, $\boldsymbol{\psi}_y = [\partial\psi_1/\partial y, \dots, \partial\psi_N/\partial y]^\top$ and $\nabla\boldsymbol{\psi} = \left[\boldsymbol{\psi}_x, \boldsymbol{\psi}_y\right]$. Analogously we define $\boldsymbol{\varphi} = [\varphi_1, \dots, \varphi_M]^\top$ and $\boldsymbol{\varphi}_{tt} = \left[\partial^2\varphi_1/\partial t^2, \dots, \partial^2\varphi_M/\partial t^2\right]^\top$.

We define the $n \times N$ matrix Ψ of evaluations of the N finite elements basis at the n space locations $\{\mathbf{p}_i; i = 1, \dots, n\}$, and the $m \times M$ matrix Φ of the evaluations of the M temporal basis functions at the m time instants $\{t_j; j = 1, \dots, m\}$, i.e.,

$$\Psi = \begin{bmatrix} \psi_1(\mathbf{p}_1) & \psi_2(\mathbf{p}_1) & \dots & \psi_N(\mathbf{p}_1) \\ \psi_1(\mathbf{p}_2) & \psi_2(\mathbf{p}_2) & \dots & \psi_N(\mathbf{p}_2) \\ \vdots & \vdots & \dots & \vdots \\ \psi_1(\mathbf{p}_n) & \psi_2(\mathbf{p}_n) & \dots & \psi_N(\mathbf{p}_n) \end{bmatrix}, \quad \Phi = \begin{bmatrix} \varphi_1(t_1) & \varphi_2(t_1) & \dots & \varphi_M(t_1) \\ \varphi_1(t_2) & \varphi_2(t_2) & \dots & \varphi_M(t_2) \\ \vdots & \vdots & \dots & \vdots \\ \varphi_1(t_m) & \varphi_2(t_m) & \dots & \varphi_M(t_m) \end{bmatrix}.$$

Then we define the matrix B obtained from the Kronecker product $\Phi \otimes \Psi$ by removing the rows corresponding to the missing observations. More precisely, if the datum at the spatial location $\mathbf{p}_i$ and time location t_j is not available, we remove the $(i + nj)$-th row from $\Phi \otimes \Psi$. In case of completely observed data $B = \Phi \otimes \Psi$. With these definitions we can write $\mathbf{f}_O = B\mathbf{f}$ and discretize the first part of the functional as $(\mathbf{z} - B\mathbf{f})^\top(\mathbf{z} - B\mathbf{f})$.

To write the discretization of the regularizing term, we introduce the matrices P_T and P_S. P_T represents the discretization of the temporal penalty term in (2.2) and is given by $P_T = \int_T \boldsymbol{\varphi}_{tt}\boldsymbol{\varphi}_{tt}^\top$, while P_S represents the discretization of the spatial penalty term in (2.2) and is given by $P_S = R_1^\top R_0^{-1} R_1$, where $R_1 = \int_{\Omega_\tau} \left(\nabla\boldsymbol{\psi}\, K\, \nabla\boldsymbol{\psi}^\top + \nabla\boldsymbol{\psi}\, \mathbf{b}\, \boldsymbol{\psi}^\top + c\boldsymbol{\psi}\boldsymbol{\psi}^\top\right)$ and $R_0 = \int_{\Omega_\tau} \boldsymbol{\psi}\boldsymbol{\psi}^\top$. The discretization of the spatial penalty is based on a variational characterization of the estimation problem and implicitly assumes homogeneous Neumann boundary conditions; see Sangalli et al. (2013) for details. Finally, let $R_t = \int_T \boldsymbol{\varphi}\boldsymbol{\varphi}^\top$ be the mass matrix in time.

3.3 Discrete estimation problem

We can thus write the discretized sum of square error functional (2.2) as

$$\begin{aligned} J =& (\mathbf{z} - B\mathbf{f})^\top(\mathbf{z} - B\mathbf{f}) + \lambda_S \mathbf{f}^\top (R_t \otimes P_S)\, \mathbf{f} + \lambda_T \mathbf{f}^\top (P_T \otimes R_0)\, \mathbf{f} \\ =& (\mathbf{z} - B\mathbf{f})^\top(\mathbf{z} - B\mathbf{f}) + \mathbf{f}^\top P \mathbf{f} \end{aligned} \tag{3.2}$$

where we indicate with P the overall penalty: $P = \lambda_S (P_S \otimes R_t) + \lambda_T (R_0 \otimes P_T)$. The vector of coefficients $\hat{\mathbf{f}}$ that minimizes the functional (3.2) is obtained deriving J with respect to $\mathbf{f}$ and setting the derivative equal to zero, leading to

$$\hat{\mathbf{f}} = \left(B^\top B + P\right)^{-1} B^\top \mathbf{z}.$$

Set $\hat{\mathbf{z}} = B\left(B^\top B + P\right)^{-1} B^\top \mathbf{z}$, the vector of predicted values at the observed data locations $(\mathbf{p}_i, t_j)$ with $(i,j) \in O$. Moreover, set $S = B\left(B^\top B + P\right)^{-1} B^\top$, the smoothing matrix. The error variance σ^2 can then be estimated by

$$\hat{\sigma}^2 = \frac{1}{\nu - \mathrm{tr}(S)} (\mathbf{z} - \hat{\mathbf{z}})^\top (\mathbf{z} - \hat{\mathbf{z}}). \tag{3.3}$$

The smoothing parameters λ_S and λ_T balance the trade-off between the fidelity to the data and the regularity of the solution. Various possibility can be employed to automatically select these parameters. Here we use Generalized Cross Validation (GCV), selecting λ_S and λ_T to minimize

$$GCV(\lambda_S, \lambda_T) = \frac{1}{\nu(1 - tr(S)/\nu)^2} (\hat{\mathbf{z}} - \mathbf{z})^T (\hat{\mathbf{z}} - \mathbf{z}).$$

3.4 Estimating the field in a new location

In the context of missing data, it is often of interest to reconstruct the signal at the missing data locations and, in general, at any unobserved spatio-temporal location $(\mathbf{p}, t) \in \Omega \times T$.

Define the vector $\mathbf{B}(\mathbf{p}, t) = \psi(\mathbf{p})^\top \otimes \varphi(t)^\top$ of evaluations of the separable basis system at the generic spatio-temporal location $(\mathbf{p}, t) \in \Omega \times T$. We thus have

$$\hat{f}(\mathbf{p}, t) = \mathbf{B}(\mathbf{p}, t) \left(B^\top B + P\right)^{-1} B^\top \mathbf{z} \tag{3.4}$$

for each $(\mathbf{p}, t) \in \Omega \times T$. The mean of the estimator is thus given by

$$E[\hat{f}(\mathbf{p}, t)] = \mathbf{B}(\mathbf{p}, t) \left(B^\top B + P\right)^{-1} B^\top \mathbf{f}_O$$

while the covariance operator in any two locations $(\mathbf{p}_1, t_1), (\mathbf{p}_2, t_2) \in \Omega \times T$ is given by

$$\begin{aligned} &\mathrm{Cov}\left[\hat{f}(\mathbf{p}_1, t_1), \hat{f}(\mathbf{p}_2, t_2)\right] = \\ &\quad \sigma^2 \mathbf{B}(\mathbf{p}_1, t_1) \left(B^\top B + P\right)^{-1} B^\top B \left(B^\top B + P\right)^{-1} \mathbf{B}(\mathbf{p}_2, t_2)^\top. \end{aligned}$$

3.5 Model with covariates

Space-time varying covariates can be introduced in the model using a semi-parametric approach. Let $\mathbf{w}_{ij}$ be a vector of q covariates associated with the observation z_{ij}, at location

$\mathbf{p}_i$ and time instant t_j. Let $\boldsymbol{\beta} \in \mathbb{R}^q$ be an unknown vector of regression coefficients that indicates the effect of the covariates on the mean of the variable of interest. We consider the semi-parametric model

$$z_{ij} = \mathbf{w}_{ij}^T \boldsymbol{\beta} + f(\mathbf{p}_i, t_j) + \varepsilon_{ij}, \quad (i,j) \in O. \tag{3.5}$$

We then estimate $\boldsymbol{\beta}$ and f minimising a functional as (2.2), where the sum of squared error term is replaced by $\sum_{(i,j)\in O} \left(z_{ij} - \mathbf{w}_{ij}^T \boldsymbol{\beta} - f(\mathbf{p}_i, t_j)\right)^2$. This term can be written in matrix form as $(\mathbf{z} - W\boldsymbol{\beta} - B\mathbf{f})^T(\mathbf{z} - W\boldsymbol{\beta} - B\mathbf{f})$, where W is the $\nu \times q$ matrix whose rows are the vectors $\mathbf{w}_{ij}$. Let $H = W\left(W^TW\right)^{-1}W^T$ and $Q = I_\nu - H$, where I_ν is the identity matrix of dimension ν. Deriving the functional with respect to $\boldsymbol{\beta}$ and $\mathbf{f}$, and setting the derivatives to zero, we then obtain the estimates

$$\begin{aligned}
\hat{\mathbf{f}} &= \left(B^TQB + P\right)^{-1} B^TQ\mathbf{z}, \\
\hat{\boldsymbol{\beta}} &= \left(W^TW\right)^{-1} W^T(\mathbf{z} - B\hat{\mathbf{f}}).
\end{aligned}$$

The estimate of the field f at any spatio-temporal location, and its distributional properties, follow as for the model without covariates, replacing the matrix $\left(B^TB + P\right)^{-1}B^T$ by $\left(B^TQB + P\right)^{-1}B^TQ$. Regarding $\hat{\boldsymbol{\beta}}$ we have

$$\begin{aligned}
E[\hat{\boldsymbol{\beta}}] &= \boldsymbol{\beta} + \left(W^TW\right)^{-1} W^T\left(I_{nm} - S_Q\right)\mathbf{f}_\nu, \\
\mathrm{Var}[\hat{\boldsymbol{\beta}}] &= \sigma^2\left(W^TW\right)^{-1} + \sigma^2\left(W^TW\right)^{-1} W^TS_QS_Q^TW\left(W^TW\right)^{-1},
\end{aligned}$$

where $S_Q = B\left(B^TQB + P\right)^{-1}B^TQ$. The value of a new observation at location point $\mathbf{p} \in \Omega$ and time instant $t \in T$ and with associated covariates $\mathbf{w}$ can be predicted by $\mathbf{w}^T\hat{\boldsymbol{\beta}} + \hat{f}(\mathbf{p}, t)$.

In this work we do not discuss inference on $\hat{\boldsymbol{\beta}}$ and $\hat{f}$, which can for instance be performed following classical undersmoothing approaches (see, e.g., the literature review in Hall and Horowitz, 2013) and the approaches in Gray (1994), Yu and Ruppert (2002) and Holland (2017). The next section establishes the asymptotic properties of the estimators.

4 Asymptotic properties

We now study the infill asymptotic properties of the discrete ST-PDE estimators. In this section, for simplicity of exposition, we reparametrize the estimation functional (2.2) as

$$\begin{aligned}
J(f, \boldsymbol{\beta}) = &\frac{1}{\nu}\sum_{(i,j)\in O}\left(z_{ij} - \mathbf{w}_{ij}^T\boldsymbol{\beta} - f(\mathbf{p}_i, t_j)\right)^2 \\
&+ \lambda_{S,\nu}\int_T\int_\Omega (Lf - u)^2 d\mathbf{p}dt + \lambda_{T,\nu}\int_\Omega\int_T\left(\frac{\partial^2 f}{\partial t^2}\right)^2 dtd\mathbf{p}.
\end{aligned} \tag{4.1}$$

This parametrization is more convenient for the study of the asymptotic properties. Clearly, this problem is equivalent to the one so far considered, setting $\lambda_S = \nu\lambda_{S,\nu}$ and $\lambda_T = \nu\lambda_{T,\nu}$. The discretization of (4.1) leads to the following expression for the estimators

$$\hat{\mathbf{f}}_\nu = \left(B^T Q B/\nu + P\right)^{-1} B^T Q \mathbf{z}/\nu$$
$$\hat{\boldsymbol{\beta}}_\nu = \left(W^T W\right)^{-1} W^T (\mathbf{z} - B\hat{\mathbf{f}}_\nu),$$

where we highlight the dependence of the estimators on the sample size ν.

We consider a fixed $NM-$dimensional spatio-temporal basis, rich enough to accurately describe the spatio-temporal structure of the problem. We make the following assumption.

Assumption 1. For sufficiently large ν, the $NM \times NM$ matrix $B^\top Q B$ is nonsingular.

Before enunciating the main theorems, we discuss under which conditions Assumption 1 is satisfied. We recall that B is the $\nu \times NM$ collocation matrix that contains the evaluations of the tensor product basis functions $\psi_l \phi_k$ in the locations $(\mathbf{p}_i, t_j)$. The Schoenberg-Whitney Theorem (Schoenberg and Whitney, 1953) establishes the conditions ensuring that the B-splines collocation matrix has full rank. A similar result can be derived for the finite element basis. In particular, a sufficient and necessary condition for B to be full rank is to have at least $\nu \geq NM$ observations, with at least one observation in each support of the tensor product basis function $\psi_l \phi_k$, for $l = 1, \ldots, N$, $k = 1, \ldots, M$. When no covariates are available (i.e., Q is the identity matrix), if B is full rank, then also $B^\top Q B$ is full rank, and thus Assumption 1 is satisfied. When covariates are available, then Q has rank $\nu - q$. This means that, in the product $B^\top Q B$, the matrix Q has the effect of annihilating q rows of B. In general having q additional observations is enough to make Assumption 1 satisfied. In any case, a sufficient condition to have $B^\top Q B$ full rank, is to have at least $\nu \geq (q+1)NM$ observations, with at least $q+1$ observations in each support of $\psi_l \phi_k$, for $l = 1, \ldots, N$, $k = 1, \ldots, M$. Thus, in practical applications, it is sufficient to choose a discretization that ensures this property. Moreover, the non-singularity of the matrix can also be verified in the code.

Under Assumption 1, we can define the matrices $\Sigma_\nu = W^\top W/\nu$ and $A_\nu = \left(B^\top Q B/\nu\right)^{-1}$. The following theorems establish the asymptotic properties of $\hat{\mathbf{f}}_\nu$ and $\hat{\boldsymbol{\beta}}_\nu$.

Theorem 1. *Let $\{\hat{\mathbf{f}}_\nu\}$ be a sequence of ST-PDE estimators. Assume that a nonsingular limit $A = \lim_\nu A_\nu$ exists. If $\lambda_{S,\nu} \to 0$ and $\lambda_{T,\nu} \to 0$, then $\hat{\mathbf{f}}_\nu$ is a consistent estimator for $\mathbf{f}$. Moreover, if $\lambda_{S,\nu} = o\left(\nu^{-1/2}\right)$ and $\lambda_{T,\nu} = o\left(\nu^{-1/2}\right)$*

$$\sqrt{\nu}(\hat{\mathbf{f}} - \mathbf{f})|W \xrightarrow{d} \mathcal{N}_{NM}\left(0, \sigma^2 A\right),$$

where $\xrightarrow{d}$ denote the convergence in distribution.

Theorem 2. *Let $\{\hat{\boldsymbol{\beta}}_\nu\}$ be a sequence of ST-PDE estimators. Assume $\Sigma = \lim_\nu \Sigma_\nu$ exists and is nonsingular. Then, under the hypothesis of Theorem 1, the estimator $\hat{\boldsymbol{\beta}}_\nu$ is consistent for $\boldsymbol{\beta}$. Moreover, if $\lambda_{S,\nu} = o\left(\nu^{-1/2}\right)$ and $\lambda_{T,\nu} = o\left(\nu^{-1/2}\right)$*

$$\sqrt{\nu}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})|W \xrightarrow{d} \mathcal{N}_q\left(0, \sigma^2\left(\Sigma^{-1} + (1/\nu^2)\Sigma^{-1} W^\top B A B^\top W \Sigma^{-1}\right)\right).$$

The proofs of Theorems 1 and 2 are deferred to the supplementary material. These theorems provide the consistency and asymptotic normality also for the estimators in Bernardi et al. (2017).

5 Simulation studies

In this section we show the good performances of the proposed ST-PDE method in different missing data scenarios, described in Section 5.2, comparing it with various alternative methods, detailed in Section 5.3. The methods are tested considering two test fields, outlined in Section 5.1.

5.1 Test fields

We consider two different test fields:

- a simple spatio-temporal field on a non-convex C-shaped domain (some temporal snapshots in Figure 4; Analytical expression in the supplementary material);
- a complicated spatio-temporal field on a simple squared domain (some temporal snapshots in the first row of Figure 10; analytical expression in the supplementary material).

In both cases the time domain is the interval $[0, 1]$.

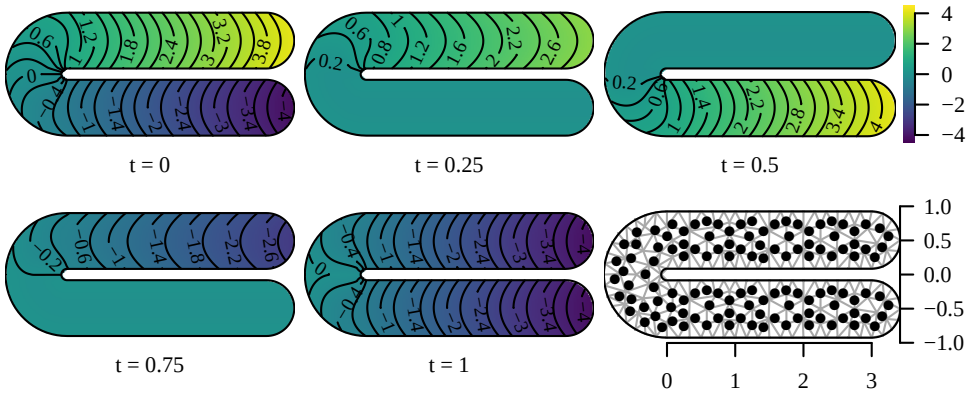

Figure 4: Test field on the C-shaped domain. For some time instants the field assumes very different values on the two arms of the C. The last panel shows the observational spatial grid corresponding to the full (uncensored) design, and the triangulation used for ST-PDE estimate.

The field on the C domain is the same considered in Bernardi et al. (2017), and based in turns on the one considered in Ramsay (2002) and Wood et al. (2008), and is constructed in such a way that it can assume quite distant values on the two arms of the C. The time and space components are present in a non separable way. This simple field puts in difficulty classical methods that are blind to the shape of the spatial domain, since this smooth field

assumes quite different values at locations that are close in the Euclidean distance, but separated by a boundary of the spatial domain.

The square is a simpler domain and there are many classical techniques working on it. On this domain, we consider a field that exhibits more complicated features and localized behaviours.

5.2 Data generation

We generate data from the test fields in Section 5.1, according to model (2.1), or model (3.5) when also considering covariates. The test fields are evaluated on a set of spatio-temporal locations. Specifically, in space, for the field on the C-shaped domain we consider the sampling grid of 120 points shown in the last panel of Figure 4, and for the field on the squared domain we consider a sampling grid of 225 points on a regular lattice. In time, we consider an equispaced grid of 41 sampling points over the interval $[0, 1]$. To the values of the test field evaluated at the sampling grid we add incorrelated gaussian errors with zero mean and standard deviation equal to 0.4 for to the field the C-shaped domain, and equal to 0.1 for the field on the C-shaped domain, corresponding in both cases to about 5% of test field range.

Finally, we censor the data according to one of the four following censoring schemas, that reproduce different missing data scenarios:

(a) independent censoring in time and space;
(b) dependent censoring in time and independent censoring in space;
(c) independent censoring in time and dependent censoring in space;
(d) dependent censoring in time and space.

We now illustrate these schemas in more detail.

5.2.1 Censoring (a): independent in time and space

The first censoring schema corresponds to a sparse spatio-temporal data setting. We randomly sample a proportion p of indexes $(i, j) \in \{1, \dots, n\} \times \{1, \dots, m\}$ and we set to `NA` the corresponding observations. In this way the spatio-temporal position of each missing value is independent of the others.

The left panels of Figure 5 (censoring independent in time) and the left panels of Figure 6 (censoring independent in space) show the plot of the noisy data obtained under censoring (a), with $p = 0.5$, for the test function on the C-shaped domain. The top left panel in Figure 5 (censoring independent in time) displays the temporal profiles corresponding to each spatial location, where values observed at two consecutive time instants are linked by a line; three temporal profiles, corresponding to three random locations, are highlighted in colors. The left panels in Figure 6 (censoring independent in space) displays the spatial observations sampled at three different time instants. These plots show that the observational process is highly fragmented in both space and time, with observations sparsely covering the whole temporal and spatial domains.

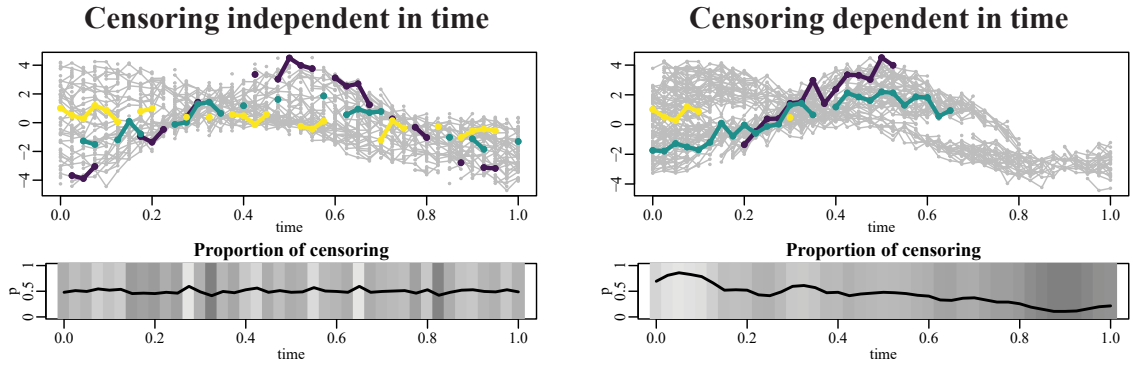

Figure 5: Censoring independent in time (left) *vs* censoring dependent in time (right). Plots obtained for test on C-shaped domain, $p = 50\%$. Top panels display the temporal profiles corresponding to each spatial location, where values observed at two consecutive time instants are linked by a line; three temporal profiles, corresponding to three random locations, are highlighted in colors. Bottom panels display the overall proportion of censoring at each sampling time instant, for this simulation replicate. Under schemas (a) and (c), where data undergo a censoring independent in time (left), the temporal profiles are highly fragmented, with sparse observations covering the whole temporal domain. Under schemas (b) and (d), where data undergo a censoring dependent in time (right), each temporal profile is typically observed for a longer fragment, while it is totally unexplored over large portions of the temporal domain.

5.2.2 Censoring (b): dependent in time and independent in space

The second censoring schema corresponds to the setting considered in partially observed functional data with spatial dependence (see, e.g., Gromenko et al., 2017). In this setting, we can look at data as evaluations of spatially dependent curves, where each curve is observed only on certain intervals of the time domain, and the observed region may be independent of that of neighbouring spatial locations.

To model this situation, independently for each spatial location $\mathbf{p}_i$, we generate a random decomposition of the temporal domain in K sub-intervals of T; for a proportion p of censoring, we then set all the observations corresponding to pK randomly selected intervals to `NA`.

The right panels of Figure 5 (censoring dependent in time) shows the typical temporal profiles observed under this censoring schema: the temporal profiles are typically observed for longer fragments, with respect to what is observed when the censoring is independent in time, while they are instead totally unobserved over large portions of the temporal domain. In space, instead, data are sparsely observed over the whole spatial domain, as in the left panels of Figure 6 (censoring independent in space), similarly to the censoring schema (a).

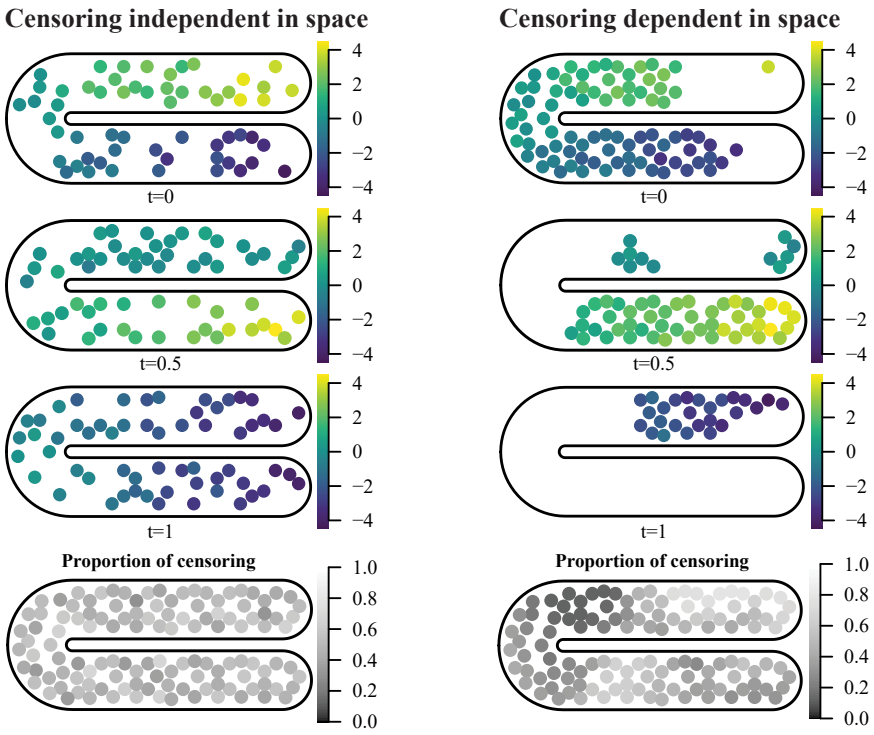

Figure 6: Censoring independent in space (left) *vs* censoring dependent in space (right). Plots obtained for test on C-shaped domain, $p = 50\%$. The plots shows the observed values at three different time instants. Bottom panels display the overall proportion of censoring at each spatial location, for this simulation replicate. Under schemas (a) and (b), where data undergo a censoring independent in space (left), the observations are sparsely observed over the whole spatial domain. Under schemas (b) and (d), where data undergo a censoring dependent in space (right), the field is observed over all the grid points in some regions of the spatial domain, while it is totally unexplored over large portions of the spatial domain.

5.2.3 Censoring (c): dependent in space and independent in time

In the third censoring schema we swap the roles of space and time with respect to censoring (b). In this case, it is natural to look at data as evaluations of time dependent surfaces, where these surfaces may be unobserved over large regions of the spatial domain, and the regions where they are observed may be independent of that of neighbouring time instants.

To simulate this scenario, independently for each time instant, we consider a random decomposition of the spatial domain Ω into K subdomains; we then set all the observations corresponding to pK randomly selected subdomains to `NA`.

As a result, in the spatial dimension the field is observed over all the grid points in some regions of the spatial domain, while in other regions of the spatial domain it is totally unexplored, as shown in the right panels of Figure 6 (censoring dependent in space). In the temporal dimension, instead, the data are sparely observed, as shown in the left panels of

Figure 5 (censoring independent in time), likewise for censoring (a).

5.2.4 Censoring (d): dependent in space and time

The forth censoring schema corresponds to the most difficult scenario, where the observations are missing over large portions of the spatial as well as temporal domains. This corresponds to the scenario observed in the application to lake surface water temperature data detailed in Section 6.

To simulate this scenario we consider a random decomposition of the spatio-temporal domain into K subdomains and we set all the observations corresponding to pK randomly selected subdomains to `NA`.

The resulting observational patterns in space and time are those observed in the right panels of Figure 5 (censoring dependent in time) and Figure 6 (censoring dependent in space).

5.3 Competing methods

We compare the following approaches to spatio-temporal field estimation.

KRIG: Kriging. We consider a separable variogram marginally gaussian in space and exponential in time, with parameters estimated at each simulation replicate from the empirical variogram. This is implemented using the functions `krigeST` and `fit.StVariogram` of the R package `gstat` (Pebesma and Graeler, 2020).

TPS and SOAP: These methods solve analogous estimation problems as ST-PDE, considering two separate penalization terms to account for the regularity in time and in space. In time, both methods considers as regularizing term the L^2 norm of the second temporal derivative, as in (2.2), and both methods adopt a cubic B-spline basis. In space, instead, the two methods differ. TPS considers as spatial penalty the thin plate spline energy, and uses a thin plate spline basis (see Wahba, 1990). SOAP considers as spatial penalty the L^2 norm of the Laplacian, and implements the discretization of soap film smoothing (Wood et al., 2008; Marra et al., 2012; Augustin et al., 2013). Both methods are implemented using the function `gam` of the `R` package `mgcv` (Wood, 2019), choosing the smoothness parameters at each simulation replicate via GCV.

DINEOF: The Data INterpolating Empirical Orthogonal Functions method (see, e.g., Beckers and Rixen, 2003; Alvera-Azcárate et al., 2005). This method is considered only for the test on the squared domain, since it has been proposed to deal with data observed on uniform fine grids, as in remote sensing problems. It is implemented using the function `dineof()` in the R package `sinkr` (Taylor, 2017).

INLA: The sPDE approach in Cameletti et al. (2013), implemented using the `R` package `INLA`. For the simulation on the C-shaped domain, where the geometry of the domain is relevant for the accurate estimation of the field, we consider the model with the barrier approach proposed in Bakka et al. (2019).

ST-PDE: the proposed method, using as spatial penalty the L^2 norm of the Laplacian (i.e., $L = \Delta$), likewise for SOAP, and homogeneous Neumann boundary conditions. The method is implemented using the function `smooth.FEM.time` in the R package `fdaPDE` (Lila et al., 2020), choosing the smoothness parameters at each simulation replicate via GCV.

5.4 Tests on C-shaped domain

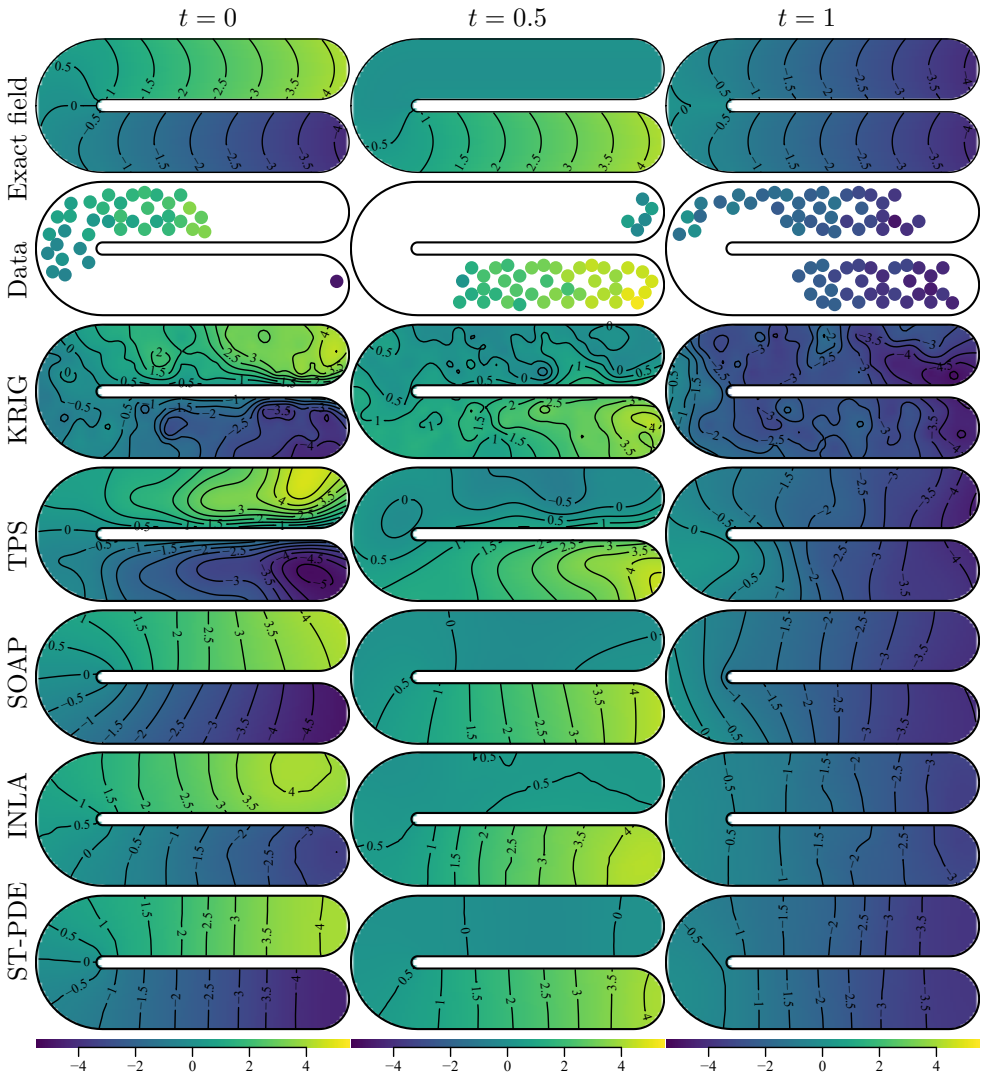

Figure 7: Test on C-shaped domain, censoring (d), proportion of censoring $p = 50\%$. The three columns of the figure correspond to different time instants. The first row shows the test function, the second row the simulated data and subsequent rows the estimates of the spatio-temporal field obtained by kriging (KRIG), thin-plate spline (TPS), soap film smoothing (SOAP), stochastic PDE (INLA), and the proposed spatio-temporal regression with PDE regularization (ST-PDE).

Figure 7 illustrates the results obtained on the C-shaped domain under the most difficult missing data scenario, censoring (d), for $p = 0.5$ and $k = 40$. The figure shows the test function, the sampled data and the estimates provided by kriging (KRIG), thin-plate spline (TPS), soap film smoothing (SOAP), sPDE (INLA) and the proposed spatio-temporal regression with PDE regularization (ST-PDE) at three different time instants. A visual inspection highlights that the methods that do not account for the shape of the domain, namely KRIG and TPS, perform poorly when the test function takes different values on the two branches of the C-shaped domain, whilst the methods that explicitly consider for the domain geometry, namely SOAP, INLA and ST-PDE, do not suffer of this problem and return accurate estimates. Focusing on these methods, we can observed that the isolines of the estimates obtained by SOAP, INLA and ST-PDE have different shapes. This is due to the different conditions imposed on the boundary. ST-PDE tends to enforce isolines perpendicular to the boundary due to Neumann b.c. here considered, while INLA tends to enforce isolines directed to the right in the concavity of the C, due to the barrier model; SOAP, instead, estimates the boundary conditions form the data, and this results in a poor estimation of the field at the boundary where data are missing, as for instance in the bottom part of the C-shaped domain for $t = 0$.

To quantitatively assess the goodness of the competing methods in terms of prediction accuracy, we compute the Root Mean Square Errors (RMSE) of the space-time field estimates $\hat{f}$. The RMSE

$$RMSE(\hat{f}) = \sqrt{\sum_{k=1}^{n^*}\sum_{l=1}^{m^*} \frac{\left(f(\mathbf{p}_k^*, t_l^*) - \hat{f}(\mathbf{p}_k^*, t_l^*)\right)^2}{n^* m^*}}.$$

is computed over a fine grid of spatio-temporal locations $\{(\mathbf{p}_k^*, t_l^*) \in \Omega \times T : k = 1 \dots n^*, l = 1, \dots, m^*\}$, of dimension 306×41. We also consider a rich set of indices concerning both predictive accuracy and uncertainty quantification. In particular, beside RMSE, we also compute the Mean Absolute Error (MAE), the Continuous Rank Probability Score (CRPS; see Gneiting and Raftery, 2007), the Interval score (INT; see Gneiting and Raftery, 2007), and the confidence interval coverage (CVG). All the indeces are computed on the same grid as the RMSE.

Method	RMSE	MAE	CRPS	INT	CVG
KRIG	0.616 (0.193)	0.652 (0.086)	0.313 (0.098)	3.079 (4.053)	0.902 (0.220)
TPS	0.371 (0.155)	0.471 (0.048)	0.171 (0.039)	2.577 (0.642)	0.749 (0.039)
SOAP	0.189 (0.072)	0.354 (0.037)	0.091 (0.021)	**1.215** (0.387)	0.786 (0.051)
INLA	0.254 (0.096)	0.381 (0.073)	0.203 (0.023)	3.133 (0.107)	**1.000** (0.001)
ST-PDE	**0.168** (0.065)	**0.324** (0.035)	**0.080** (0.023)	1.327 (0.929)	0.757 (0.068)

Table 1: Test on C-shaped domain, censoring (d). For each competing method we report the median (and inter quartile range), over 30 simulation repetitions, of Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Continuous Rank Probability Score (CRPS), Interval score (INT), interval coverage (CVG).

Table 1 contains the median (and inter quartile range) of these indices over 30 simulation

repetitions. The proposed ST-PDE gives the best results in terms of RMSE, MAE and CRPS. SOAP attains the best performances in terms of INT. Regarding the CVG, all the methods but INLA lead to undercoverage with respect to the nominal value 95%, while INLA lead to overcoverage. However, the low coverage of the standard parametric Wald-type confidence intervals here considered is a well known issue of nonparametric regression, and several approaches can be followed to construct better confidence intervals, such as bias correction and undersmoothing approaches (see, e.g., the reviews in Eubank and Speckman, 1993; Hall and Horowitz, 2013).

Figure 8 shows the boxplots of the RMSE over 30 repetitions of the same simulation, considering all the censoring schemas presented in Section 5.2, with $p = 50\%$ and $K = 8, 12, 40$ for censoring (b), (c) and (d) respectively. In general, the estimates provided by methods that appropriately account for the shape of the domain, namely SOAP, INLA and ST-PDE, have significantly lower RMSEs with respect to methods that instead are blind to the domain shape, namely KRIG and TPS. In particular, ST-PDE appears the best method under all scenarios, followed by INLA and SOAP, hinting that ST-PDE is particularly suited to cases when the shape of the domain influences the signal. From the ranges of the RMSE in the four plots, we can also see that censoring (d) is the most challenging for all the methods; under this censoring the differences in the performances of the various methods partly reduce, while still being significant.

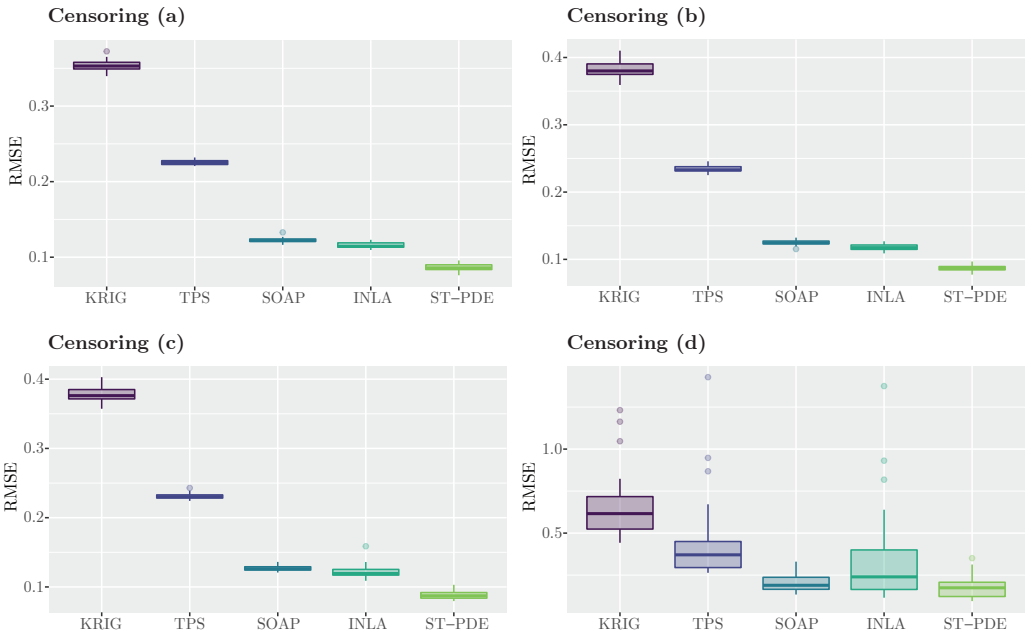

Figure 8: Tests on C-shaped domain, proportion of censoring $p = 50\%$. Boxplots of the RMSE over 30 simulation replicates of the estimates of the spatio-temporal field obtained by kriging (KRIG), thin-plate spline (TPS), soap film smoothing (SOAP), sPDE (INLA) and the proposed spatio-temporal regression with PDE regularization (ST-PDE), under the different censoring schemas.

To investigate the behaviour of the estimators with respect to the percentage of missing data, we perform an analogous simulations when $p = 0.25$ and $p = 0.75$. Figure 9, left panel, reports the mean RMSE with respect to p, for the four considered methods under censoring (a). The same behaviour is observed under all the other censoring schemas. The RMSE increases when the proportion of censoring p increases; ST-PDE attains the lowest RMSEs under all censoring proportions.

Finally, we investigate the behaviour of TPS, SOAP, INLA and the proposed method ST-PDE in presence of covariates, studying a problem on the C-shaped domain, with the same field f used in the previous simulations, and also including the space-time varying covariate $\mathbf{w}_{ij} = \cos(5x_i)\sqrt{(t_j+1)/5}$ with true $\boldsymbol{\beta}$ equal to 1. Figure 9, right panel, shows the RMSE of $\hat{\boldsymbol{\beta}}$, namely $|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}|$, over 30 repetitions of this simulation, confirming the same ordering of the methods also for the estimation of $\boldsymbol{\beta}$. The boxplots of the RMSE of $\hat{f}$, not reported here for sake of space, show the same behavior highlighted in Figure 8.

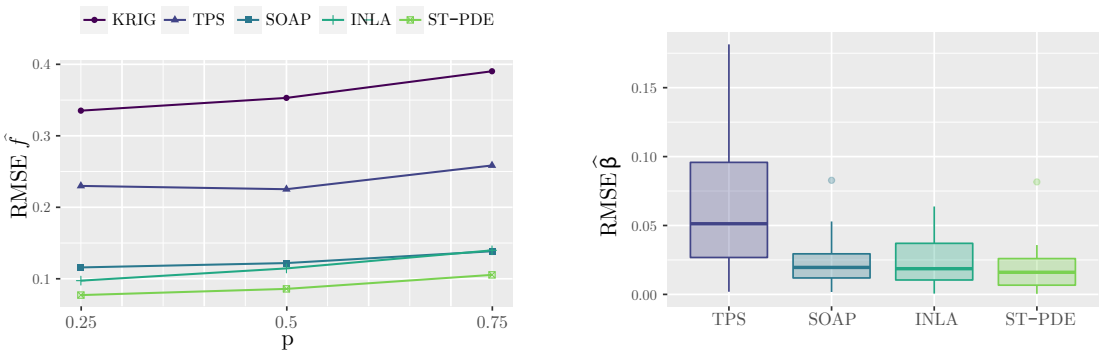

Figure 9: Tests on C-shaped domain, censoring (a). Left: different proportions of censoring, $p = 25\%, 50\%, 75\%$; median RMSE of $\hat{f}$ over 30 simulation replicates. Right: test with inclusion of covariates, $p = 50\%$; RMSE of $\hat{\boldsymbol{\beta}}$ over 30 simulation replicates.

5.5 Tests on squared domain

In previous section we commented on the results obtained in tests where the non-convex shape of the spatial domain raises difficulties for the classical methods such as kriging and space-time smoothing based on thin-plate splines. Here we consider instead a simple spatial domain, a square, but with a field displaying strong spatio-temporal variations. Also in this case, the proposed ST-PDE method is able to capture the features of the signal better than the other techniques.

Figure 10 reports the image plots of the test field on the square domain, the sample data and the estimates obtained with the competing methods, under censoring (a), when $p = 0.5$. With respect to the simulation on the C-shaped domain, we here also add a comparison to DINEOF, which can only be used for gridded data. From a visual inspection is difficult to compare the estimates; however, we can for instance notice that both TPS and SOAP tend to oversmooth the isolines where the test function has sharp variations.

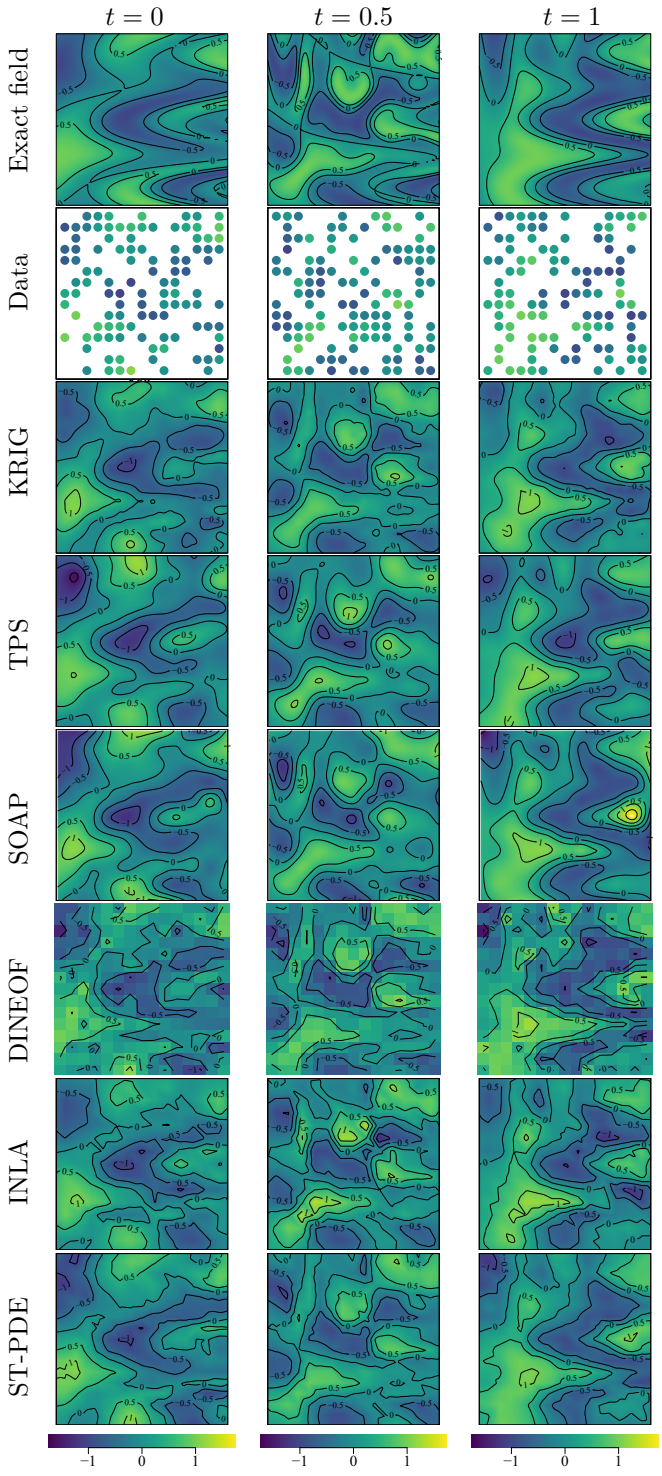

Figure 10: Censoring (a). Spatio-temporal test function, simulated data and estimates obtained by kriging (KRIG), thin-plate spline (TPS), soap film smoothing (SOAP), data interpolating empirical orthogonal functions DINEOF, stochastic PDE (INLA), and the proposed spatio-temporal regression with PDE regularization ST-PDE.

As for the previous simulation study, we assess the methods considering RMSE, MAE, CRPS, INT and CVG. All indices are evaluated on a regular lattice of 225×41 spatio-temporal locations, corresponding to the full data design. DINEOF does not estimate the standard deviation of the field; therefore, for this technique, it is not possible to compute CRPS, INT and CVG. Table 2 contains the median (and inter quartile range) of these indices over 30 simulation repetitions. With respect to the previous simulation study, we observe that in this setting, where the shape of the domain does not influences the signal, the methods that do not consider the shape of the domain have more similar performances to the others. As in the previous simulation, the proposed ST-PDE gives the best results in terms of RMSE, MAE, CRPS and INT, highlighting the ability of the method to well estimate signals with strong spatio-temporal variations. Concerning CVG, all methods but INLA lead to undercoverage with respect to the nominal value 95%; bias correction and undersmoothing approaches can be used to improve these results, as commented for the previous simulation. Finally, the Supplementary Material reports the same indices in simulation studies as the one here described, but with different signal-to-noise ratios. In all cases, the same considerations here made remain valid.

Method	RMSE	MAE	CRPS	INT	CVG
KRIG	0.123 (0.002)	0.303 (0.003)	0.083 (0.001)	2.262 (0.053)	0.496 (0.001)
TPS	0.166 (0.001)	0.348 (0.002)	0.093 (0.001)	1.416 (0.028)	0.752 (0.006)
SOAP	0.176 (0.001)	0.356 (0.001)	0.099 (0.001)	1.675 (0.037)	0.713 (0.007)
DINEOF	0.137 (0.018)	0.308 (0.015)	-	-	-
INLA	0.158 (0.002)	0.331 (0.002)	0.084 (0.001)	0.909 (0.024)	**0.9611** (0.002)
ST-PDE	**0.087** (0.002)	**0.253** (0.002)	**0.046** (0.001)	**0.471** (0.023)	0.876 (0.007)

Table 2: Test on square domain, censoring (a). For each competing method we report the median (and inter quartile range), over 30 simulation repetitions, of Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Continuous Rank Probability Score (CRPS), Interval score (INT), interval coverage (CVG).

6 Application to lake surface water temperature

In this section we study the surface water temperature of Lake Victoria using ST-PDE. As discussed in Hulley and Ghent (2019) many studies have been conducted on lakes temperature. Lakes have a high heat capacity that dampens short-term temperature variability and enhances long-term variations. In particular, lake surface water temperature (LSWT) is internationally recognised as an Essential Climate Variable and a proxy for climate change. Moreover, also ecologists are interested in studying LSWT data, to investigate environmental system dynamics.

In this context, to provide the basis for long term records of the physical state of lakes, many satellite temperature data are considered. Among them, the ARC-Lake database (see MacCallum and Merchant, 2011) offers a range of satellite-derived LSWT data sets, for thousands of lakes around the world. ARC-Lake is a project funded by the European

Space Agency, and developed by the Earth Observation and the Space Research Division at the University of Reading. The available remote sensors include the series of Advanced Along Track Scanning Radiometers, which are in principle capable to be a highly accurate source of information on LSWTs on a systematic global basis. However, due to technical problems and particular meteorological conditions, such as the presence of clouds in the atmosphere, the recorded data present a substantial amount of missing observations.

6.1 Data

The specific data set we consider here is the monthly LSWT of Lake Victoria; see Figures 1 and 2. More precisely, each datum is an average of LWST over a pixel of 0.05° longitude by 0.05° latitude, and over a month time (not considering nights). When a cell falls in part on the main land or an island, the LSWT is the average temperature over the water portion only. This average is assigned to the spatial coordinates of the centre of the pixel, resulting in 2313 locations over the lake. The observation period consists of 203 months, from June 1995 to April 2012.

The data set is characterized by a large proportion of missing values, about 45%. The bottom right panel of Figure 1 shows that the proportion of missing data is not uniform across the lake, while the bottom panel of Figure 2 highlights the high proportion of missing data in several months, including an interval of some consecutive months where no datum is available, across the whole lake.

6.2 Estimation of the temperature field via ST-PDE

We fit a ST-PDE model to the incomplete Lake Victoria data in order to reconstruct the spatio-temporal pattern over the unobserved regions. Differently from Gong et al. (2018), who restrict the analysis on a smaller rectangular portion of the lake, we here consider its entire domain, appropriately complying with its complicated shape, highlighted in Figure 3. In particular, as seen in the left panel of Figure 3, the shoreline of the lake is very jagged and there are many islands, most of them very small. Keeping this level of detail in the domain boundary is unnecessary, since the signal does not display a comparable level of detail. The right panel of the same figure shows a simplification of the domain, that maintains the most relevant features of the boundary and the main islands (for which the corresponding observations are `NA`). This simplification is obtained using the function `ms_simplify` from the R package `rmapshaper` with the Visvalingam & Whyatt Line Simplification Algorithm (Teucher and Russell, 2020). Starting from this simplified boundary, we then obtain a triangulation of the lake using the function `create.mesh.2D` of the package `fdaPDE`; this triangulation is hence refined using the function `refine.mesh.2D` imposing a minimum angle of 30° and a maximum area of 0.03 for each element of the triangulation. The resulting mesh is reported in the right panel of Figure 3.

We then fit the ST-PDE model with homogeneous Neumann boundary conditions, penalizing the Laplacian and selecting the smoothing parameters via GCV. Figure 11 shows the estimate in the first three months, which display rather different amounts of missing values.

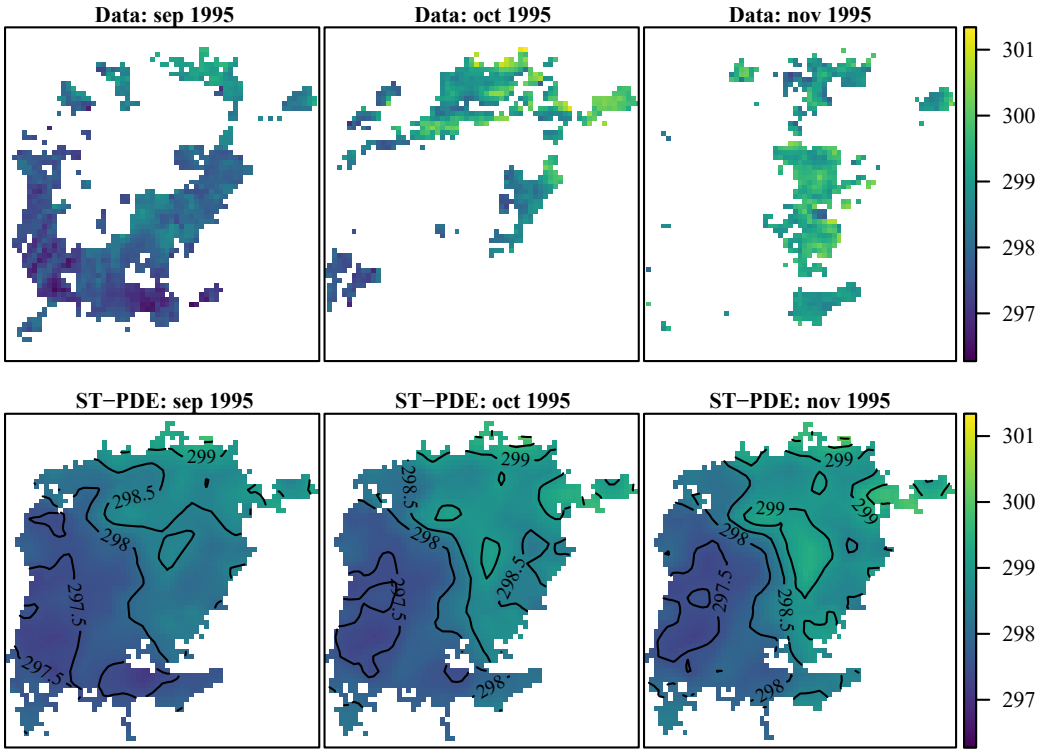

Figure 11: LWST data and corresponding ST-PDE estimate, for three months.

The figure highlights that the method is able to capture very well the features of the signal where data are available. A video of the obtained estimate is available at `t.ly/5hpH` and in the Supplementary Material.

7 Discussion and future research directions

The proposed methodology is capable to accurately estimate spatio-temporal fields from data affected by complicated missingness patterns. Simulation studies show that ST-PDE has competitive advantages over available techniques. The method is able to capture complicated signals, with strongly localized features, as well as signals observed over irregular domains.

ST-PDE can be extended in various directions. A first interesting development is to consider curved spatial domains with non-standard geometries. This extension to manifold domains can be obtained by combining the proposed method with the technique for spatial-only data in Ettinger et al. (2016); Lila et al. (2016); Wilhelm et al. (2016). This would open the possibility to consider environmental applications concerning data on the globe, but also fascinating life-sciences applications, as for instance the study of neuroimaging data associated with neural activity on the highly folded brain surface.

Moreover, the generalization to other differential models would further broaden the modeling of complex spatio-temporal phenomena. Particularly interesting in this direction would be the extensions toward PDEs parameters that are not-only space-varying but also time-varying, and more generally to non-linear PDEs (considered in the framework of state-space models by, e.g., Stroud et al., 2010).

Although we do not tackle this aspect in the current work, the PDE terms K, $\mathbf{b}$ and c could as well depend upon some unknown parameters, to be estimated from data. A possible approach in this direction relies on parameter cascading, a generalized profiling estimation procedure originally introduced by Ramsay et al. (2007) to estimate the parameters of an ordinary differential equation, starting from noisy measurements of the solution to the differential equation, and later extended by Xun et al. (2013) to the context of partial differential equations. This approach has already been successfully employed by Bernardi et al. (2018), that considers the simpler case of spatial regression models over space only, with a regularizing term involving the anisotropic diffusion operator $\nabla \cdot (K \nabla f)$, where the symmetric and positive definite diffusion tensor K, that determines the direction and intensity of anisotropy, is estimated from data. The estimation of the PDE parameters, combined with an appropriate choice of the smoothing parameters λ_S and λ_T, along the lines described by Bernardi et al. (2018), increases significantly the computational cost of the methodology. The study of appropriate computational techniques, to reduce this increased cost, will be the subject of future research.

ST-PDE could also handle multiple datasets. In particular, the principal component analysis method in Lila et al. (2016) could be naturally extended to space-time data. Moreover, the proposed method could be generalized to a mixed effect regression settings (similar to those in, e.g., Guo, 2002; Wood, 2006; Chen and Wang, 2011). This would also permit to consider possible clustering structures in the multiple datasets.

Future research will also concern the study of the asymptotic properties of the estimator in functional norms, when also the number of basis functions goes to infinity, together with number of observations, along the lines of the study in Arnone et al. (2021), which considered the case of spatial-only data.

References

Aguilera-Morillo, M. C., Durbán, M., and Aguilera, A. M. (2017). Prediction of functional data with spatial dependence: a penalized approach. *Stochastic environmental research and risk assessment*, **31**(1), 07–22.

Alvera-Azcárate, A., Barth, A., Rixen, M., and Beckers, J.-M. (2005). Reconstruction of incomplete oceanographic data sets using empirical orthogonal functions: application to the adriatic sea surface temperature. *Ocean Modelling*, **9**(4), 325–346.

Arnone, E., Azzimonti, L., Nobile, F., and Sangalli, L. M. (2019). Modeling spatially dependent functional data via regression with differential regularization. *Journal of Multivariate*

Analysis, **170**, 275–295.

Arnone, E., Kneip, A., Nobile, F., and Sangalli, L. M. (2021). Some first results on the consistency of spatial regression with partial differential equation regularization. *Statistica Sinica*. doi: 10.5705/ss.202019.0346.

Augustin, N. H., Trenkel, V. M., Wood, S. N., and Lorance, P. (2013). Space-time modelling of blue ling for fisheries stock management. *Environmetrics*, **24**(2), 109–119.

Azzimonti, L., Sangalli, L. M., Secchi, P., Domanin, M., and Nobile, F. (2015). Blood flow velocity field estimation via spatial regression with pde penalization. *Journal of the American Statistical Association*, **110**(511), 1057–1071.

Bakka, H., Vanhatalo, J., Illian, J. B., Simpson, D., and Rue, H. v. (2019). Non-stationary Gaussian models with physical barriers. *Spat. Stat.*, **29**, 268–288.

Beckers, J.-M. and Rixen, M. (2003). Eof calculations and data filling from incomplete oceanographic datasets. *Journal of Atmospheric and oceanic technology*, **20**(12), 1839–1856.

Berliner, L. M. (2003). Physical-statistical modeling in geophysics. *Journal of Geophysical Research: Atmospheres*, **108**(D24).

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

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

Cameletti, M., Lindgren, F., Simpson, D., and Rue, H. v. (2013). Spatio-temporal modeling of particulate matter concentration through the SPDE approach. *AStA Adv. Stat. Anal.*, **97**(2), 109–131.

Chen, H. and Wang, Y. (2011). A penalized spline approach to functional mixed effects model analysis. *Biometrics*, **67**(3), 861–870.

Cressie, N. and Wikle, C. K. (2015). *Statistics for spatio-temporal data*. John Wiley & Sons.

Delicado, P., Giraldo, R., Comas, C., and Mateu, J. (2010). Statistics for spatial functional data: some recent contributions. *Environmetrics*, **21**(3-4), 224–239.

Eilers, P. H. and Marx, B. D. (1996). Flexible smoothing with b-splines and penalties. *Statistical science*, **11**(2), 89–121.

Eilers, P. H. and Marx, B. D. (2021). *Practical Smoothing: The Joys of P-splines*. Cambridge University Press.

Ettinger, B., Perotto, S., and Sangalli, L. M. (2016). Spatial regression models over two-dimensional manifolds. *Biometrika*, **103**(1), 71–88.

Eubank, R. L. and Speckman, P. L. (1993). Confidence bands in nonparametric regression. *Journal of the American Statistical Association*, **88**(424), 1287–1301.

Ferraty, F. and Vieu, P. (2006). *Nonparametric functional data analysis*. Springer Series in Statistics. Springer, New York.

Fuglstad, G.-A., Lindgren, F., Simpson, D., and Rue, H. (2015). Exploring a new class of non-stationary spatial gaussian random fields with varying local anisotropy. *Statistica Sinica*, pages 115–133.

Gneiting, T. and Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American statistical Association*, **102**(477), 359–378.

Gong, M., Miller, C., and Scott, M. (2018). Spatio-temporal modelling of remote-sensing lake surface water temperature data. *33rd International Workshop on Statistical Modelling (IWSM 2018)*, pages 106–111.

Gong, M., Miller, C., Scott, M., O'Donnell, R., Simis, S., Groom, S., Tyler, A., Hunter, P., and Spyrakos, E. (2021). State space functional principal component analysis to identify spatiotemporal patterns in remote sensing lake water quality. *Stochastic Environmental Research and Risk Assessment*, pages 1–16.

Gray, R. J. (1994). Spline-based tests in survival analysis. *Biometrics*, pages 640–652.

Gromenko, O., Kokoszka, P., Sojka, J., et al. (2017). Evaluation of the cooling trend in the ionosphere using functional regression with incomplete curves. *The Annals of Applied Statistics*, **11**(2), 898–918.

Guo, W. (2002). Functional mixed effects models. *Biometrics*, **58**(1), 121–128.

Hall, P. and Horowitz, J. (2013). A simple bootstrap method for constructing nonparametric confidence bands for functions. *Ann. Statist.*, **41**(4), 1892–1921.

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

Holland, A. D. (2017). Penalized spline estimation in the partially linear model. *J. Multivariate Anal.*, **153**, 211–235.

Horváth, L. and Kokoszka, P. (2012). *Inference for functional data with applications*. Springer Series in Statistics. Springer, New York.

Hulley, G. and Ghent, D. (2019). *Taking the Temperature of the Earth 1st Edition. Steps towards Integrated Understanding of Variability and Change*. Elsevier.

Kokoszka, P. and Reimherr, M. (2017). *Introduction to functional data analysis*. Chapman and Hall/CRC.

Kuhnert, P. M. (2014). Physical-statistical modeling. *Wiley StatsRef: Statistics Reference Online*, pages 1–5.

Lila, E., Aston, J. A., Sangalli, L. M., et al. (2016). Smooth principal component analysis over two-dimensional manifolds with an application to neuroimaging. *The Annals of Applied Statistics*, **10**(4), 1854–1879.

Lila, E., Sangalli, L. M., Arnone, E., Ramsay, J., and Formaggia, L. (2020). fdaPDE: Functional data analysis and partial differential equations (PDE); statistical analysis of functional and spatial data, based on regression with pde regularization. R package version 1.0-8.

Lindgren, F., Rue, H. v., and Lindström, J. (2011). An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, **73**(4), 423–498.

Liu, C., Ray, S., and Hooker, G. (2017). Functional principal component analysis of spatially correlated data. *Statistics and Computing*, **27**(6), 1639–1654.

MacCallum, S. N. and Merchant, C. J. (2011). Arc-lake algorithm theoretical basis document–arc-lake v1. 1, 1995–2009 [dataset].

Malmberg, A., Arellano, A., Edwards, D. P., Flyer, N., Nychka, D., and Wikle, C. (2008). Interpolating fields of carbon monoxide data using a hybrid statistical-physical model. *Ann. Appl. Stat.*, **2**(4), 1231–1248.

Marra, G., Miller, D. L., and Zanin, L. (2012). Modelling the spatiotemporal distribution of the incidence of resident foreign population. *Statistica Neerlandica*, **66**(2), 133–160.

Mateu, J. and Romano, E. (2017). Advances in spatial functional statistics. *Stochastic Environmental Research and Risk Assessment*, **31**, 1–6.

Menafoglio, A. and Secchi, P. (2017). Statistical analysis of complex and spatially dependent data: a review of Object Oriented Spatial Statistics. *European J. Oper. Res.*, **258**(2), 401–410.

Pebesma, E. and Graeler, B. (2020). Spatial and Spatio-Temporal Geostatistical Modelling, Prediction and Simulation. R package version 2.0-5.

Ping, B., Su, F., and Meng, Y. (2016). An improved dineof algorithm for filling missing values in spatio-temporal sea surface temperature data. *PloS one*, **11**(5).

Ramsay, J. O. and Silverman, B. W. (2005). *Functional data analysis*. Springer Series in Statistics. Springer, New York, second edition.

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

Ramsay, T. (2002). Spline smoothing over difficult regions. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, **64**(2), 307–319.

Richardson, R. A. (2017). Sparsity in nonlinear dynamic spatiotemporal models using implied advection. *Environmetrics*, **28**(6), e2456, 16.

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

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

Schoenberg, I. J. and Whitney, A. (1953). On pólya frequence functions. iii. the positivity of translation determinants with an application to the interpolation problem by spline curves. *Transactions of the American Mathematical Society*, **74**(2), 246–259.

Sigrist, F., Künsch, H. R., and Stahel, W. A. (2015). Stochastic partial differential equation based modelling of large space-time data sets. *J. R. Stat. Soc. Ser. B. Stat. Methodol.*, **77**(1), 3–33.

Stroud, J. R., Stein, M. L., Lesht, B. M., Schwab, D. J., and Beletsky, D. (2010). An ensemble Kalman filter and smoother for satellite data assimilation. *J. Amer. Statist. Assoc.*, **105**(491), 978–990.

Taylor, M. (2017). sinkr package. https://github.com/marchtaylor/sinkr. Accessed: 2020-09-06.

Teucher, A. and Russell, K. (2020). rmapshaper: Client for 'mapshaper' for 'geospatial' operations. R package version 0.4.4.

Ugarte, M., Goicoa, T., Militino, A., and Durbán, M. (2009). Spline smoothing in small area trend estimation and forecasting. *Computational Statistics & Data Analysis*, **53**(10), 3616–3629.

Ugarte, M., Goicoa, T., and Militino, A. (2010). Spatio-temporal modeling of mortality risks using penalized splines. *Environmetrics: The official journal of the International Environmetrics Society*, **21**(3-4), 270–289.

Wahba, G. (1990). *Spline models for observational data*. SIAM.

Whittle, P. (1954). On stationary processes in the plane. *Biometrika*, pages 434–449.

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

Wikle, C. K. and Hooten, M. B. (2010). A general science-based framework for dynamical spatio-temporal models. *TEST*, **19**(3), 417–451.

Wilhelm, M., Dedè, L., Sangalli, L. M., and Wilhelm, P. (2016). IGS: an IsoGeometric approach for smoothing on surfaces. *Comput. Methods Appl. Mech. Engrg.*, **302**, 70–89.

Wood, S. (2019). Mixed GAM Computation Vehicle with Automatic Smoothness Estimation. R package version 1.8-31.

Wood, S. N. (2006). Low-rank scale-invariant tensor product smooths for generalized additive mixed models. *Biometrics*, **62**(4), 1025–1036.

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

Xu, K. and Wikle, C. K. (2007). Estimation of parameterized spatio-temporal dynamic models. *J. Statist. Plann. Inference*, **137**(2), 567–588.

Xun, X., Cao, J., Mallick, B., Maity, A., and Carroll, R. J. (2013). Parameter estimation of partial differential equation models. *J. Amer. Statist. Assoc.*, **108**(503), 1009–1020.

Yu, Y. and Ruppert, D. (2002). Penalized spline estimation for partially linear single-index models. *Journal of the American Statistical Association*, **97**(460), 1042–1054.

Supplementary material for the paper: Smoothing spatio-temporal data with complex missing data patterns

Eleonora Arnone [1], Laura M. Sangalli [1], and Andrea Vicini [1]

[1] MOX - Dipartimento di Matematica, Politecnico di Milano, Italy

Address for correspondence: Laura M. Sangalli, MOX - Dipartimento di Matematica, Politecnico di Milano, Piazza Leonardo da Vinci 32, 20133 Milano, Italy.
E-mail: laura.sangalli@polimi.it.
Phone: (+39) 02 2399 4554.
Fax: (+39) 02 2399 4568.

Abstract: Supplemental material.

Key words: functional data with spatial dependence; incomplete and partially observed functional data; nonparametric regression with partial differential equation regularization; smoothing with roughness penalties

1 Proofs of asymptotic results

Proof of Theorem 1. We first prove that the ST-PDE estimator is consistent. Consider the mean square error for the estimator

$$\mathbb{E}(\hat{\mathbf{f}} - \mathbf{f})^2 = b_\nu(\lambda_{S,\nu}, \lambda_{T,\nu}) b_\nu(\lambda_{S,\nu}, \lambda_{T,\nu})^\top + \mathrm{Var}_\nu(\lambda_{S,\nu}, \lambda_{T,\nu})$$

where $b_\nu(\lambda_{S,\nu}, \lambda_{T,\nu})$ and $\mathrm{Var}_\nu(\lambda_{S,\nu}, \lambda_{T,\nu})$ are the bias and the variance of the estimator obtained with ν observations and $\lambda_{S,\nu}$ and $\lambda_{T,\nu}$ as smoothing parameters, respectively. To obtain the consistency, we need $\mathbb{E}(\hat{\mathbf{f}} - \mathbf{f})^2 \to 0$.

For the bias we have:

$$\begin{aligned}
b_\nu(\lambda_{S,\nu}, \lambda_{T,\nu}) &= \mathbb{E}(\hat{\mathbf{f}}_\nu) - \mathbf{f} \\
&= \left[\left(B^\top Q B/\nu + \lambda_{S,\nu} P_S + \lambda_{T,\nu} P_T \right)^{-1} B^\top Q B/\nu - I \right] \mathbf{f} \\
&= \left[\left(A_\nu^{-1} + \lambda_{S,\nu} P_S + \lambda_{T,\nu} P_T \right)^{-1} A_\nu^{-1} - I \right] \mathbf{f} \\
&= -\lambda_{S,\nu} A_\nu P_s \mathbf{f} - \lambda_{T,\nu} A_\nu P_T \mathbf{f} + o\left(\lambda_{S,\nu} + \lambda_{T,\nu}\right)
\end{aligned}$$

where we have used Taylor expansion and Assumption 1. Analogously, for the variance we have:

$$\begin{aligned}
\mathrm{Var}_\nu\left(\lambda_{S,\nu},\lambda_{T,\nu}\right) &= \frac{\sigma^2}{\nu}\left(B^\top QB/\nu + P\right)^{-1} B^\top QB/\nu \left(B^\top QB/\nu + P\right)^{-1} \\
&= \frac{\sigma^2}{\nu}\left(A_\nu^{-1} + P\right)^{-1} A_\nu^{-1}\left(A_\nu^{-1} + P\right)^{-1} \\
&= \frac{\sigma^2}{\nu}\left(I - A_\nu P\right)^{-1}\left(I - A_\nu P\right)^{-1} A_\nu = \frac{\sigma^2}{\nu}\left(I - A_\nu P\right)^{-2} A_\nu \\
&= \frac{\sigma^2}{\nu}\left(I - \lambda_{S,\nu} A_\nu P_S - \lambda_{T,\nu} A_\nu P_T\right)^{-2} A_\nu \\
&= \frac{\sigma^2}{\nu} A_\nu + o\left(\lambda_{S,\nu} + \lambda_{T,\nu}\right).
\end{aligned}$$

Under the hypothesis that both $\lambda_{S,\nu}$ and $\lambda_{T,\nu}$ goes to zero we thus have that $\mathbb{E}(\hat{\mathbf{f}} - \mathbf{f})^2 \to 0$ as $\nu \to \infty$.

In order to obtain the asymptotic distribution of $\sqrt{\nu}(\hat{\mathbf{f}} - \mathbf{f})$, first of all note that $\mathbb{E}(\sqrt{\nu}(\hat{\mathbf{f}} - \mathbf{f}))$ goes to zero only if $\lambda_{S,\nu}$ and $\lambda_{T,\nu}$ are $o(\nu^{-1/2})$. Then, recall that

$$\begin{aligned}
\hat{\mathbf{f}}_\nu &= \left(B^\top QB/\nu + P\right)^{-1} B^\top Qz/\nu \\
&= \left(B^\top QB/\nu + P\right)^{-1} B^\top Q(W\boldsymbol{\beta} + B\mathbf{f} + \boldsymbol{\epsilon})/\nu.
\end{aligned}$$

Rearranging the above equation and noting that QW has all entries equal to zero we obtain

$$\left(B^\top QB/\nu + P\right)(\hat{\mathbf{f}}_\nu - \mathbf{f}) + P\mathbf{f} = B^\top Q\boldsymbol{\epsilon}/\nu.$$

The right side of the equation above can be used as a pivot to derive the asymptotic distribution of the estimator.

□

Proof of Theorem 2. Given $\hat{\mathbf{f}}$, the vector $\hat{\boldsymbol{\beta}}$ is the solution of the score equation

$$\frac{1}{\nu} W^\top(\mathbf{z} - W\hat{\boldsymbol{\beta}} - B\hat{\mathbf{f}}) = 0.$$

Recalling that $\mathbf{z} = W\boldsymbol{\beta} + \mathbf{f} + \boldsymbol{\epsilon}$, we get

$$\frac{1}{\nu} W^\top W(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) + \frac{1}{\nu} W^\top B(\hat{\mathbf{f}} - \mathbf{f}) = \frac{1}{\nu} W^\top \boldsymbol{\epsilon}$$

where $\boldsymbol{\epsilon}$ represents the vector of i.i.d. errors. We thus obtain

$$\hat{\Sigma}_\nu(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = \frac{1}{\nu} W^\top \boldsymbol{\epsilon} - \frac{1}{\nu} W^\top B(\hat{\mathbf{f}} - \mathbf{f}). \tag{1.1}$$

Recalling that, thanks to Theorem 1, $\hat{\mathbf{f}}$ is a consistent estimator for $\mathbf{f}$, we can use the right side of the equation above as a pivoting quantity to derive the asymptotic normality of $\hat{\boldsymbol{\beta}}$.

In order to prove the consistency of $\hat{\boldsymbol{\beta}}$ it remains to show that $\mathbb{E}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^2 \to 0$. By squaring and taking the expectation of equation (1.1), and recalling that the term $(\hat{\mathbf{f}} - \mathbf{f})$ is independent from $\boldsymbol{\epsilon}$ by construction, we have that $\mathbb{E}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^2$ converges to zero if both $\frac{1}{\nu^2}\mathbb{E}(\boldsymbol{\epsilon}^\top W W^\top \boldsymbol{\epsilon})$ and $\mathbb{E}(\hat{\mathbf{f}} - \mathbf{f})^2$ goes to zero. The first term goes to zero by the central limit theorem, while the second term goes to zero thanks to Theorem 1.

□

2 Test field on C-shaped domain

The test field $f(\mathbf{p}, t)$ on the C-shaped domain is defined as

$$f(x,y,t) = \begin{cases} \cos(t)(q+x) + (y-r)^2 & \text{if} \quad x \geq 0 \wedge y > 0 \\ \cos(2t)(-q-x) + (-y-r)^2 & \text{if} \quad x \geq 0 \wedge y \leq 0 \\ \cos(t)\left(-\arctan\left(\frac{y}{x}\right) r\right) + (\sqrt{x^2+y^2} - r)^2 K(x,y) & \text{if} \quad x < 0 \wedge y > 0 \\ \cos(2t)\left(-\arctan\left(\frac{y}{x}\right) r\right) + (\sqrt{x^2+y^2} - r)^2 K(x,y) & \text{if} \quad x < 0 \wedge y \leq 0 \end{cases}$$

where

$$K(x,y) = \left(\frac{y}{r_0} 1_{|y| \leq r_0 \wedge x > -r} + 1_{|y| > r_0 \vee x \leq -r}\right)^2$$

and $r_0 = 0.1$, $r = 0.5$ and $q = \pi r/2$.

3 Test field on squared domain

The test field $f(\mathbf{p}, t)$ on the squared domain is shown in Figure 12 and is defined as

$$\begin{aligned} f(x,y,t) = &\sin\left(2\pi\left(h(y)x\cos\left(\frac{9}{5}t - 2\right) - y\sin\left(\frac{9}{5}t - 2\right)\right)\right) \\ &\cdot \cos\left(2\pi\left(h(y)x\cos\left(\frac{9}{5}t - 2 + \frac{\pi}{2}\right) + h(x)y\sin\left(\left(\frac{9}{5}t - 2\right)\frac{\pi}{2}\right)\right)\right) \end{aligned}$$

where $h(x) = \frac{1}{2}\sin(5\pi x)e^{-x^2} + 1$.

3.1 Tests for different signal-to-noise ratio

Table 3 and 4 contain the results for simulation studies on the square domain, as in Section 5.5 of the main paper, but with different signal-to-noise ratio, setting the standard deviation σ of the error respectively to $0.20 \cdot$ `range(exact data)` and to $0.01 \cdot$ `range(exact data)`. As for the simulation case considered in Section 5.5 of the main paper, ST-PDE achieves the best results in terms of RMSE, MAE, CRPS and INT in all cases, with a stronger

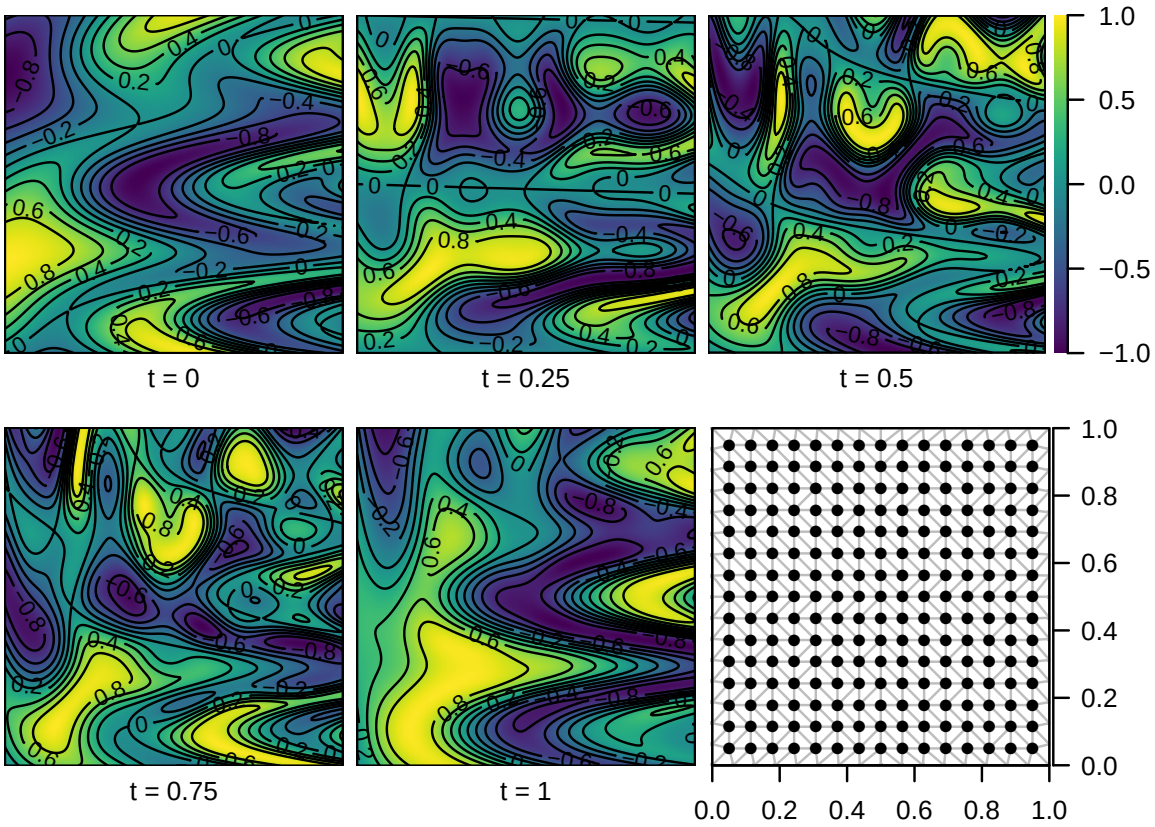

Figure 12: Test field on the c-shaped domain. The field exhibits a complicated behaviors and localized features. The last panel shows the observational spatial grid corresponding to the full (uncensored) design, and the triangulation used for ST-PDE estimate.

advantage over the competing methods for high signal-to-noise ratio. TPS, SOAP and INLA have comparable performances in terms of RMSE and MAE. INLA always achieves the best CVG, with coverage slightly higher then the nominal value 0.95. ST-PDE is the second best method in terms of CVG. However, as already noted in the main paper the standard parametric Wald-type confidence intervals here considered have low coverage for the nonparametric regressors (TPS, SOAP and ST-PDE); this is a well known issue, and several approaches can be followed to construct better confidence intervals, such as bias correction and undersmoothing strategies.

Method	RMSE	MAE	CRPS	INT	CVG
KRIG	0.367 (0.004)	0.539 (0.003)	0.252 (0.003)	7.155 (0.106)	0.483 (0.006)
TPS	0.221 (0.004)	0.416 (0.004)	0.125 (0.002)	1.255 (0.027)	0.861 (0.005)
SOAP	0.228 (0.003)	0.420 (0.004)	0.129 (0.003)	1.429 (0.096)	0.831 (0.013)
DINEOF	0.370 (0.014)	0.539 (0.014)	-	-	-
INLA	0.222 (0.003)	0.416 (0.003)	0.126 (0.001)	1.171 (0.017)	**0.983** (0.002)
ST-PDE	**0.193** (0.005)	**0.389** (0.005)	**0.109** (0.003)	**1.080** (0.054)	0.849 (0.014)

Table 3: Test on square domain, censoring (a), noise standard deviation $\sigma = 0.20 \cdot$ `range(exact data)`. For each competing method we report the median (and inter quartile range), over 20 simulation repetitions, of Root Mean Square Error (RMSE), mean absolute error (MAE), continuous rank probability score (CRPS), interval score (INT), interval coverage (CVG).

Method	RMSE	MAE	CRPS	INT	CVG
KRIG	0.081 (0.004)	0.211 (0.004)	0.041 (0.001)	0.717 (0.014)	0.495 (0.002)
TPS	0.160 (0.001)	0.336 (0.002)	0.089 (0.001)	1.504 (0.019)	0.727 (0.004)
SOAP	0.170 (0.001)	0.346 (0.001)	0.096 (0.001)	1.789 (0.029)	0.683 (0.007)
DINEOF	0.066 (0.021)	0.166 (0.015)	-	-	-
INLA	0.148 (0.001)	0.308 (0.004)	0.077 (0.001)	0.894 (0.014)	**0.955** (0.004)
ST-PDE	**0.038** (0.003)	**0.142** (0.002)	**0.015** (0.001)	**0.196** (0.021)	0.922 (0.009)

Table 4: Test on square domain, censoring (a), noise standard deviation $\sigma = 0.01 \cdot$ `range(exact data)`. Results reported as in Table 3.