



MOX-Report No. 29/2024

**Functional principal component analysis for incomplete space-time  
data**

Palummo, A.; Arnone, E.; Formaggia, L.; Sangalli, L.M.

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

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

<https://mox.polimi.it>

# Functional principal component analysis for incomplete space-time data

Alessandro Palummo<sup>1</sup>, Eleonora Arnone<sup>2</sup>, Luca Formaggia<sup>1</sup>,  
Laura M. Sangalli<sup>1\*</sup>

<sup>1</sup>MOX - Department of Mathematics, Politecnico di Milano, Italy.

<sup>2</sup>Department of Management, Univesità degli Studi di Torino, Italy.

\*Corresponding author(s). E-mail(s): [laura.sangalli@polimi.it](mailto:laura.sangalli@polimi.it);  
Contributing authors: [alessandro.palummo@polimi.it](mailto:alessandro.palummo@polimi.it);  
[eleonora.arnone@unito.it](mailto:eleonora.arnone@unito.it); [luca.formaggia@polimi.it](mailto:luca.formaggia@polimi.it);

## Abstract

Environmental signals, acquired, e.g., by remote sensing, often present large gaps of missing observations in space and time. In this work, we present an innovative approach to identify the main variability patterns, in space-time data, when data may be affected by complex missing data structures. We formalise the problem in the framework of Functional Data Analysis, proposing an innovative method of functional Principal Component Analysis (fPCA) for incomplete space-time data. The functional nature of the proposed method permits to borrow information from measurements observed at nearby spatio-temporal locations. The resulting functional principal components are smooth fields over the considered spatio-temporal domain, and can lead to interesting insights in the spatio-temporal dynamic of the phenomenon under study. Moreover, they can be used to provide a reconstruction of the missing entries, also under severe missing data patterns. The proposed model combines a weighted rank-one approximation of the data matrix with a roughness penalty. We show that the estimation problem can be solved using a Majorize-Minimization approach, and we provide a numerically efficient algorithm for its solution. Thanks to a discretization based on finite elements in space and B-splines in time, the proposed method can handle multidimensional spatial domains with complex shapes, such as water bodies with complicated shorelines, or curved spatial regions with complex orography. As shown by simulation studies, the proposed space-time fPCA is superior to alternative techniques for Principal Component Analysis with missing data. We further highlight the potentiality of the proposed method for environmental problems, by applying space-time fPCA to the study of the Lake Water Surface Temperature (LWST) of Lake Victoria, in Central Africa, starting from satellite measurements with

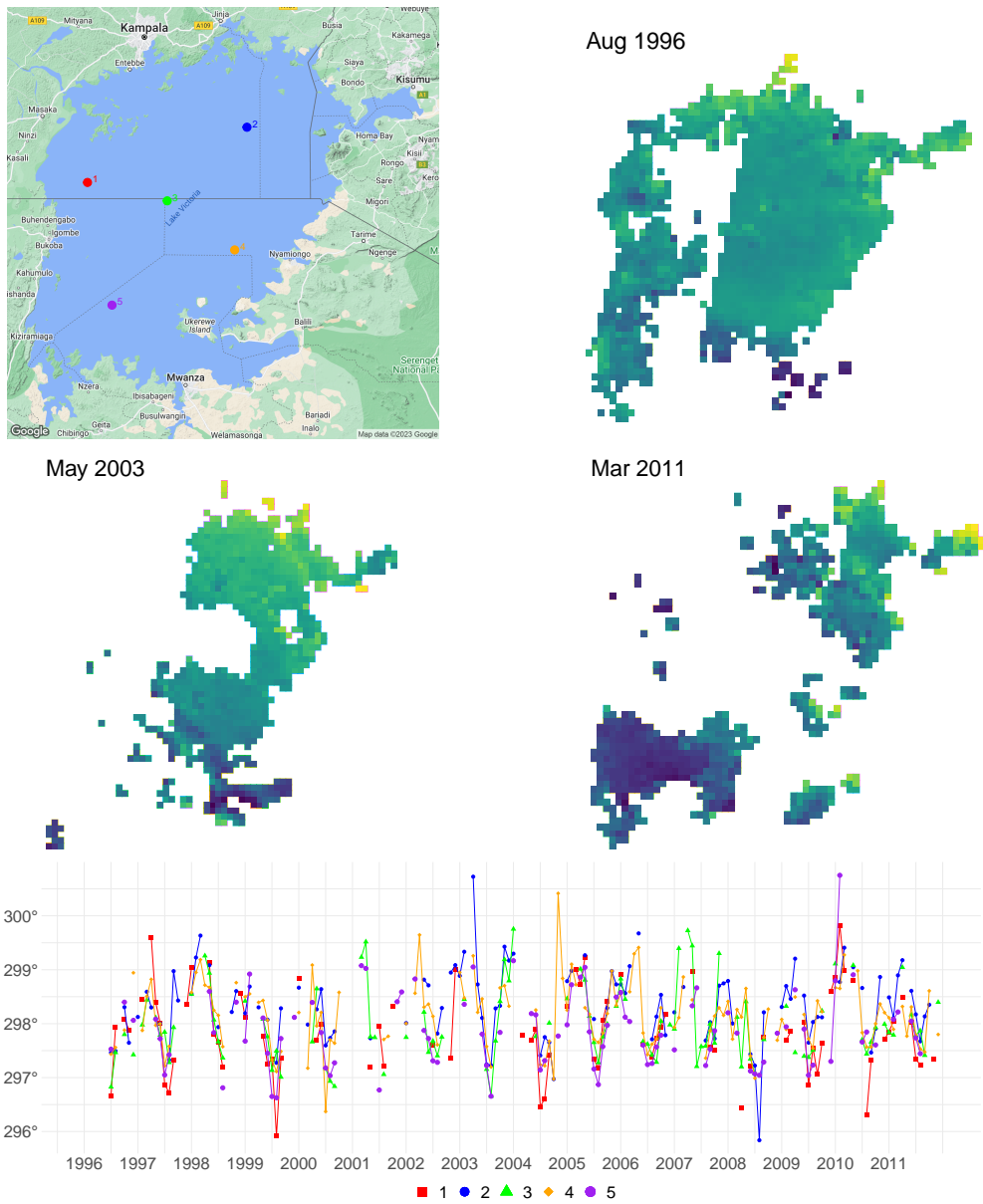
large gaps. LWST is considered one of the fundamental indicators of how climate change is affecting the environment, and is recognized as an Essential Climate Variable.

**Keywords:** Functional Data Analysis, Majorization-Minimization, Missing data, Roughness Penalties

## 1 Introduction

In environmental and ecological sciences, it is of fundamental interest to analyze signals acquired across space and time, using remote sensing or other measuring devices. However, such signals are often only partially observed, over the spatio-temporal domain, and may present complex missing data patterns. Air pollution datasets, for instance, often display a high percentage of missing values, due to faults in the measuring devices. Satellite remote sensing data, that can be used to explore vegetation indices or surface temperature over lands, seas or lakes, are often affected by large gaps, in space and time, which might be caused, e.g., by ice coverage, presence of clouds or other meteorological conditions. Figure 1 offers an example in this sense. We here display the spatio-temporal profile of the water surface temperature of Lake Victoria, in Central Africa. These data, analyzed, for example, by [Gong et al. \(2018\)](#) and [Gong et al. \(2021\)](#), are provided by the ARC-Lake database (see, e.g., [MacCallum and Merchant, 2011](#)). Although the data consist of monthly averaged measurements, data may be missing for many consecutive months, on large portions of the lake.

In this work, we are interested in investigating the main patterns of variability in spatio-temporal signals, which may be affected by complex missing data structures, such as those highlighted above. We do so in the framework of Principal Component Analysis (PCA). In this respect, it should be noted that the presence of missing data may challenge or invalidate standard approaches to PCA. For this reason, alternative strategies have been explored in the literature, to perform PCA with missing data, relying on iterative procedures, which combine PCA with missing data imputation. These iterative PCA techniques are motivated by the results of, e.g. [Gabriel and Zamir \(1979\)](#) and [Kiers \(1997\)](#), in the context of a weighted low-rank approximation. For example, the Data INterpolating Empirical Orthogonal Function (DINEOF) method ([Beckers and Rixen, 2003](#)) updates its reconstruction of the missing entries by Singular Value Decomposition on the imputed data, until convergence. An analogous technique is proposed by [Josse et al. \(2011\)](#) and [Josse and Husson \(2012\)](#), who describe a regularized iterative PCA algorithm, which reduces the possibility of overfit. These approaches are extensively employed in environmental and ecological applications, where satellite remote sensing data are of interest (see, e.g., [Hilborn and Costa, 2018](#); [Wang and Liu, 2014](#); [Alvera-Azcárate et al., 2007, 2005](#)). DINEOF is arguably the most popular approach in these fields. However, these techniques work on a multivariate assumption, and do not take advantage of the spatio-temporal nature of the phenomena under study.



**Fig. 1:** Monthly averaged satellite measurements of Lake Water Surface Temperature (LWST), Lake Victoria, Central Africa. Top left: map of Lake Victoria. Top right and center: LWST in the months of August 1996, May 2003, and March 2011. Bottom: LWST at the 5 spatial locations in the lake, indicated by color markers in the map in the top-left panel.

Here, we propose an innovative PCA method for incomplete spatio-temporal signals. To appropriately borrow information from measurements observed at nearby spatio-temporal locations, we formalize the problem of PCA in a Functional Data Analysis framework (Ramsay and Silverman, 2005; Ferraty, 2006; Kokoszka and Reimherr, 2017). Functional PCA approaches for space-time data are considered, for instance, in Li and Guan (2014), where a method based on Poisson maximum likelihood is proposed to provide an estimation of the covariance function for the spatio-temporal data generation process, from which principal components are extracted by means of an eigenvalue decomposition. In Liu et al. (2017) the authors develop a technique for the functional principal component analysis of spatially correlated functional data, which is then used as curve reconstruction method in the context of partially observed functional data.

Our proposal of functional Principal Component Analysis (fPCA) originates from a different literature, based on penalized rank-one approximations of the data matrix (see, e.g., Huang et al., 2008). In particular, we consider an estimation functional that combines a weighted rank-one approximation of the data matrix with roughness penalties based on partial differential operators over space and time. The obtained functional principal components are smooth fields over the considered spatio-temporal domain. They are easy to interpret and can lead to interesting insights in the spatio-temporal dynamic of the phenomenon under study. Moreover, they can be used to provide a reconstruction of the missing entries, also under severe missing data patterns.

To minimize the considered fPCA estimation functional, we develop an appropriate Majorization-Minimization algorithm (see, e.g., Lange, 2016). This approach is used in a variety of statistical methods, such as multidimensional scaling (Heiser, 1987a) and correspondence analysis (Heiser, 1987b). In particular, the popular Expectation-Maximization method, widely used in all areas of statistics, is a special case of the Majorization-Minimization approach (Lange and Zhou, 2022). An interesting property of these optimization approaches is that they guarantee convergence to a local optimum (Wu, 1983). For the proposed fPCA problem, we show that the estimation functional of fPCA with incomplete space-time data can be majorized by the estimation functional of fPCA with fully observed space-time data. Moreover, the latter estimation problem can be seen as an extension to space-time settings of the fPCA approaches considered by Lila et al. (2016) and Arnone et al. (2023) over space-only domains. We discretize the estimation problem using B-splines over the temporal domain, and finite elements defined over a triangulation of the spatial domain. This enables the methodology to deal with data observed over spatial domains with complex shapes, such as water bodies with complicated shorelines, or curved spatial regions with complex orography. The proposed fPCA is a new addition to the class of Physics-Informed Spatial and Functional Data Analysis methods, reviewed in Sangalli (2021), and is implemented in the R/C++ library `fdapDE` (Arnone et al., 2023).

Simulation studies demonstrate the good performances of the proposed fPCA for incomplete space-time data, and its advantages over state-of-the-art techniques for PCA with missing data. These simulation studies consider different incomplete data settings, including sparse data and data with large gaps in space and time, as in the case of the considered application to the study of water surface temperature of Lake Victoria.

The paper is organized as follows. Section 2 introduces the proposed fPCA for incomplete spatio-temporal data. Section 3 describes the discretization of the estimation problem. Section 4 details the Majorize-Minimization algorithm for the minimization of the proposed estimation functional. Section 5 reports the simulation studies, that compare the proposal to popular approaches for PCA with missing data. Section 6 illustrates the application to the study of the surface water temperature of Lake Victoria. Some concluding remarks are drawn in Section 7. All proofs are deferred to A and B.

## 2 Mathematical framework

In this section we introduce the fPCA problem for incomplete space-time data. In Section 2.1 we give the theoretical definition of functional principal component analysis, for a random field defined over a spatio-temporal domain. In Section 2.2 we introduce the fPCA estimation problem for incomplete space-time data.

### 2.1 Functional Principal Component Analysis of a random field on a spatio-temporal domain

Let  $\mathcal{D}$  be a bounded and possibly non-convex subset of  $\mathbb{R}^d$ , with  $d = 2, 3$ , and let  $T \subset \mathbb{R}$  be a time interval. Introduce the space of square-integrable functions on the spatio-temporal domain  $\mathcal{D} \times T$

$$L^2(\mathcal{D} \times T) = \left\{ f : \mathcal{D} \times T \rightarrow \mathbb{R} : \left( \int_{\mathcal{D} \times T} f(\mathbf{p}, t)^2 \right)^{\frac{1}{2}} < \infty \right\}$$

with inner product  $\langle f, g \rangle_{\mathcal{D} \times T} = \int_{\mathcal{D} \times T} f(\mathbf{p}, t)g(\mathbf{p}, t)d\mathbf{p}dt$ . Consider a random field  $\mathcal{X}$  taking values in  $L^2(\mathcal{D} \times T)$ , with mean  $\mu = \mathbb{E}[\mathcal{X}]$ , and assume it has a finite second moment, i.e.,  $\int_{\mathcal{D} \times T} \mathbb{E}[\mathcal{X}^2] < \infty$ , and a square integrable covariance function  $K((\mathbf{p}_1, t_1), (\mathbf{p}_2, t_2))$ . Define the covariance operator  $V : L^2(\mathcal{D} \times T) \rightarrow L^2(\mathcal{D} \times T)$  as  $Vf = \int_{\mathcal{D}} \int_T K(\cdot, (\mathbf{p}, t))f(\mathbf{p}, t)d\mathbf{p}dt$ . Thanks to Mercer's Lemma (Riesz and Nagy, 2012), there exists an orthonormal sequence  $\{f_k\}_k$  of eigenfunctions and a non-increasing sequence  $\{\xi_k\}_k$  of eigenvalues such that the following eigenvalue problem holds

$$(Vf_k)(\mathbf{p}, t) = \xi_k f_k(\mathbf{p}, t) \quad \forall (\mathbf{p}, t) \in \mathcal{D} \times T, k \in \mathbb{N}.$$

Moreover we can express  $K((\mathbf{p}_1, t_1), (\mathbf{p}_2, t_2)) = \sum_{k=1}^{\infty} \xi_k f_k(\mathbf{p}_1, t_1)f_k(\mathbf{p}_2, t_2)$  and  $\mathcal{X}(\mathbf{p}, t) = \mu(\mathbf{p}, t) + \sum_{k=1}^{\infty} s^{[k]} f_k(\mathbf{p}, t)$ , where  $\{s^{[k]}\}_k$  is a sequence of zero-mean uncorrelated random variables, with  $s^{[k]} = \int_{\mathcal{D} \times T} (\mathcal{X} - \mu)(\mathbf{p}, t)f_k(\mathbf{p}, t)$ . The functions  $\{f_k\}_k$  are named Principal Component (PC) functions, whereas the random variables  $\{s^{[k]}\}_k$  are named Principal Component scores. PC functions  $\{f_k\}_k$  identify the strongest modes of variation in the random field  $\mathcal{X}$ . In fact, it can be shown that  $f_1$  is such that

$$f_1 = \arg \max_{f: \|f\|_{L^2(\mathcal{D} \times T)}=1} \text{Var} \left[ \langle \mathcal{X} - \mu, f \rangle_{L^2(\mathcal{D} \times T)} \right],$$

and subsequent PCs  $f_k$ , with  $k > 1$ , solve the same problem with the additional constraint  $\langle f_k, f_h \rangle_{L^2(\mathcal{D} \times T)} = 0$ , for  $h = 1, \dots, k-1$ , i.e.,  $f_k$  must be orthogonal to all the previous PCs.

Another characterization of the PCs goes under the name of best  $M$ -basis approximation property: for any positive integer  $M$ , the first  $M$  PCs satisfy

$$\{f_k\}_{k=1}^M = \arg \min_{\substack{\langle f_k, f_h \rangle_{L^2(\mathcal{D} \times T)} = \delta_{kh}, \\ h=1, \dots, k}} \mathbb{E} \left[ \int_{\mathcal{D} \times T} \left( \mathcal{X} - \mu - \sum_{r=1}^M \left( \int_{\mathcal{D} \times T} \mathcal{X} f_r \right) f_k \right)^2 \right] \quad (1)$$

where  $\delta_{kh}$  denotes the Kronecker delta, i.e.,  $\delta_{kh} = 1$  if and only if  $k = h$ , and  $\delta_{kh} = 0$  otherwise.

## 2.2 fPCA estimation problem for incomplete space-time data

Assume  $L$  realizations  $x_1, x_2, \dots, x_L$  of the random field  $\mathcal{X}$  were available, observed continuously over the spatio-temporal domain  $\mathcal{D} \times T$ , and without noise. We could then compute the sample covariance operator  $\hat{V}$  and obtain the first  $M$  PCs from its (numerical) spectral decomposition.

In real-world applications, however, we never observe realizations of the random field  $\mathcal{X}$  continuously over  $\mathcal{D} \times T$  and without noise, but only their noisy measurements, at some spatio-temporal locations. Specifically, let  $\mathbf{p}_1, \dots, \mathbf{p}_n$ , be  $n$  locations in the spatial domain  $\mathcal{D}$ , and  $t_1, \dots, t_m$ , be  $m$  time instants in  $T$ . Denote by  $x_{lij}$  the noisy measurement of the  $l$ -th statistical unit  $x_l$  at the spatio-temporal location  $(\mathbf{p}_i, t_j)$ . In the Functional Data Analysis community, a common approach to solve the fPCA problem, starting from the noisy and discrete measurements of the statistical units, consists in first obtaining smooth representatives of  $x_1, x_2, \dots, x_L$ , by appropriate smoothing procedures, and then computing the resulting sample covariance operator, with its spectral decomposition. However, such a pre-smoothing approach may fail in the context of missing data, especially in the presence of complex missing data patterns, as highlighted in [Palumbo et al. \(2023\)](#).

We here follow a different approach that starts from the characterization of the PCs given in Equation (1). Specifically, for  $l = 1, \dots, L$ , let  $\mathcal{O}_l$  be the set of all index pairs  $(i, j)$  where  $x_{lij}$  is not missing. Assume, for simplicity of exposition, that the data have already been centered around the mean, at each spatio-temporal location. Then, the sample version of the objective functional in Equation (1), for  $M = 1$ , is given by

$$\sum_{l=1}^L \sum_{(i,j) \in \mathcal{O}_l} [x_{lij} - s_l f(\mathbf{p}_i, t_j)]^2. \quad (2)$$

The estimation of the infinite-dimensional PC function  $f$ , starting from the discrete measurements  $x_{lij}$ , through minimization of (2), is though an ill-posed problem, unless a proper regularization is introduced. To this end, we add to the objective functional (2) a proper regularizing term, which seeks smoothness in the PC function  $f$ . In particular, following the approach presented by [Bernardi et al. \(2017\)](#) in the context

of spatio-temporal smoothing, we consider the space-time roughness penalty

$$\mathcal{P}_{\lambda_{\mathcal{D}}, \lambda_T}(f) = \lambda_{\mathcal{D}} \int_{\mathcal{D}} \int_{\mathcal{D}} (\Delta f)^2 + \lambda_T \int_{\mathcal{D}} \int_T \left( \frac{\partial^2 f}{\partial t^2} \right)^2, \quad (3)$$

where  $\Delta f = \sum_{h=1}^d \frac{\partial^2 f}{\partial p_h^2}$  denotes the Laplacian of  $f$ , while  $\lambda_{\mathcal{D}} > 0$  and  $\lambda_T > 0$  are two smoothness parameters, which control the roughness of the PC function in space and time. Therefore, we propose to estimate the first PC function  $f_1 : \mathcal{D} \times T \rightarrow \mathbb{R}$  and the corresponding PC scores vector  $\mathbf{s}_1 \in \mathbb{R}^L$  by solving the following minimization problem

$$\begin{aligned} \arg \min_{\mathbf{s} \in \mathbb{R}^L, f \in \mathbb{H}} \quad & \sum_{l=1}^L \sum_{(i,j) \in \mathcal{O}_l} [x_{lij} - s_l f(\mathbf{p}_i, t_j)]^2 + \mathbf{s}^\top \mathbf{s} \mathcal{P}_{\lambda_{\mathcal{D}}, \lambda_T}(f) \\ \text{Subject to} \quad & \|\mathbf{s}\|_2 = 1 \end{aligned} \quad (4)$$

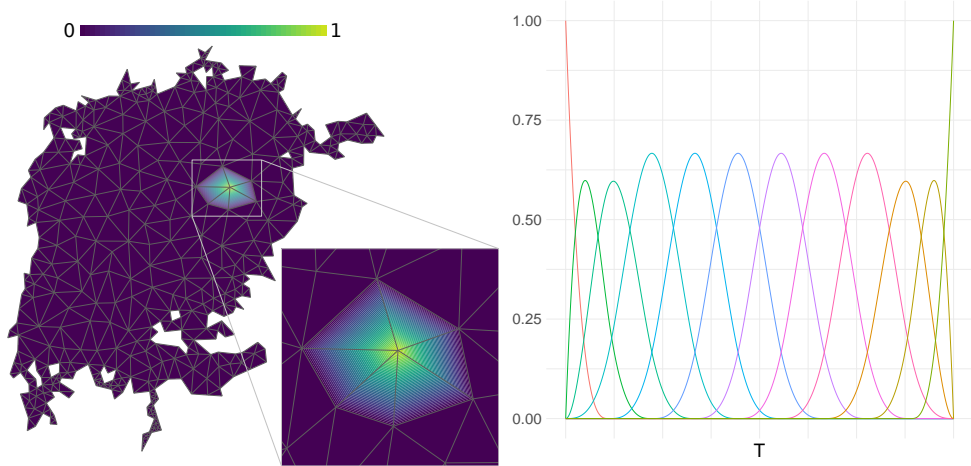
where  $\mathbb{H} \subset L^2(\mathcal{D} \times T)$  is a proper space of functions where the objective functional is well-posed (see, e.g., [Arnone, 2018](#)). As discussed in Section 4.1, the first term of the objective functional in Equation (4) corresponds to a weighted rank-one approximation of the data matrix, and encourages the PC function  $f$  to capture the strongest mode of variation in the observed data. The second term controls the smoothness of  $f$  in space and time. Moreover, the term  $\mathbf{s}^\top \mathbf{s}$  is justified by invariance considerations on the objective functional, similar to what was done in [Huang et al. \(2008\)](#), while the normalization constraint  $\|\mathbf{s}\|_2 = 1$  is set to make the representation unique. Subsequent PCs are estimated by solving the same estimation problem, but where missing entries have been imputed, as detailed in Section 4.2.

The estimation problem (4) presents various challenging aspects. First of all, it is an infinite-dimensional estimation problem, involving the infinite-dimensional unknown  $f$ , and it does not enjoy a closed form solution. This calls for an appropriate numerical discretization that will be presented in Section 3. Second, it is non-convex in  $(\mathbf{s}, f)$ . This requires the development of an appropriate iterative scheme, that will be detailed in Section 4. In this respect, the presence of missing data raises another complication. Indeed, the iterative approaches formerly considered for fPCA for fully observed data by, e.g., [Zou et al. \(2006\)](#); [Huang et al. \(2008\)](#); [Lila et al. \(2016\)](#); [Arnone et al. \(2023\)](#), and explored in [Stefanucci et al. \(2018\)](#) in the context of partially observed functional data, may instead be inadequate in the presence of complex missing data patterns, which requires the more complex iterative scheme proposed in Section 4.

### 3 Discretization of the infinite dimensional problem

We here present a numerical discretization of the functional in Equation (4) which allow us to consider spatial domains with complex shapes, such as water bodies with complicated shorelines or curved spatial regions with complex orography. This discretisation is based on finite elements in space and splines in time.





**Fig. 2:** Left: linear finite element basis function  $\psi(\mathbf{p})$  defined over a triangulation of the Lake Victoria. Right: cubic B-spline basis functions  $\{\phi_k(t)\}_k$  over a time interval  $T$ .

### 3.1 Spatio-temporal basis system

Let  $\mathcal{T}$  be a triangulation of the spatial domain of interest  $\mathcal{D}$ , i.e., a finite union of non-overlapping triangles approximating  $\mathcal{D}$  (Hjelle and Dæhlen, 2006). The left panel of Figure 2 shows an example. We define, on such triangulation, the space of finite element functions  $V_{\mathcal{T}}^r(\mathcal{D})$  as the space of continuous functions which are polynomials of degree  $r$ , once restricted to any triangle in  $\mathcal{T}$ ; see, e.g. Ciarlet (2002); Quarteroni (2017). For simplicity, in this work, we consider the case of linear finite elements ( $r = 1$ ). Let  $\xi_1, \xi_2, \dots, \xi_{N_{\mathcal{D}}}$  be the nodes of the triangulation that, for linear elements, coincide with the vertices of the triangles in  $\mathcal{T}$ . Therefore, we can introduce a basis system  $\psi_1(\mathbf{p}), \psi_2(\mathbf{p}), \dots, \psi_{N_{\mathcal{D}}}(\mathbf{p})$  for  $V_{\mathcal{T}}^r(\mathcal{D})$ , where each basis function is Lagrangian, that is,  $\psi_k(\xi_h) = 1$  if and only if  $k = h$ , and  $\psi_k(\xi_h) = 0$  otherwise. The left panel of Figure 2 shows a linear finite element basis, defined over a triangulation of Lake Victoria. Let  $\boldsymbol{\psi} = (\psi_1, \psi_2, \dots, \psi_{N_{\mathcal{D}}})^{\top}$  be the  $N_{\mathcal{D}}$ -vector of finite element basis. Any function  $v \in V_{\mathcal{T}}^r(\mathcal{D})$  can be written as a finite linear combination of these basis, i.e.,  $v(\mathbf{p}) = \mathbf{v}^{\top} \boldsymbol{\psi}(\mathbf{p})$ . An interesting property of Lagrangian elements is  $\mathbf{v} = (v(\xi_1), v(\xi_2), \dots, v(\xi_{N_{\mathcal{D}}}))^{\top}$ .

For the time dimension, we adopt a set of  $N_T$  cubic B-spline basis functions  $\phi_1(t), \phi_2(t), \dots, \phi_{N_T}(t)$ , defined over the time interval  $T$ ; see, e.g. De Boor (1978). The right panel of Figure 2 shows such a basis system.

Finally, we represent the spatio-temporal PC function  $f(\mathbf{p}, t)$  over  $\mathcal{D} \times T$  as

$$f(\mathbf{p}, t) = \sum_{k=1}^{N_{\mathcal{D}}} \sum_{h=1}^{N_T} c_{kh} \psi_k(\mathbf{p}) \phi_h(t). \quad (5)$$

where  $\{c_{kh}\}_{kh}$  are the coefficients of the expansion of  $f$  with respect to the considered spatio-temporal basis.

### 3.2 Discretization of the differential penalty

Define the  $n \times N_{\mathcal{D}}$  matrix  $\Psi = [\Psi]_{ik} = \psi_k(\mathbf{p}_i)$  of the evaluation of the  $N_{\mathcal{D}}$  finite element basis at the  $n$  spatial locations, and the  $N_{\mathcal{D}} \times N_{\mathcal{D}}$  matrices

$$R_0 = [R_0]_{kh} = \int_{\mathcal{D}} \psi_k \psi_h \quad R_1 = [R_1]_{kh} = \int_{\mathcal{D}} \nabla \psi_k^\top \nabla \psi_h.$$

Moreover, define the  $m \times N_T$  matrix  $\Phi = [\Phi]_{jk} = \phi_k(t_j)$  of the evaluation of the  $N_T$  spline basis functions at the  $m$  temporal locations, and the  $N_T \times N_T$  matrices

$$R_t = [R_t]_{kh} = \int_T \phi_k \phi_h \quad P_t = [P_t]_{kh} = \int_T \frac{d^2 \phi_k}{dt^2} \frac{d^2 \phi_h}{dt^2}.$$

Let  $\otimes$  be the Kronecker tensor product between matrices, and set  $\tilde{\Psi} = \Psi \otimes \Phi$ . Moreover, introduce the vector  $\mathbf{c} = (c_{11}, c_{12}, \dots, c_{N_{\mathcal{D}}N_T})^\top$  of the expansion coefficients of  $f$  in (5). Set  $P = \lambda_{\mathcal{D}}(R_1^\top R_0^{-1} R_1 \otimes R_t) + \lambda_T(P_t \otimes R_0)$ . Following the approach in Bernardi et al. (2017), we can discretize the differential penalty in Equation (3) by

$$\lambda_{\mathcal{D}} \int_T \int_{\mathcal{D}} (\Delta f)^2 + \lambda_T \int_{\mathcal{D}} \int_T \left( \frac{\partial^2 f}{\partial t^2} \right)^2 \approx \mathbf{c}^\top P \mathbf{c}. \quad (6)$$

## 4 Iterative solution of the estimation problem

In this section, we propose an iterative procedure, in the family of Majorize-Minimization (MM) algorithms (see, e.g., Lange, 2004; Lange and Zhou, 2022), which permits the efficient numerical solution of the estimation problem (4), and hence the identification of the principal component, and corresponding scores, from a set of partially observed space-time data.

### 4.1 Data loss

Before introducing the MM algorithm that solves Equation (4), we highlight that the first term in (4), i.e., the data loss term, can be interpreted as a weighted rank-one approximation of the data matrix. To this end, let  $X$  be the  $L \times nm$  data matrix, whose  $l$ -th row contains the noisy measurements of the  $l$ -th statistical units, at the  $nm$  spatio-temporal locations, i.e.,  $(x_{l11}, x_{l12}, \dots, x_{lnm})$ . Denote by  $W$  the binary matrix  $L \times nm$ , whose  $l$ -th row  $(w_{l11}, w_{l12}, \dots, w_{lnm})$  has  $w_{lij} = 1$  if and only if  $(i, j) \in \mathcal{O}_l$ , that is, when the datum  $x_{lij}$  is observed and  $w_{lij} = 0$  otherwise. Denote instead by  $W^C$  the binary matrix  $L \times nm$  with zeros indicating observed values and ones indicating missing observations. Finally, let  $\|\cdot\|_F$  be the Frobenious norm, and  $*$  be the Hadamard (or element-wise) product between matrices. Then, the data loss term in Equation (4) can be written as

$$\sum_{l=1}^L \sum_{(i,j) \in \mathcal{O}_l} [x_{lij} - s_l f(\mathbf{p}_i, t_j)]^2 = \|W * (X - \mathbf{s} \mathbf{f}_{nm}^\top)\|_F^2, \quad (7)$$

where  $\mathbf{s} = (s_1, s_2, \dots, s_L) \in \mathbb{R}^L$  is the scores vector, and  $\mathbf{f}_{nm} = (f(\mathbf{p}_1, t_1), f(\mathbf{p}_2, t_1), \dots, f(\mathbf{p}_n, t_m))^\top \in \mathbb{R}^{nm}$  is the vector of the evaluations of the PC function  $f$  at the  $nm$  spatio-temporal data locations, i.e.,  $\mathbf{f}_{nm} = \tilde{\Psi}\mathbf{c}$ . The formulation on the right-hand side of (7) is not uncommon in multivariate analysis, where the associated minimization problem is usually formalized as an approximation problem of the data matrix  $X$  by another matrix of lower rank (see, e.g., [Gabriel and Zamir, 1979](#)). For the unweighted case, i.e., in the special case where the data are fully observed and  $W$  is a matrix of all ones, the Eckart–Young–Mirsky theorem ([Eckart and Young, 1936](#)) guarantees that the best rank- $M$  matrix, with  $M \geq 1$ , approximating  $X$  is provided by its Singular Value Decomposition (SVD). For a general weight matrix  $W$ , even if binary, there is no analytic solution, and the problem is solved resorting to iterative methods, such as, for instance, Non-linear Iterative Partial Least Squares (NIPALS) ([Wold, 1966](#)) or Criss-Cross regression ([Gabriel and Zamir, 1979](#)).

## 4.2 Majorize-Minimization scheme

Thanks to Equations (6) and (7), we can rewrite the objective functional in (4) as

$$\|W * (X - \mathbf{s}\mathbf{f}_{nm}^\top)\|_F^2 + \mathbf{s}^\top \mathbf{s}\mathbf{c}^\top P\mathbf{c} = \|W * (X - \mathbf{s}(\tilde{\Psi}\mathbf{c})^\top)\|_F^2 + \text{Tr}[\mathbf{s}\mathbf{c}^\top P\mathbf{c}\mathbf{s}^\top] \quad (8)$$

where  $\text{Tr}[\cdot]$  denotes the matrix trace operator. Now, define the  $L \times N_{\mathcal{D}}N_T$  matrix  $U = \mathbf{s}\mathbf{c}^\top$ . Noting that  $\mathbf{s}\mathbf{f}_{nm}^\top = U\tilde{\Psi}^\top$ , we can further rewrite (8) as a functional of  $U$  as

$$h(U) = \|W * (X - U\tilde{\Psi}^\top)\|_F^2 + \text{Tr}[UPU^\top]. \quad (9)$$

At this point, we show that (9) can be minimized in  $U$  by an appropriate Majorize-Minimization (MM) scheme. An MM procedure seeks to minimize an objective function  $h : \mathbb{U} \rightarrow \mathbb{R}$ , where  $\mathbb{U}$  denotes some parameter space, by iterative minimization of a simpler function, whose optimization is more computationally tractable. In particular, starting from an initial guess  $U^0$ , an MM algorithm builds a sequence  $U^1, U^2, \dots, U^s$  in  $\mathbb{U}$ , which converges to a local optimum of the objective functional  $h(\cdot)$  (see, e.g. [Wu, 1983](#)). For each iteration index  $s \geq 0$ ,  $U^{s+1}$  is selected as the minimizer of a function  $g(U|U^s)$ , which is taken to be a majorization of  $h(\cdot)$  at  $U^s$ , that is, such that  $g(U|U^s) \geq h(U)$  for all  $U \neq U^s$ , with the additional condition  $h(U^s) = g(U^s|U^s)$ . Under this update rule, an MM procedure forces  $h(U)$  to decrease, as we have

$$h(U^{s+1}) \leq g(U^{s+1}|U^s) \leq g(U^s|U^s) = h(U^s).$$

The next result shows that a majorizing functional for the objective  $h(U)$  in (9) corresponds to the estimation functional of a fPCA on completely observed space-time data.

**Proposition 1.** *The functional in Equation (9) is majorized by*

$$g(U|U^s) = \|(W * X + W^C * U^s\tilde{\Psi}^\top) - U\tilde{\Psi}^\top\|_F^2 + \text{Tr}[UPU^\top] + \zeta \quad (10)$$

where  $\zeta \in \mathbb{R}$  is a constant not depending on  $U$ , and  $Y^s = W * X + W^C * U^s \tilde{\Psi}^\top$  is the data matrix obtained by imputing the missing observations in  $X$  with the reconstructed signal  $U^s$ , provided by the PCs estimated at the  $s$ -th iteration.

The proof is reported in Appendix A. According to Proposition 1, at each iteration of the MM procedure, we have to solve the following minimization problem.

$$\begin{aligned} \arg \min_{\mathbf{s} \in \mathbb{R}^L, \mathbf{c} \in \mathbb{R}^{N_D N_T}} \quad & \|Y^s - \mathbf{s}(\tilde{\Psi}\mathbf{c})^\top\|_F^2 + \text{Tr}[\mathbf{s}\mathbf{c}^\top P\mathbf{c}\mathbf{s}^\top] \\ \text{subject to} \quad & \|\mathbf{s}\|_2 = 1 \end{aligned} \quad (11)$$

This is an fPCA problem for completely observed spatio-temporal functional data, which extends, to the space-time setting, the models presented in Lila et al. (2016) and Arnone et al. (2023). Section 4.3 details the numerical algorithm to solve (11) and extract the first  $M$  functional principal components from the imputed data  $Y^s$ . Once the PCs of  $Y^s$  are estimated, the reconstructed signal  $U^s$  is updated, and the minimization problem (11) is repeated in  $Y^{s+1}$ , until convergence.

In the fPCA approaches for fully observed data, described in Lila et al. (2016); Arnone et al. (2023); Huang et al. (2008), the first  $M$  principal components are computed one at a time, solving problems similar to (11), on residualized data matrices, as detailed in Section 4.3, without any need for the MM algorithm proposed here. However, the missing data scenario here considered is much more challenging. Indeed, in this setting, we have to recursively apply the MM procedure, and repeat the estimation of all the first  $M$  PCs, at each iteration of the MM algorithm, from the imputed data  $Y^s$ , using as a starting point  $U^0$  the reconstructed signal  $U^s$ , obtained at convergence of the previous  $M - 1$  PCs. This recursive procedure permits to improve the quality of the  $M$ -th estimated PC, and the overall signal reconstruction, while preserving the same quality of estimation on the first  $M - 1$  PCs.

### 4.3 Minimization of the majorizing functional

We solve the estimation problem (11) by an iterative two-step algorithm, where we alternate the estimation of  $\mathbf{s}$  given  $\mathbf{c}$ , and the estimation of  $\mathbf{c}$  given  $\mathbf{s}$ . This iterative scheme is based on the following results.

**Proposition 2.** (*Estimation of  $\mathbf{s}$  given  $\mathbf{c}$* ). *Given  $\mathbf{c} \in \mathbb{R}^{N_D N_T}$ , there exists a unique estimator  $\hat{\mathbf{s}} \in \mathbb{R}^L$ , with  $\|\hat{\mathbf{s}}\|_2 = 1$ , which solves (11). Moreover,*

$$\hat{\mathbf{s}} = \frac{Y^s \tilde{\Psi} \mathbf{c}}{\|Y^s \tilde{\Psi} \mathbf{c}\|_2}. \quad (12)$$

**Proposition 3.** (*Estimation of  $\mathbf{c}$  given  $\mathbf{s}$* ). *Given  $\mathbf{s} \in \mathbb{R}^L$ , with  $\|\mathbf{s}\|_2 = 1$ , there exists a unique estimator  $\hat{\mathbf{c}} \in \mathbb{R}^{N_D N_T}$  which solves (11). Moreover,*

$$\hat{\mathbf{c}} = (\tilde{\Psi}^\top \tilde{\Psi} + P)^{-1} \tilde{\Psi}^\top Y^s \mathbf{s}.$$

The proofs of Propositions 2 and 3 are reported in Appendix B. Proposition 2 highlights that the problem of estimating the scoring vector  $\hat{\mathbf{s}}$  is equivalent to that of finding the scores, given the loadings, in multivariate PCA. Proposition 3 shows that

the problem of estimating  $\hat{\mathbf{c}}$ , given the scores, corresponds to the problem of estimating a smooth field, starting from noisy observations at the spatio-temporal data locations  $(\mathbf{p}_i, t_j)$ .

It is worth noting that the problem (11) is nonconvex in  $(\mathbf{s}, f)$ . However, Proposition 2 states the uniqueness of the minimizer  $\hat{\mathbf{s}}$ , given  $\mathbf{c}$ , while Proposition 3 guarantees the uniqueness of the minimizer  $\hat{\mathbf{c}}$ , given  $\mathbf{s}$ . This implies that the objective in Equation (11) is non-decreasing under the update rule of the proposed algorithm.

Subsequent principal components of the complete data matrix  $Y^s$  are computed by solving the estimation problem in Equation (11) on the residual matrix  $Y^s - \mathbf{s}\mathbf{f}_{nm}^\top$ .

#### 4.4 Selection of the optimal smoothing parameters

The presence of the pair of smoothing parameters,  $\lambda_{\mathcal{D}}$  and  $\lambda_T$ , in the penalty term, allows for a further degree of flexibility of the modeling, as we can select the optimal level of smoothing, in space and time, of the PC functions. The accurate selection of these smoothing parameters is crucial to achieve optimal results. Too high values of the smoothing parameters can lead to oversmoothed solutions, leaving meaningful patterns in the residuals. In contrast, a too low value causes the estimated PCs to also fit the noise.

We select the optimal pair of smoothing parameters by K-fold cross validation. Specifically, let  $\mathcal{O}$  be the set  $\{(i, j, l) : (i, j) \in \mathcal{O}_l, l = 1, \dots, L\}$ , i.e., the set of all index triplets for which  $x_{lij}$  is observed. We random permute the set  $\mathcal{O}$ , and partition it in  $K$  non-overlapping folds. For each  $k = 1, \dots, K$ , let  $\mathcal{O}^k$  be the  $k$ -th fold, and define the  $L \times nm$  binary matrix  $W^{-k}$  as the matrix having, on the  $l$ -th row,  $w_{lij} = 1$  if and only if  $(i, j, l) \in \mathcal{O} \setminus \mathcal{O}^k$ , and  $w_{lij} = 0$  otherwise. Therefore, we define the training set  $X^{-k}$  as the matrix  $W^{-k} * X$ . Similarly, letting  $W^k$  be the  $L \times nm$  binary matrix having, on the  $l$ -th row,  $w_{lij} = 1$  if and only if  $(i, j, l) \in \mathcal{O}^k$ , and  $w_{lij} = 0$  otherwise, we define the validation set  $X^k$  as  $W^k * X$ . Finally, for each pair of smoothing parameters, we calculate the scores matrix  $S^k = [\mathbf{s}_1^k, \dots, \mathbf{s}_M^k]$  and the loadings matrix  $F_{nm}^k = [(\mathbf{f}_1^k)_{nm}, \dots, (\mathbf{f}_M^k)_{nm}]$  on the training set  $X^{-k}$ , and we select the pair of parameters  $(\lambda_{\mathcal{D}}, \lambda_T)$  that minimizes the reconstruction error over the validation set  $X^k$ , averaged over the  $K$  folds:

$$\text{CV}(\lambda_{\mathcal{D}}, \lambda_T) = \frac{1}{K} \sum_{k=1}^K \left[ \frac{1}{|\mathcal{O}^k|} \|W^k * (X^k - S^k(F_{nm}^k)^\top)\|_F^2 \right]. \quad (13)$$

As commented by Hastie et al. (2009) in more classical settings, a too high value of the number of folds  $K$  might lead to an approximately unbiased estimate for the reconstruction error, having high variance. On the contrary, a too low value of  $K$  might lead to an estimated error with low variance, but high bias. To avoid these two opposite suboptimal solutions, in this work we set  $K = 10$ , following the general recommendation in Hastie et al. (2009).

## 4.5 Selection of the optimal number of principal components

Determining the appropriate number of principal components that characterize the data is a critical aspect of Principal Component Analysis, when used for dimensionality reduction. In this work, we select the number of principal components on the basis of the total explained variance, following a standard elbow approach. Specifically, we use the notion of adjusted total variance of the computed principal components, proposed by Zou et al. (2006), and detailed in Lila et al. (2016) for the modeling framework here considered.

## 5 Simulation studies

In this section, we assess the performances of the proposed fPCA approach for incomplete functional data, compared to other methods for PCA in presence of missing observations, and under different missing data settings.

### 5.1 Data generation

We consider the spatio-temporal domain  $\mathcal{D} \times T$ , with  $\mathcal{D} = [0, 1]^2$ , and  $T = [0, 1]$ . We consider 3 orthonormal cosinusoidal functions  $f_1(\mathbf{p}, t)$ ,  $f_2(\mathbf{p}, t)$ ,  $f_3(\mathbf{p}, t)$  on  $\mathcal{D} \times T$  of the form  $\cos(\alpha\pi p_1) \cos(\beta\pi p_2) \cos(\gamma\pi t)$ , for  $\mathbf{p} = (p_1, p_2)^\top \in \mathcal{D}$  and  $t \in T$ , where  $(\alpha, \beta, \gamma)$  are set to  $(1, 1, 2)$ ,  $(1, 3, 2)$  and  $(4, 2, 3)$  for  $f_1$ ,  $f_2$  and  $f_3$ , respectively. Based on these functions, which shall play the role of the principal components, we generate  $L = 50$  fields as

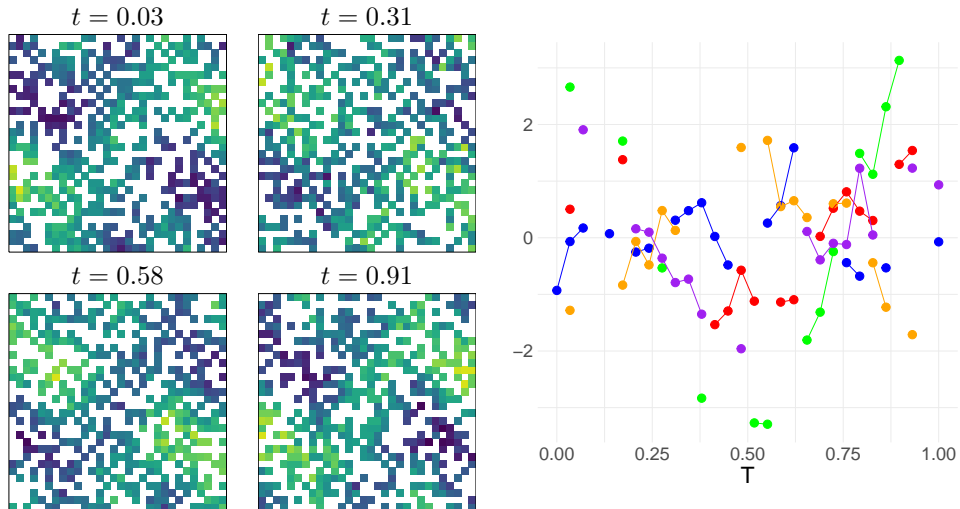
$$x_l(\mathbf{p}, t) = s_l^1 f_1(\mathbf{p}, t) + s_l^2 f_2(\mathbf{p}, t) + s_l^3 f_3(\mathbf{p}, t) \quad l = 1, \dots, L$$

with scores  $s_l^i \sim \mathcal{N}(0, \sigma_i^2)$  and  $\sigma_1 > \sigma_2 > \sigma_3$ , setting  $\sigma_1 = 0.4$ ,  $\sigma_2 = 0.3$  and  $\sigma_3 = 0.2$ . The  $L$  functions are evaluated on a regular grid of  $15 \times 15$  points over the spatial domain  $[0, 1]^2$  and at 15 equidistant time points over the temporal domain  $[0, 1]$ . Finally, data are obtained adding to each  $x_l(\mathbf{p}_i, t_j)$  uncorrelated Gaussian errors  $\epsilon_l(\mathbf{p}_i, t_j)$  with zero mean and standard deviation equal to 10% of the range of the data, obtaining the following noisy and discrete measurements of the  $L$  statistical units,

$$x_{lij} = s_l^1 f_1(\mathbf{p}_i, t_j) + s_l^2 f_2(\mathbf{p}_i, t_j) + s_l^3 f_3(\mathbf{p}_i, t_j) + \epsilon_l(\mathbf{p}_i, t_j),$$

for  $l = 1, \dots, L$ ,  $i = 1, \dots, n$ , and  $j = 1, \dots, m$ . This leads to the  $L \times nm$  complete data matrix  $X$ . To simulate the presence of missing observations, we consider two different missing data settings: an independent censoring in space and time, obtained as in the censoring scheme (a) of Arnone et al. (2022); a dependent censoring in space and time, in which data might be missing for large regions of the spatio-temporal domain, obtained as in the censoring scheme (d) of Arnone et al. (2022). Figures 3 shows the profile of a sparsely observed space-time signal, resulting from an independent censoring in space and time. Figure 4 instead depicts the same data subject to dependent censoring in space and time.

Data generation is repeated 100 times, sampling different scores and errors, and simulating different missing data profiles. The proportion of missing observation is set to 50% for both the considered missing data scenarios.



**Fig. 3:** Independent censoring in space and time. Left: visualization in space of one sampled signal, at 4 time instants. Right: visualization in time of the same signal, at 5 spatial locations. Data are sparsely observed over the spatio-temporal domain.

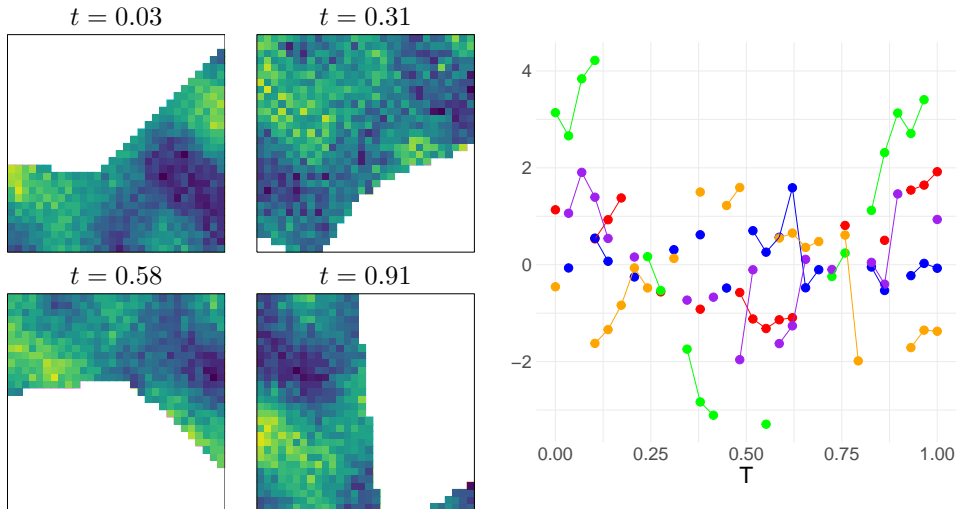
## 5.2 Competing methods

We compare the following competing methods for the Principal Component Analysis of incomplete data.

- PPCA: the Probabilistic PCA approach (Tipping and Bishop, 1999), as implemented by the R package `pcaMethods` (Stacklies et al., 2007), followed by a multivariate PCA.
- DINEOF: the Data INterpolating Empirical Orthogonal Function (DINEOF) method (Beckers and Rixen, 2003), as implemented in the R package `sinkr` (Taylor, 2022), followed by a multivariate PCA.
- IPPCA: the multivariate Iterative Penalized PCA approach (Josse and Husson, 2012), provided by the R package `missMDA` (Josse and Husson, 2016), followed by a multivariate PCA.
- fPCA: the proposed fPCA approach for incomplete functional data, implemented in the R package `fdaPDE` (Arnone et al., 2023). We use as nodes of the triangulation the  $15 \times 15$  grid of observations, and consider 15 equidistant nodes over the time domain  $[0, 1]$ . The optimal pair of smoothing parameters is selected using the K-fold cross validation approach detailed in Section 4.4.

## 5.3 Performance measures

Let  $(\hat{f}_{nm})_k$  and  $\hat{s}_k$  be the estimates of the  $k$ -th PC and corresponding scores. To compare the quality of the estimates produced by the considered methods, we consider the Root Mean Square Error (RMSE) of the PCs, evaluated at the spatio-temporal



**Fig. 4:** Dependent censoring in space and time. Left: visualization in space of one sampled signal, at 4 time instants. Right: visualization in time of the same signal, at 5 spatial locations. Data display large gaps in space and time.

data locations, i.e.,

$$\text{RMSE}((\hat{\mathbf{f}}_{nm})_k) = \sqrt{\frac{\sum_{i=1}^n \sum_{j=1}^m (f_k(\mathbf{p}_i, t_j) - \hat{f}_{kij})^2}{nm}} \quad (14)$$

where  $\hat{f}_{kij}$  denotes the evaluation of the estimated  $k$ -th PC  $(\hat{\mathbf{f}}_{nm})_k$ , at the spatio-temporal location  $(\mathbf{p}_i, t_j)$ . Moreover, we compute the RMSE on the scores as  $\frac{1}{\sqrt{L}} \|\mathbf{s}_k - \hat{\mathbf{s}}_k\|_2$ .

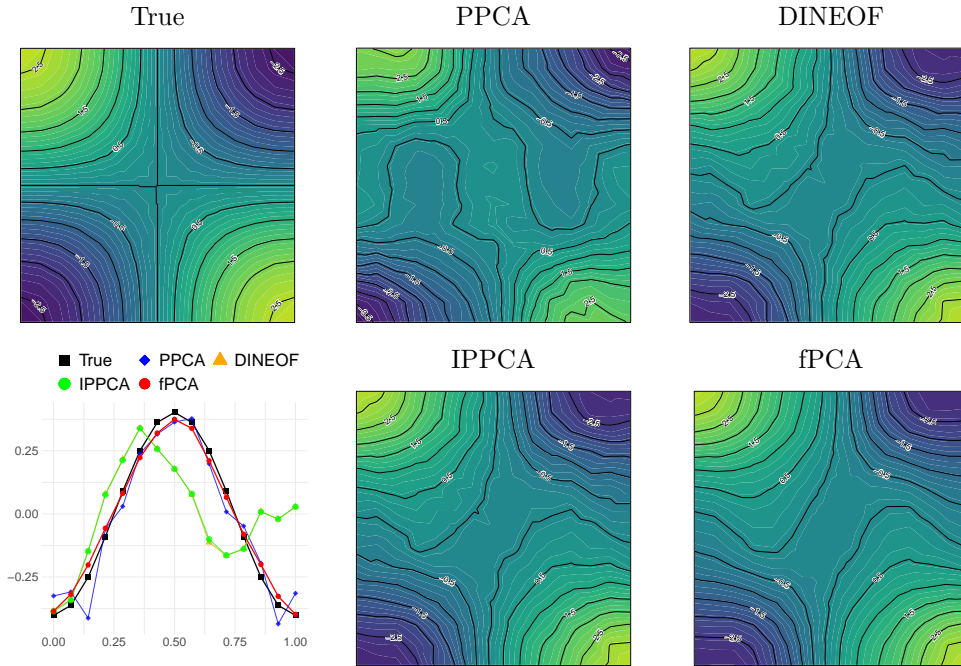
We also evaluate the performances of the methods in reconstructing the original data. Let  $\hat{S} = [\hat{\mathbf{s}}_1, \hat{\mathbf{s}}_2, \dots, \hat{\mathbf{s}}_M]$  and  $\hat{F}_{nm} = [(\hat{\mathbf{f}}_{nm})_1, (\hat{\mathbf{f}}_{nm})_2, \dots, (\hat{\mathbf{f}}_{nm})_M]$  be the computed scores and loadings matrices, and define the reconstructed signal as  $\hat{X} = \bar{X} + \hat{S}\hat{F}_{nm}^\top$ . We define the signal reconstruction error as  $\frac{1}{\sqrt{Lnm}} \|X - \hat{X}\|_F$ , where  $X$  is the matrix of fully observed data.

Finally, we assess the accuracy of the methods in reconstructing the space spanned by the principal components, considering the principal angle between the space spanned by the true principal components and the estimated ones, computed with the command `subspace` of the R package `pracma` (Borchers, 2022).

## 5.4 Results

We here compare the results obtained by the various methods using  $M = 3$  principal components. The data are indeed generated using 3 orthonormal functions, as detailed in Section 5.1. Moreover, all the considered methods correctly select  $M = 3$  components, following an elbow analysis of the total explained variance. Figure 5 reports some

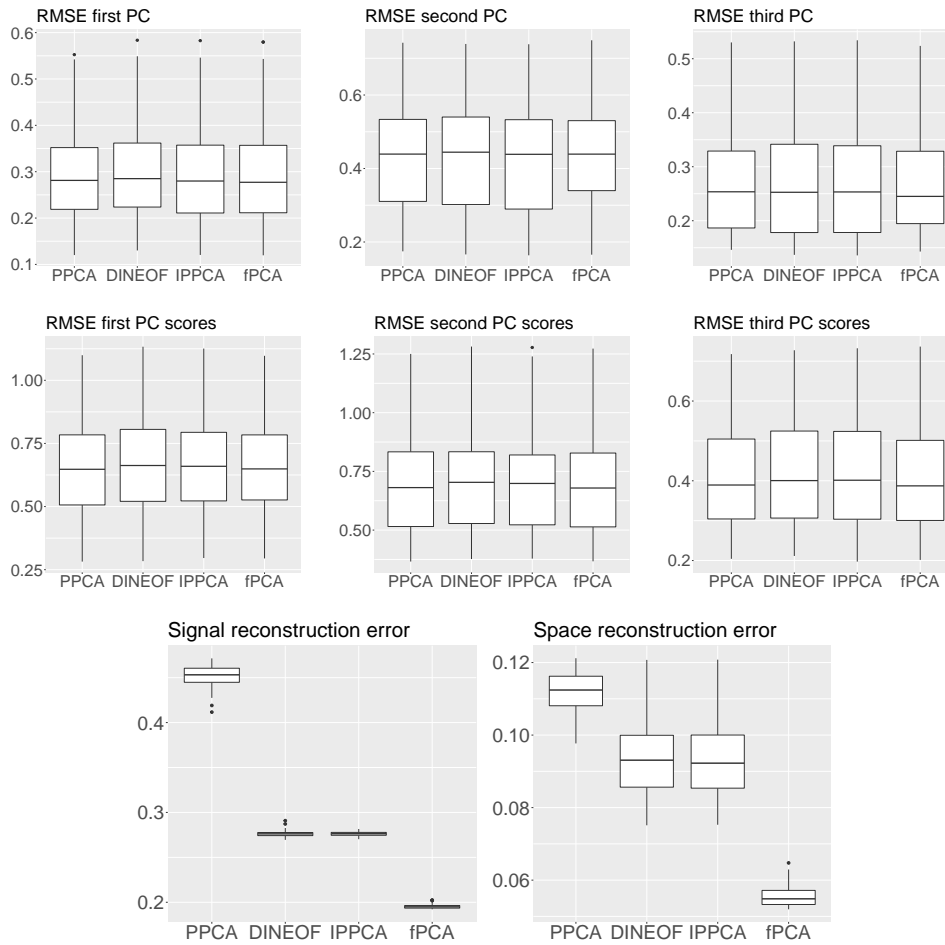




**Fig. 5:** Dependent censoring in space and time, first PC. Top-left: spatial profile of the true first PC, at a fixed time step. Center and right panels: spatial profile of the estimated first PC, at the same time step. Bottom-left: temporal profile of the true and estimated first PC, at a fixed spatial location.

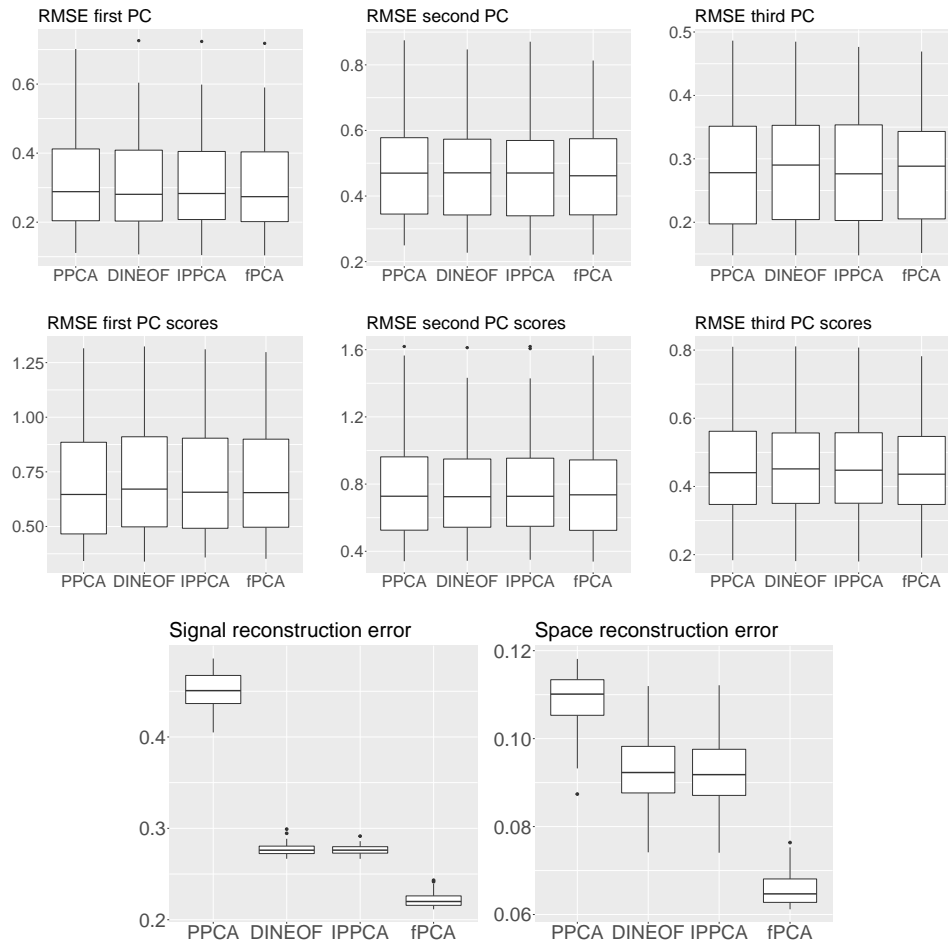
visualizations of the true first PC, and its estimates provided by PPCA, DINEOF, IPPCA and the proposed fPCA, in the most challenging scenario of dependent censoring in space and time, as illustrated in Figure 4. Figure C1 in Appendix C reports instead the estimates provided by the various methods in the independent censoring scenario. The panels in the center and right column display the spatial profile of the estimated first PC, at a fixed time step. We observe that all methodologies are able to capture the overall spatial profile of the true principal component, with fPCA producing the smoothest results. The bottom-left panel on the same figure shows instead the temporal profile of the true and estimated first PC, at a fixed spatial location. We note that the proposed fPCA is able to correctly follow the smooth behavior of the true PC function, while PPCA, DINEOF, and IPPCA, which do not account for any temporal correlation in the data, produce an irregular and less accurate estimate.

Figures 6 and 7 show the boxplots of the considered performance measures, for the estimates obtained by the various competing methods, across the 100 simulation repetitions. The boxplots show that the proposed fPCA performs equally or better than the competitors, along all performance measures. In particular, it produces the best results in terms of signal reconstruction as well as space reconstruction, both for sparsely observed data and for data presenting large gaps in space-time. In particular,



**Fig. 6:** Independent censoring in space and time: boxplots of the errors for Probabilistic PCA (PPCA), Data INterpolating Empirical Orthogonal Function (DINEOF), multivariate Iterative Penalized PCA (IPPCA) and the proposed functional PCA (fPCA). Top: RMSE on the three PCs. Center: RMSE on scores of the three PCs. Bottom: signal reconstruction error and space reconstruction error.

in the simulation considered with sparsely observed data, the signal reconstruction error of the proposed approach is, on average, 83% smaller than that of PPCA and 35% smaller than that of DINEOF and IPPCA. For the most challenging scenario of space-time dependent censoring, fPCA was able to reduce the average reconstruction error of 28% when compared to IPPCA and DINEOF, and to reduce the space reconstruction error of approximately 45%. We point out that we here reported the performances in the signal reconstruction over the whole space-time grid. However, the ordering of the



**Fig. 7:** Dependent censoring in space and time: boxplots of error for Probabilistic PCA (PPCA), Data INterpolating Empirical Orthogonal Function (DINEOF), multivariate Iterative Penalized PCA (IP-PCA) and the proposed functional PCA (fPCA). Left: signal reconstruction error. Right: space reconstruction error.

methods is the same when considering the RMSE computed over the missing locations only.

## 6 Application to Lake Victoria satellite data

Lakes possess a remarkable ability to stabilize short-term temperature fluctuations, while accentuating long-term variations. Notably, the Lake Surface Water Temperature (LSWT) is internationally acknowledged as an Essential Climate Variable (ECV), serving as a surrogate indicator for climate change and representing an important

indicator of lake hydrology and biogeochemistry. Furthermore, ecologists are also interested in understanding the spatial and temporal patterns of LSWT, to gain deeper insights into the dynamics of the environmental system.

With the aim of establishing a comprehensive long-term record of lakes' physical conditions, several satellite temperature data products have been developed. Among these, the ARC-Lake database (MacCallum and Merchant, 2011) offers an array of satellite-derived LSWT datasets, encompassing thousands of lakes worldwide. ARC-Lake is a project funded by the European Space Agency (ESA) and developed by the Earth Observation and Space Research Division at the University of Reading. The ESA's Earth Observing missions, including the series of (Advanced) Along Track Scanning Radiometers, (A)ATSRs, hold the potential to serve as highly accurate sources of information concerning LSWTs on a systematic global scale. However, due to technical challenges and specific meteorological conditions, such as the presence of atmospheric clouds, ice covering, and snow, the recorded data contain a significant number of missing observations. Accurate LSWT reconstructions are crucial in many applied fields, such as climate monitoring and numerical weather prediction. Concerning the latter, for instance, the increasing spatial resolution of weather simulations implies that it is no longer possible to neglect the presence of inland water bodies, nor it suffices to provide coarse approximations of their behavior.

In this section, we apply the proposed fPCA model to investigate the main spatio-temporal patterns in the surface temperature of Lake Victoria, starting from the noisy and incomplete observations shown in Figure 1, considering its complex shoreline and the spatio-temporal correlation in the data. Using the estimated smooth principal components, we can provide a reconstruction of the temperature field over the lake, which results to be more accurate than those provided by other signal reconstruction techniques.

## 6.1 Lake Surface Temperature data

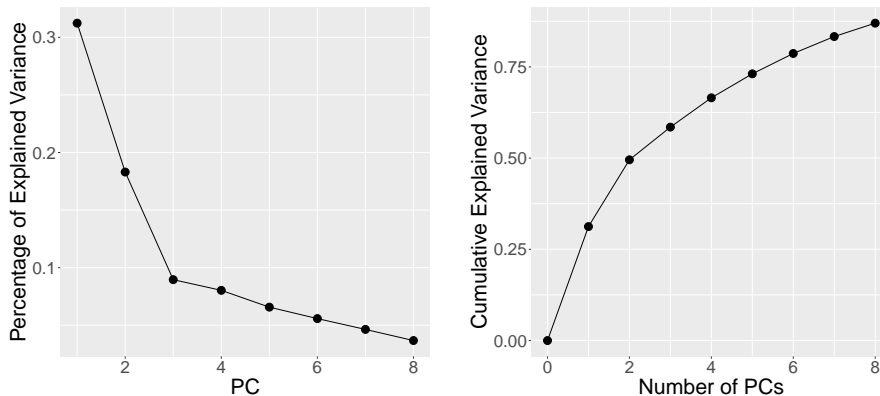
We consider the monthly averaged LSWT of Lake Victoria, in Central Africa. Specifically, each datum is an average of the surface temperature over a lake patch of  $0.05^\circ$  latitude by  $0.05^\circ$  longitude, and considering a month of observations (daytime). The resulting value is assigned to the center of the pixel, resulting in a grid of  $n = 2180$  equidistant spatial locations over the lake surface. The observation period goes from January 1996 to December 2011, for a total of 202 observations per spatial location (one per month). The considered data display a proportion of missing data equal to 45.2%, with non-trivial observation patterns, as highlighted in Figure 1.

In order to prepare the spatio-temporal water surface temperature data for the fPCA model, we split the data in  $L = 16$  statistical units, one per each calendar year of observations. Because, fixed a calendar year and a spatial location, we observe its temperature once per month, we have  $m = 12$  time instants per statistical unit. This leads us to an  $L \times nm$  data matrix  $X$ , in which each row corresponds to a calendar year. It should be pointed out that, in the given grid of observations, there are several spatio-temporal locations  $(\mathbf{p}_i, t_j)$  for which there is no observation available, for any statistical units (i.e., calendar year). These correspond to columns of the data matrix

$X$  with all missing entries. As discussed in Section 6.2, this setting is particularly challenging and invalidates some of the available state-of-the-art methods.

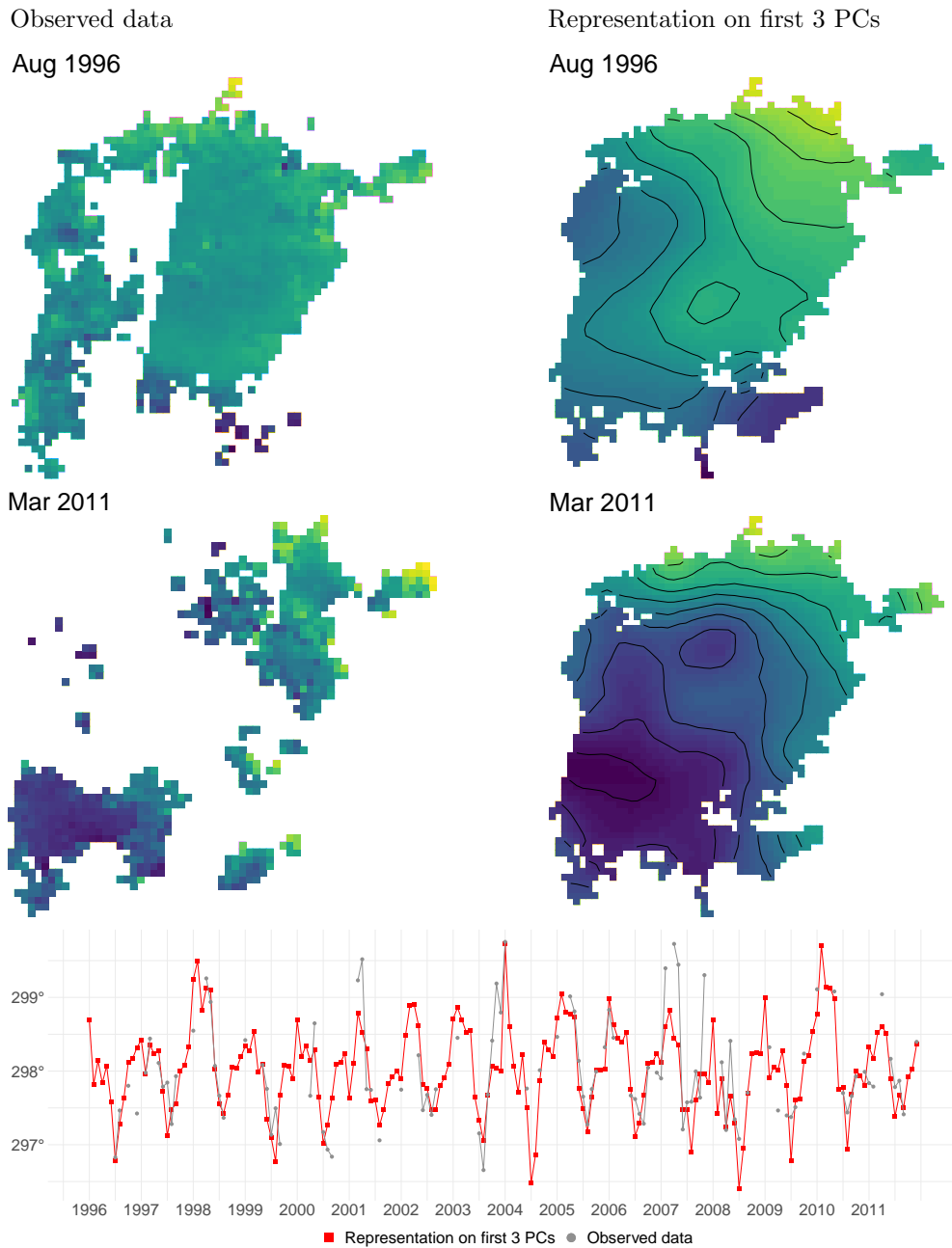
We point out that, due to the limited temporal range of the considered data, no long-term changes over the years are evident in this dataset. Consequently, we can treat each calendar year as an independent realization of the same spatio-temporal random field  $\mathcal{X}(\mathbf{p}, t)$ , so that the assumptions of the proposed fPCA are met. In presence of long-term trends, it is instead necessary to first detrend the data, to obtain signals exhibiting seasonal behaviors, where each recurrence of the yearly pattern can be regarded as an independent realization of the same spatio-temporal random field. The proposed fPCA can thus be applied to the detrended data.

## 6.2 Data analysis



**Fig. 8:** Left: empirical percentage of variance explained by the first 8 PCs. Right: cumulative percentage of explained variance captured by the first 8 PCs.

Before fPCA can be applied, the data matrix  $X$  must first be centered around its mean. If the noise in the measurements is low, the underlying spatio-temporal signal is smooth, and the data matrix  $X$  has no columns with all entries missing, then a simple point-wise mean is sufficient. However, in presence of very noisy data, with strong local variability, and for a data matrix  $X$  having columns with all entries missing, it is convenient to compute a smooth mean. This avoids removing much of the data variability in the mean, and bypasses the problem raised by spatio-temporal locations  $(\mathbf{p}_i, t_j)$  for which there is no observation available, for any statistical units. To compute a smooth estimate of the mean spatio-temporal temperature field, we use the smoothing method described in (Arnone et al., 2022) and implemented in the R package `fdaPDE` (Arnone et al., 2023). The estimate is obtained using the triangulation shown in the left panel of Figure 2, and forcing high values for the smoothing parameters. We then extract the first 8 PCs, from the centred data, using the proposed fPCA approach. To do so, we use again the triangulation shown in the left panel of Figure 2, but now



**Fig. 9:** Top left: observed LWST in August 1996. Top right: representation on the estimated first 3 PCs of LWST in August 1996. Center: same as top panels, for March 2011. Bottom: fPCA representation on the estimated first 3 PCs of LWST at the spatial location 3 displayed in the top-left map of Figure 1.

we select the optimal level of smoothness by K-fold cross validation, as detailed in Section 4.4, using  $K = 10$ . We then select  $M = 3$  PCs, since the plot of explained variance, given in the left panel of Figure 8, shows a clear elbow in correspondence of  $M = 3$ , with the first 3 PCs accounting for 60% of the total variation in the data, and the remaining PCs contributing less than 6% each. Figure 9 contrasts the observed LWST data, and the representation of LWST data on the first 3 estimated PCs. The top and central panels contrast the observed data (left) and the reconstructions on the first three estimated PCs (right), respectively, in August 1996 and March 2011. The bottom panel contrasts instead the observed temporal profile (in gray) and its reconstruction based on the first 3 estimated PCs, at the spatial location 3 displayed in the top left map of Figure 1. The accuracy in the reconstruction, using only 3 PCs, is remarkable.

It should be pointed out that, among the state-of-the-arts competitors presented in Section 5.2, only DINEOF is applicable for the considered data. Indeed, the R packages `pcaMethods` and `missMDA`, which implements respectively PPCA and IPPCA, do not consent to analyze data matrices  $X$  having at least one column without any observed entry. Subsequently, we contrast the proposed functional Principal Component Analysis (fPCA) with DINEOF, executed through the R package `sinkr` (Taylor, 2022). It is noteworthy that `sinkr` autonomously determines the most suitable number of Empirical Orthogonal Functions. Through a K-fold cross-validation with  $K = 10$  folds, the average signal reconstruction error for DINEOF is 0.38, whereas for the proposed fPCA, is 0.29, reflecting a notable reduction in error by an average of 25%.

## 7 Conclusions

We have presented an innovative method functional Principal Component Analysis of incomplete space-time data. The functional nature of the proposed method makes it able to borrow information from measurements observed at nearby spatio-temporal locations. This permits an accurate identification of the main variability patterns, in space and time, also when data are sparsely observed or in presence of large spatio-temporal gaps in the signals. The simulation studies in Section 5 demonstrate the comparative advantages of the proposed method over state-of-the-art PCA techniques for data with missing values, in both missing data scenarios. In particular, these studies highlight the superiority of the proposed fPCA in terms of signal reconstruction and space reconstruction. The ability to accurately reconstruct signals is a highly valuable feature in environmental applications, where signals are often affected by large observational gaps in space and time, as illustrated by the application to the study of LSWT of Lake Victoria.

**Acknowledgments.** L.M. Sangalli acknowledges funding by the European Union NextGenerationEU programme, within the framework of research project CoEnv - Complex Environmental Data and Modeling (PRIN2022 - CUP 2022E3RY23). A. Palumbo, L. Formaggia and L.M. Sangalli acknowledge the research project Dipartimento di Eccellenza 2023–2027, Dipartimento di Matematica, Politecnico di Milano, Italy, funded by the Ministero dell’Università e della Ricerca.

## References

- Gong, M., Miller, C., Scott, M.: Spatio-temporal modelling of remote-sensing lake surface water temperature data. 33rd International Workshop on Statistical Modelling (IWSM 2018), 106–111 (2018)
- Gong, M., Miller, C., Scott, M., O'Donnell, R., Simis, S., Groom, S., Tyler, A., Hunter, P., Spyarakos, E.: State space functional principal component analysis to identify spatiotemporal patterns in remote sensing lake water quality. *Stochastic Environmental Research and Risk Assessment* **35**, 2521–2536 (2021)
- MacCallum, S., Merchant, C.: Arc-lake v1.1 - per-lake, 1995-2009 [dataset] (2011) <https://doi.org/10.7488/ds/159>
- Gabriel, K.R., Zamir, S.: Lower rank approximation of matrices by least squares with any choice of weights. *Technometrics* **21**(4), 489–498 (1979)
- Kiers, H.A.L.: Weighted least squares fitting using ordinary least squares algorithms. *Psychometrika* **62**(2), 251–266 (1997)
- Beckers, J.M., Rixen, M.: Eof calculations and data filling from incomplete oceanographic datasets. *Journal of Atmospheric and Oceanic Technology* **20**(12), 1839–1856 (2003)
- Josse, J., Pagès, J., Husson, F.: Multiple imputation in principal component analysis. *Advances in Data Analysis and Classification* **5**(3), 231–246 (2011)
- Josse, J., Husson, F.: Handling missing values in exploratory multivariate data analysis methods. *Journal de la société française de statistique* **153**(2), 79–99 (2012)
- Hilborn, A., Costa, M.: Applications of dineof to satellite-derived chlorophyll-a from a productive coastal region. *Remote Sensing* **10**(9) (2018)
- Wang, Y., Liu, D.: Reconstruction of satellite chlorophyll-a data using a modified dineof method: a case study in the bohai and yellow seas, china. *International Journal of Remote Sensing* **35**(1), 204–217 (2014)
- Alvera-Azcárate, A., Barth, A., Beckers, J.-M., Weisberg, R.H.: Multivariate reconstruction of missing data in sea surface temperature, chlorophyll, and wind satellite fields. *Journal of Geophysical Research: Oceans* **112**(C3) (2007)
- Alvera-Azcárate, A., Barth, A., Rixen, M., Beckers, J.M.: Reconstruction of incomplete oceanographic data sets using empirical orthogonal functions: application to the adriatic sea surface temperature. *Ocean Modelling* **9**(4), 325–346 (2005)
- Ramsay, J.O., Silverman, B.W.: *Functional Data Analysis*, 2nd edn. Springer Series in Statistics, p. 426. Springer, New York (2005)



- Ferraty, F.: Nonparametric Functional Data Analysis. Springer, New York (2006)
- Kokoszka, P., Reimherr, M.: Introduction to Functional Data Analysis (1st Ed.). CRC Press, Boca Raton (2017)
- Li, Y., Guan, Y.: Functional principal component analysis of spatiotemporal point processes with applications in disease surveillance. *Journal of the American Statistical Association* **109**(507), 1205–1215 (2014)
- Liu, C., Ray, S., Hooker, G.: Functional principal component analysis of spatially correlated data. *Statistics and Computing* **27**(6), 1639–1654 (2017)
- Huang, J.Z., Shen, H., Buja, A.: Functional principal components analysis via penalized rank one approximation. *Electronic Journal of Statistics* **2** (2008)
- Lange, K.: MM Optimization Algorithms. Society for Industrial and Applied Mathematics (SIAM), Philadelphia (2016)
- Heiser, W.: Multidimensional scaling with least absolute residuals. *Proc. 1st Conf. Int. Fed. Classification Soc.* (1987)
- Heiser, W.J.: Correspondence analysis with least absolute residuals. *Computational Statistics & Data Analysis* **5**(4), 337–356 (1987)
- Lange, K., Zhou, H.: A legacy of em algorithms. *International Statistical Review* **90**(S1), 52–66 (2022)
- Wu, C.F.J.: On the Convergence Properties of the EM Algorithm. *The Annals of Statistics* **11**(1), 95–103 (1983)
- Lila, E., Aston, J.A.D., Sangalli, L.M.: Smooth Principal Component Analysis over two-dimensional manifolds with an application to neuroimaging. *The Annals of Applied Statistics* **10**(4), 1854–1879 (2016)
- Arnone, E., Negri, L., Panzica, F., Sangalli, L.M.: Analyzing data in complicated 3d domains: Smoothing, semiparametric regression, and functional principal component analysis. *Biometrics* (2023) <https://doi.org/10.1111/biom.13845>
- Sangalli, L.M.: Spatial regression with partial differential equation regularisation. *International Statistical Review* **89**(3), 505–531 (2021)
- Arnone, E., Clemente, A., Sangalli, L.M., Lila, E., Ramsay, J., Formaggia, L.: fdaPDE: Physics-Informed Spatial and Functional Data Analysis. (2023). R package version 1.1-8. <https://CRAN.R-project.org/package=fdaPDE>
- Riesz, F., Nagy, B.S.: Functional Analysis. Dover Books on Mathematics. Dover Publications, New York (2012)

- Palummo, A., Arnone, E., Formaggia, L., Sangalli, L.M.: Functional principal component analysis for space-time data. *Proceedings of the GRASPA 2023 Conference* (2023)
- Bernardi, M.S., Sangalli, L.M., Mazza, G., Ramsay, J.O.: A penalized regression model for spatial functional data with application to the analysis of the production of waste in venice province. *Stochastic Environmental Research and Risk Assessment* **31**(1), 23–38 (2017)
- Arnone, E.: Regression with pde penalization for modelling functional data with spatial and spatio-temporal dependence. PhD thesis, Politecnico di Milano (2018)
- Zou, H., Hastie, T., Tibshirani, R.: Sparse principal component analysis. *Journal of Computational and Graphical Statistics* **15**(2), 265–286 (2006)
- Stefanucci, M., Sangalli, L.M., Brutti, P.: PCA-based discrimination of partially observed functional data, with an application to AneuRisk65 data set. *Stat. Neerl.* **72**(3), 246–264 (2018)
- Hjelle, Ø., Dæhlen, M.: *Triangulations and Applications*. Springer, Berlin (2006)
- Ciarlet, P.G.: *The Finite Element Method for Elliptic Problems*. SIAM, Philadelphia (2002)
- Quarteroni, A.: *Numerical Models for Differential Problems* vol. 16. Springer, Cham (2017)
- De Boor, C.: *A Practical Guide to Splines* vol. 27. Springer, New York (1978)
- Lange, K.: The MM Algorithm, pp. 119–136. Springer, New York, NY (2004)
- Lange, K., Zhou, H.: A legacy of em algorithms. *International Statistical Review* **90**(S1), 52–66 (2022)
- Eckart, C., Young, G.: The approximation of one matrix by another of lower rank. *Psychometrika* **1**(3), 211–218 (1936)
- Wold, H.: Estimation of principal components and related models by iterative least squares. *Journal of Multivariate Analysis*, 391–420 (1966)
- Hastie, T., Tibshirani, R., Friedman, J.: *The Elements of Statistical Learning*, 2nd edn. Springer Series in Statistics, p. 745. Springer, New York (2009)
- Arnone, E., Sangalli, L.M., Vicini, A.: Smoothing spatio-temporal data with complex missing data patterns. *Statistical Modelling* (2022) <https://doi.org/10.1177/1471082X211057959>
- Tipping, M.E., Bishop, C.M.: Probabilistic principal component analysis. *Journal of*

the Royal Statistical Society. Series B (Statistical Methodology) **61**(3), 611–622 (1999)

Stacklies, W., Redestig, H., Scholz, M., Walther, D., Selbig, J.: pcamethods – a bio-conductor package providing pca methods for incomplete data. *Bioinformatics* **23**, 1164–1167 (2007)

Taylor, M.: Sinkr: Collection of Functions with Emphasis in Multivariate Data Analysis. (2022). R package version 0.7

Josse, J., Husson, F.: missmda: A package for handling missing values in multivariate data analysis. *Journal of Statistical Software* **70**(1), 1–31 (2016)

Borchers, H.W.: Pracma: Practical Numerical Math Functions. (2022). R package version 2.4.2. <https://CRAN.R-project.org/package=pracma>

Horn, R.A., Johnson, C.R.: Matrix Analysis, 2nd edn., p. 643. Cambridge University Press, Cambridge (2013)

## Appendix A Majorization of the estimation functional

The proof of Proposition 1 is based on the following lemma.

**Lemma 1.** *Let  $S$  an  $n \times n$  real symmetric matrix and  $\lambda_S$  its maximum eigenvalue, then*

$$\text{Tr}[\mathbf{e}^\top S \mathbf{e}] \leq \lambda_S \text{Tr}[\mathbf{e}^\top \mathbf{e}] \quad \forall \mathbf{e} \in \mathbb{R}^n. \quad (\text{A1})$$

*Proof:* the thesis follows directly from the fact that  $\lambda_S$  is the maximum of the Rayleigh quotient, i.e.  $\mathbf{e}^\top S \mathbf{e} \leq \lambda_S \mathbf{e}^\top \mathbf{e}$ , for any  $\mathbf{e} \in \mathbb{R}^n$  (see, e.g., [Horn and Johnson, 2013](#)). Applying the trace operator on both sides of the inequality yields the desired result.  $\square$

*Proof of Proposition 1:* We follow similar arguments as those described in ([Heiser, 1987b](#); [Kiers, 1997](#)). In particular, we provide a majorization of the functional  $h(U) = \|W * (X - U\Psi^\top)\|_F^2 + \text{Tr}[UPU^\top]$  by majorizing the data loss term  $\ell(U) = \|W * (X - U\Psi^\top)\|_F^2$ . In the following, for any  $n \times m$  matrix  $A$ , let  $\text{vec}(A)$  be the  $nm \times 1$  vector obtained by stacking the columns of  $A$  on top of one another. Define  $D_W = \text{Diag}(\text{vec}(W))$ , and rewrite the Hadamard product as a matrix product to obtain

$$\begin{aligned} \|W * (X - U\Psi^\top)\|_F^2 &= \|D_W(\text{vec}(X) - \text{vec}(U\Psi^\top))\|_2^2 = \\ &\text{Tr}[\text{vec}(X - U\Psi^\top)^\top D_W^2 \text{vec}(X - U\Psi^\top)] \end{aligned} \quad (\text{A2})$$

The last equality is obtained by recalling that, for any matrix  $A$ ,  $\text{Tr}[A^\top A] = \|A\|_F^2$ . Let  $U^s$  be the estimate produced by the MM algorithm at the  $s$ -th iteration, and set  $\mathbf{e} = \text{vec}((U - U^s)\Psi^\top)$ . We can sum and subtract  $U^s\Psi^\top$  to each operand of the  $\text{vec}(\cdot)$

operator in Equation (A2) to obtain

$$\begin{aligned} & \text{Tr}[\text{vec}(X - U\Psi^\top)^\top D_W^2 \text{vec}(X - U\Psi^\top)] = \\ & \text{Tr}[\text{vec}(X - (U^s + U - U^s)\Psi^\top)^\top D_W^2 \text{vec}(X - (U^s + U - U^s)\Psi^\top)] = \\ & \ell(U^s) + \text{Tr}[\mathbf{e}^\top D_W^2 \mathbf{e}] - 2\text{Tr}[\text{vec}(X - U^s\Psi^\top)^\top D_W^2 \mathbf{e}]. \end{aligned} \quad (\text{A3})$$

Using the bound in Equation (A1) with  $S = D_W^2$ , we can provide a majorization for the functional in Equation (A3). In particular, because  $D_W^2$  is binary and diagonal, its maximum eigenvalue  $\lambda_{D_W^2}$  equals 1, so that  $\text{Tr}[\mathbf{e}^\top D_W^2 \mathbf{e}] \leq \text{Tr}[\mathbf{e}^\top \mathbf{e}]$ . Therefore

$$h(U) \leq \ell(U^s) + \text{Tr}[\mathbf{e}^\top \mathbf{e}] - 2\text{Tr}[\text{vec}(X - U^s\Psi^\top)^\top D_W^2 \mathbf{e}] + \text{Tr}[UPU^\top].$$

Set  $g(U|U^s) = \ell(U^s) + \text{Tr}[\mathbf{e}^\top \mathbf{e}] - 2\text{Tr}[\text{vec}(X - U^s\Psi^\top)^\top D_W^2 \mathbf{e}] + \text{Tr}[UPU^\top]$  and observe that  $g(U|U^s)$  is a majorization for  $h(U)$  at  $U^s$ . Indeed  $g(U^s|U^s) = h(U^s)$ , as in this case  $\mathbf{e} = \mathbf{0}$ . Moreover, for the previous argument,  $g(U|U^s) \geq h(U)$ .

Finally, let  $\mathbf{z}$  be the vector  $D_W^2 \text{vec}(X - U^s\Psi^\top)$ , then

$$\begin{aligned} \ell(U^s) + \text{Tr}[\mathbf{e}^\top \mathbf{e}] - 2\text{Tr}[\mathbf{z}^\top \mathbf{e}] &= \ell(U^s) - \|\mathbf{z}\|_2^2 + \text{Tr}[\mathbf{e}^\top \mathbf{e} - 2\mathbf{z}^\top \mathbf{e} + \mathbf{z}^\top \mathbf{z}] = \\ \ell(U^s) - \|\mathbf{z}\|_2^2 + \text{Tr}[(\mathbf{z} - \mathbf{e})^\top (\mathbf{z} - \mathbf{e})] &= \ell(U^s) - \|\mathbf{z}\|_2^2 + \|\mathbf{z} - \mathbf{e}\|_2^2 = \\ \ell(U^s) - \|W * (X - U^s\Psi^\top)\|_F^2 + \|(W * X + W^C * U^s\Psi^\top) - U\Psi^\top\|_F^2. \end{aligned}$$

Setting  $\zeta = \ell(U^s) - \|W * (X - U^s\Psi^\top)\|_F^2$ , and observing that this quantity is not function of  $U$ , we obtain the thesis.  $\square$

## Appendix B Minimization of the majorizing functional

*Proof of Proposition 2:* We can rewrite the objective function in Equation (11) in vectorial form as  $\|Y^s - \mathbf{s}(\tilde{\Psi}\mathbf{c})^\top\|_F^2 + \mathbf{s}^\top \mathbf{s}(\mathbf{c}^\top P\mathbf{c})$ . Deriving this functional with respect to  $\mathbf{s}$  we get

$$\frac{\partial}{\partial \mathbf{s}} \left[ \|Y^s - \mathbf{s}(\tilde{\Psi}\mathbf{c})^\top\|_F^2 + \mathbf{s}^\top \mathbf{s}(\mathbf{c}^\top P\mathbf{c}) \right] = -2(Y^s - \mathbf{s}(\tilde{\Psi}\mathbf{c})^\top)\tilde{\Psi}\mathbf{c} + 2\mathbf{s}(\mathbf{c}^\top P\mathbf{c}). \quad (\text{B4})$$

We can then set to zero the last term in Equation (B4), thus obtaining

$$\begin{aligned} -2(Y^s - \mathbf{s}(\tilde{\Psi}\mathbf{c})^\top)\tilde{\Psi}\mathbf{c} + 2\mathbf{s}(\mathbf{c}^\top P\mathbf{c}) &= 0 \iff Y^s\tilde{\Psi}\mathbf{c} - \mathbf{s}\mathbf{c}^\top\tilde{\Psi}^\top\tilde{\Psi}\mathbf{c} = \mathbf{s}(\mathbf{c}^\top P\mathbf{c}) \\ &\iff \mathbf{s}(\|\tilde{\Psi}\mathbf{c}\|_2^2 + \mathbf{c}^\top P\mathbf{c}) = Y^s\tilde{\Psi}\mathbf{c}. \end{aligned}$$

Using the last equality, we obtain the following expression for the estimator  $\hat{\mathbf{s}}$

$$\hat{\mathbf{s}} = \frac{Y^s\tilde{\Psi}\mathbf{c}}{\|\tilde{\Psi}\mathbf{c}\|_2^2 + \mathbf{c}^\top P\mathbf{c}}. \quad (\text{B5})$$

On the other hand, for fixed  $\mathbf{c}$ , the objective in Equation (11) is convex in  $\mathbf{s}$  as  $\frac{\partial^2}{\partial \mathbf{s}^2} (\|Y^s - \mathbf{s}(\tilde{\Psi}\mathbf{c})^\top\|_F^2 + \mathbf{s}^\top \mathbf{s}(\mathbf{c}^\top P\mathbf{c})) = \|\tilde{\Psi}\mathbf{c}\|_2^2 + \mathbf{c}^\top P\mathbf{c} \geq 0$ . We conclude that the expression in Equation (B5), for fixed  $\mathbf{c}$ , is the unique minimizer of the functional in Equation (11). Finally, normalizing the expression in (B5) we recover the expression for  $\hat{\mathbf{s}}$  in Equation (12).  $\square$

The proof of Proposition 3 is based on the following lemma.

**Lemma 2.** *Minimizing (11) with respect to  $\mathbf{c} \in \mathbb{R}^{NM}$ , for fixed  $\mathbf{s} \in \mathbb{R}^L$  such that  $\|\mathbf{s}\|_2 = 1$ , is equivalent to minimize the following functional*

$$\mathbf{c}^\top (\tilde{\Psi}^\top \tilde{\Psi}) \mathbf{c} - 2(\tilde{\Psi}\mathbf{c})^\top Y^{s^\top} \mathbf{s} + \mathbf{c}^\top P\mathbf{c}. \quad (\text{B6})$$

*Proof:* consider the objective function in Equation (11), and develop the term  $\|Y^s - \mathbf{s}(\tilde{\Psi}\mathbf{c})^\top\|_F^2$  to obtain  $\|Y^s\|_F^2 + \|\mathbf{s}\|_2^2 \mathbf{c}^\top (\tilde{\Psi}^\top \tilde{\Psi}) \mathbf{c} - 2(\tilde{\Psi}\mathbf{c})^\top Y^{s^\top} \mathbf{s}$ . Thus, we can rewrite Equation (4) as

$$\|Y^s\|_F^2 + \|\mathbf{s}\|_2^2 \mathbf{c}^\top (\tilde{\Psi}^\top \tilde{\Psi}) \mathbf{c} - 2(\tilde{\Psi}\mathbf{c})^\top Y^{s^\top} \mathbf{s} + \|\mathbf{s}\|_2^2 (\mathbf{c}^\top P\mathbf{c}).$$

Since  $\|\mathbf{s}\|_2 = 1$  and  $\|Y^s\|_F^2$  do not depend on  $\mathbf{c}$ , we get the thesis.  $\square$

*Proof of Proposition (3):* let  $\mathbf{z} = Y^{s^\top} \mathbf{s}$ . By summing and subtracting  $\mathbf{z}^\top \mathbf{z}$ , we can rewrite the functional in Equation (B6) as

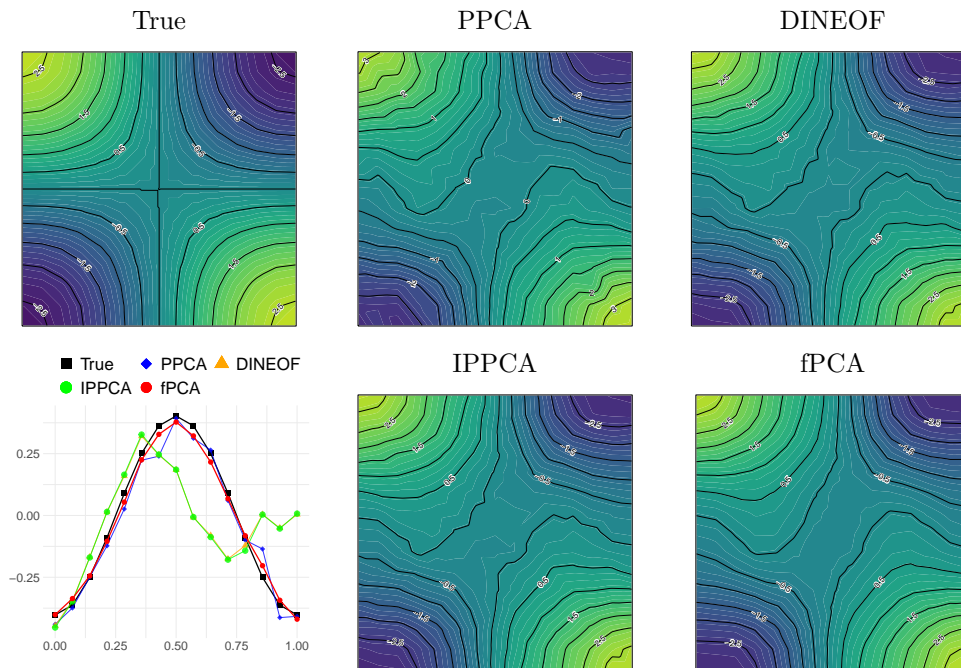
$$\|\mathbf{z} - \tilde{\Psi}\mathbf{c}\|_2^2 + \mathbf{c}^\top P\mathbf{c} - \mathbf{z}^\top \mathbf{z}. \quad (\text{B7})$$

Because  $\mathbf{z}^\top \mathbf{z}$  does not depend on  $\mathbf{c}$ , minimizing (B7) with respect to  $\mathbf{c}$ , for fixed  $\mathbf{s}$ , is equivalent to minimize  $\|\mathbf{z} - \tilde{\Psi}\mathbf{c}\|_2^2 + \mathbf{c}^\top P\mathbf{c}$ . On the other hand, as detailed in (Bernardi et al., 2017), estimating  $\mathbf{c}$  as the minimizer of this functional corresponds to fit a smooth spatio-temporal field to the vector of noisy observations  $\mathbf{z}$ , using the differential penalty in Equation (3). Moreover, the solution  $\hat{\mathbf{f}}$  to this estimation problem is equal to  $(\tilde{\Psi}^\top \tilde{\Psi} + P)^{-1} \tilde{\Psi}^\top \mathbf{z}$ . Finally, thanks to Proposition 2.1 in Arnone (2018), we have the existence and uniqueness of the estimator.  $\square$

## Appendix C Independent censoring simulation

In this section we report the plots of the first PC function as estimated by the proposed fPCA and the competing methods discussed in Section 5.2, under the independent space-time censoring illustrated in Figure 3. The panels in the center and right column display the spatial profile of the estimated PC, at a fixed time step. In analogy to what discussed in Section 5.4 of the paper, we observe that all multivariate methodologies are not able to produce regular estimates, even when data are sparsely observed over the spatio-temporal domain. The proposed fPCA approach, instead, is able to estimate a smooth spatial field, thanks to the regularisation induced by the differential penalty. The bottom-left panel of the same figure report the estimated temporal profile of the PC function at a fixed spatial locations and for all the considered methods. We point out that only fPCA, which properly regularise the estimated field in time, is able to

capture the smooth behavior of the true PC, while other methodologies produce less accurate and regular estimates.



**Fig. C1:** Independent censoring in space and time, first PC. Top-left: spatial profile of the true first PC, at a fixed time step. Center and right panels: spatial profile of the estimated first PC, at the same time step. Bottom-left: temporal profile of the true and estimated first PC, at a fixed spatial location. Plots are produced using the same spatial location and time instant considered for producing Figure 5.

## MOX Technical Reports, last issues

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

- 28/2024** Magri, M.; Riccobelli, D.  
*Modelling of initially stressed solids: structure of the energy density in the incompressible limit*
- Magri, M.; Riccobelli, D.  
*Modelling of initially stressed solids structure of the energy density in the incompressible limit*
- 27/2024** Antonietti, P.F.; Beirao da Veiga, L.; Botti, M.; Vacca, G.; Verani, M.  
*A Virtual Element method for non-Newtonian fluid flows*
- 26/2024** Bucelli, M.; Regazzoni, F.; Dede', L.; Quarteroni, A.  
*Robust radial basis function interpolation based on geodesic distance for the numerical coupling of multiphysics problems*
- 25/2024** Enrico Ballini e Luca Formaggia e Alessio Fumagalli e Anna Scotti e Paolo Zunino  
*Application of Deep Learning Reduced-Order Modeling for Single-Phase Flow in Faulted Porous Media*
- 22/2024** Gatti, F.; de Falco, C.; Fois, M.; Formaggia, L.  
*A scalable well-balanced numerical scheme for a depth-integrated lava flow model*
- 21/2024** Caldana, M.; Antonietti P. F.; Dede' L.  
*Discovering Artificial Viscosity Models for Discontinuous Galerkin Approximation of Conservation Laws using Physics-Informed Machine Learning*
- 19/2024** Torzoni, M.; Manzoni, A.; Mariani, S.  
*A multi-fidelity surrogate model for structural health monitoring exploiting model order reduction and artificial neural networks*
- 20/2024** Torzoni, M.; Manzoni, A.; Mariani, S.  
*Structural health monitoring of civil structures: A diagnostic framework powered by deep metric learning*
- 17/2024** Fois, M.; de Falco, C.; Formaggia, L.  
*A semi-conservative depth-averaged Material Point Method for fast flow-like landslides and mudflows*