Prediction in function-on-function linear model with partially observed functional covariate and response

Christophe Crambes, Chayma Daayeb, Ali Gannoun, Yousri Henchiri

Prediction in function-on-function linear model with partially observed functional covariate and response

Christophe Crambes[†] and Chayma Daayeb[†,‡] and Ali Gannoun[†] and Yousri Henchiri[‡,♮,*]

July 11, 2023

Abstract

In this work, we are interested in a function-on-function linear model in which the response and the covariate are partially observed curves. First, we reconstruct the missing part of the covariate using the observed parts. Then, we consider two strategies for dealing with the missing part of the response. The first one consists in a reconstruction in the same way as for the covariate. The second one uses regression imputation. Once the dataset is reconstructed, we estimate the slope function and give the mean square prediction error for a new observation of the covariate. Both methods are compared from a theoretical and a practical point of view.

Keywords. Functional linear model, Missing parts, Imputation, Partially observed functional data, Functional Principal Components, Reconstruction operator.

1 Introduction

Functional data analysis (FDA) is becoming progressively more and more important in Statistic (Bosq, 2000; Ramsay and Silverman, 2005; Ferraty and Vieu, 2006; Hsing and Eubank, 2015; Kokoszka and Reimherr, 2018). The functional linear model is specially a very popular model in both theoretical and applied research such as climatology, meteorology, economy, image analysis and many other fields. There is a large amount of work done on the functional linear model. In the case where the response is a real variable, the model has been widely studied (e.g. Cardot et al., 2003; Cai and Hall, 2006; Li and Hsing, 2007; Hall and Horowitz, 2007; Crambes et al., 2009; Comte and Johannes, 2012; Cai and Yuan, 2012; Brunel et al.,

*Corresponding author. E-mail: yousri.henchiri@umontpellier.fr. Contributing authors: christophe.crambes@umontpellier.fr, chayma.daayeb@umontpellier.fr, ali.gannoun@umontpellier.fr.
[†]Institut Montpelliérain Alexander Grothendieck (IMAG), Université de Montpellier, France. [‡]Université de Tunis El Manar, Laboratoire de Modélisation Mathématique et Numérique dans les Sciences de l'Ingénieur (LAMSIN), Tunis, Tunisie. [♮]Université de la Manouba, Institut Supérieur des Arts Multimédia de la Manouba (ISAMM), Tunisie.

2016). In contrast, fewer researchers have tackled the problem of function-on-function linear models where the covariate and the response are both functional (e.g. Ramsay and Silverman, 2005; Yao et al., 2005; Prchal and Sarda., 2007; Aguilera et al., 2008; Lian, 2011; Ferraty et al., 2012; Crambes and Mas, 2013; Crambes et al., 2016; Luo and Qi, 2017; Benatia et al., 2017; Imaizumi and Kato, 2018; Sun et al., 2018).

In recent years, applications producing partially observed functional data have emerged. Sometimes each individual trajectory is collected only over individual specific subintervals, densely or sparsely, within the whole domain of interest. Several recent works have begun addressing the estimation of covariance functions for short functional segments observed at sparse and irregular grid points, called *functional snippets* (Lin and Wang, 2022; Lin et al., 2021) or for *fragmented functional data* observed on small subintervals (Delaigle et al., 2020). For densely observed partial data, existing studies have focused on estimating the unobserved part of curves (Kneip and Liebl, 2020; Kraus and Stefanucci, 2020), prediction (Goldberg et al., 2014), classification (Kraus and Stefanucci, 2018; Park and Simpson, 2019), functional regression (Gellar et al., 2014), and inferences (Kraus, 2019; Park et al., 2022).

To our knowledge, few articles investigate the theoretical properties of the slope operator or kernel estimators in the framework of partially observed data. In the framework of a real response, we can notice the works from Crambes and Henchiri (2019) and Crambes et al. (2023). The objective of this paper is to study the prediction problem when the covariate and the response are partially observed in the function-on-function linear regression setting, for which we propose two methods. The paper is organized as follows. We present the model with fully observed functional data in section 2. In section 3, we study the case of partially observed covariate and response. We reconstruct the missing parts of the functional explanatory variable and the functional response using the work Kneip and Liebl (2020), and we get theoretical results for the prediction error rate. In section 4, we propose an alternative method which consists in imputing the functional response after having reconstructed the missing parts of the functional covariate as in the previous section. We also obtain a convergence rate for the mean square prediction error. In section 5, we conduct a numerical study over simulated data in order to compare the methods in practice. Section 6 is devoted to a real dataset application. Finally, all the proofs are postponed to section 7.

2 The centered function-on-function linear model

2.1 Functional principal components regression

Let $f : \mathscr{T} \to \mathbb{R}$ and $g : \mathscr{T} \times \mathscr{S} \to \mathbb{R}$ be two square integrable functions defined in the Hilbert space $\mathbb{L}^2(\mathscr{T})$ $\left(\text{resp.}\, \mathbb{L}^2(\mathscr{T} \times \mathscr{S})\right)$ i.e. the space of square integrable functions on the interval $\mathscr{T} \subseteq \mathbb{R}$ $\left(\text{resp. } \mathscr{S} \subseteq \mathbb{R}\right)$. For all $t \in \mathscr{T}$ $\left(\text{resp. } s \in \mathscr{S}\right)$, we set $\|f\| = \left(\int_{\mathscr{T}} f^2(t)\mathrm{d}t\right)^{1/2}$ and $\|g\| = \left(\int_{\mathscr{S}}\int_{\mathscr{T}} g^2(t,s)\mathrm{d}t\mathrm{d}s\right)^{1/2}$.

We use the following notation for the tensor product, defining $u \otimes v : \mathbb{L}^2(\mathscr{T}) \to \mathbb{L}^2(\mathscr{T})$, by $u \otimes v = \langle u, . \rangle v$, for any functions $u, v \in \mathbb{L}^2(\mathscr{T})$. We define the Hilbert-Schmidt integral operator $G : \mathbb{L}^2(\mathscr{T} \times \mathscr{S}) \to \mathbb{L}^2(\mathscr{T})$ by $(G \cdot u)(s) = \int_{\mathscr{T}} u(t) g(t,s) \mathrm{d}t$ for all g belonging to $\mathbb{L}^2(\mathscr{S} \times \mathscr{T})$ and $\int_{\mathscr{S}} \int_{\mathscr{T}} |g(t,s)|^2 \mathrm{d}t \mathrm{d}s < \infty$. The function g is the Hilbert-Schmidt kernel corresponding to the operator G. Moreover, the operator G is continuous (i.e bounded) and compact.

We assume here that $\mathscr{X}$ takes values in the space $\mathbb{L}^2(\mathscr{T})$ and $\mathscr{Y}$ in $\mathbb{L}^2(\mathscr{S})$, $\mathscr{X} \triangleq \{\mathscr{X}(t) \mid t \in \mathscr{T}\}$ is the predictor variable and $\mathscr{Y} \triangleq \{\mathscr{Y}(s) \mid s \in \mathscr{S}\}$ the response variable. We observe a sample $\{(\mathscr{X}_i, \mathscr{Y}_i)\}_{i=1}^n$, of identically distributed and independent copies $(\mathscr{X}, \mathscr{Y})$, with $\mathbb{E}(\|\mathscr{X}\|^2) < \infty$ and $\mathbb{E}(\|\mathscr{Y}\|^2) < \infty$. In practice, functional responses $\mathscr{Y}_i(s)$ and functional covariates $\mathscr{X}_i(t)$ are observed on grid points $\boldsymbol{s}_i = (s_{i1}, \ldots, s_{iq_i})$ and $\boldsymbol{t}_i = (t_{i1}, \ldots, t_{ip_i})$. For the sake of simplicity, we assume those to be identical vectors $\boldsymbol{s}$, $\boldsymbol{t}$ with lengths q and p, respectively, for each observation i.

We consider a centered function-on-function linear regression model (FFLRM) characterizing a linear relationship between the functional response and the functional predictor

$$\big[\mathscr{Y} - \mathbb{E}\big[\mathscr{Y}\big]\big](s) = \int_{\mathscr{T}} \big[\mathscr{X} - \mathbb{E}\big[\mathscr{X}\big]\big](t) \theta(t,s) \mathrm{d}t + \epsilon(s),$$

where the bivariate functional coefficient $(t,s) \mapsto \theta(t,s)$ is assumed to be in $\mathbb{L}^2(\mathscr{T} \times \mathscr{S})$, that is, $\int_{\mathscr{S}} \int_{\mathscr{T}} |\theta(t,s)|^2 \mathrm{d}t \mathrm{d}s < \infty$. Natural estimators of $\mathbb{E}\big[\mathscr{X}\big]$ and $\mathbb{E}\big[\mathscr{Y}\big]$ are the empirical means, $\overline{\mathscr{X}} \triangleq 1/n \sum_{i=1}^n \mathscr{X}_i$ and $\overline{\mathscr{Y}} \triangleq 1/n \sum_{i=1}^n \mathscr{Y}_i$. The centered random error function ϵ is assumed to be independent of $\mathscr{X}$, and $\mathbb{E}\|\epsilon\|^2 < \infty$.
We define the elements of the centered model as follows, $X \triangleq \mathscr{X} - \mathbb{E}\big[\mathscr{X}\big]$ and $Y \triangleq \mathscr{Y} - \mathbb{E}\big[\mathscr{Y}\big]$, and the centered FFLRM writes

$$Y(s) = \int_{\mathscr{T}} X(t) \theta(t,s) \mathrm{d}t + \epsilon(s). \tag{2.1}$$

Denote

$$\mathrm{C}_X(t_1, t_2) = \mathbb{E}\Big(X(t_1) X(t_2)\Big), \quad t_1, t_2 \in \mathscr{T},$$

and

$$\mathrm{C}_Y(s_1, s_2) = \mathbb{E}\Big(Y(s_1) Y(s_2)\Big), \quad s_1, s_2 \in \mathscr{S},$$

as the covariance functions of X and Y, respectively. As $\mathrm{C}_X \in \mathbb{L}^2(\mathscr{T}^2)$ and $\mathrm{C}_Y \in \mathbb{L}^2(\mathscr{S}^2)$ are a positive self-adjoint operators, then, according to Mercer's theorem (Hsing and Eubank, 2015, Theorem 4.6.5), C_X and C_Y admit spectral expansions in $\mathbb{L}^2(\mathscr{T}^2)$ and $\mathbb{L}^2(\mathscr{S}^2)$ respectively

$$\mathrm{C}_X(t_1, t_2) = \sum_{k=1}^{+\infty} \lambda_k \phi_k(t_1) \phi_k(t_2) \quad \text{and} \quad \mathrm{C}_Y(s_1, s_2) = \sum_{j=1}^{+\infty} \mu_j \psi_j(s_1) \psi_j(s_2),$$

where $\lambda_1 > \lambda_2 > \ldots > 0$ and $\mu_1 > \mu_2 > \ldots > 0$ are the eigenvalue sequences of the covariance functions C_X and C_Y, respectively, while $\{\phi_k\}_{k\geqslant 1}$ and $\{\psi_j\}_{j\geqslant 1}$ are the corresponding orthonormal bases of eigenfunctions in $\mathbb{L}^2(\mathscr{T})$ and $\mathbb{L}^2(\mathscr{S})$.

The Karhunen–Loève (KL) expansions of the curves X and Y (Hsing and Eubank, 2015, Theorem 7.3.5) in $\mathbb{L}^2(\mathscr{T})$ and $\mathbb{L}^2(\mathscr{S})$ are respectively

$$X(t) = \sum_{k=1}^{+\infty} \xi_k \phi_k(t) \quad \text{and} \quad Y(s) = \sum_{j=1}^{+\infty} \beta_j \psi_j(s),$$

for all $t \in \mathscr{T}$ and $s \in \mathscr{S}$, where $\xi_k = \int_{\mathscr{T}} X(t)\phi_k(t)\mathrm{d}t$ and $\beta_j = \int_{\mathscr{S}} Y(s)\psi_j(s)\mathrm{d}s$ are uncorrelated random variables with zero mean and variances $\mathbb{E}(\xi_k^2) = \lambda_k$ and $\mathbb{E}(\beta_j^2) = \mu_j$ for all $k, j \geqslant 1$. These coefficients ξ_k and β_j are called functional principal components scores. By Parseval's identity and Fubini's theorem, we have

$$\sum_{k=1}^{+\infty} \mathbb{E}(\xi_k^2) < \infty \quad \text{and} \quad \sum_{j=1}^{+\infty} \mathbb{E}(\beta_j^2) < \infty.$$

Ramsay and Silverman (2005, Chapter 16, Section 1.1), Park and Qian (2012), Crambes and Mas (2013) and Imaizumi and Kato (2018) suggest that $\theta(t, s)$ can be expressed in terms of a double basis expansion

$$\theta(t, s) = \sum_{k=1}^{+\infty} \sum_{j=1}^{+\infty} \theta_{k,j} \phi_k(t) \psi_j(s),$$

where $\theta_{k,j} = \int_{\mathscr{S}} \int_{\mathscr{T}} \phi_k(t)\theta(t, s)\psi_j(s)\mathrm{d}t\mathrm{d}s$.
Noticing that

$$\int_{\mathscr{T}} \theta(t, s)X(t)\mathrm{d}t = \sum_{k=1}^{+\infty} \sum_{j=1}^{+\infty} \theta_{k,j}\xi_k\psi_j(s),$$

and

$$\mathbb{E}\Big(\xi_k Y(s)\Big) = \lambda_k \sum_{j=1}^{+\infty} \theta_{k,j}\psi_j(s),$$

we obtain the following characterization of the bivariate functional coefficient $\theta(t, s)$ as

$$\theta(t, s) = \sum_{k=1}^{+\infty} \frac{\mathbb{E}\Big(\xi_k Y(s)\Big)}{\lambda_k} \phi_k(t).$$

This characterization leads to a method for estimating $\theta(t, s)$. For example, Park and Qian (2012) and Crambes and Mas (2013) use the following truncation estimator with $k_n \to \infty$ as $n \to \infty$,

$$\widehat{\theta}(t, s) = \frac{1}{n} \sum_{i=1}^{n} \sum_{k=1}^{k_n} \frac{\widehat{\xi}_{ik} Y_i(s)}{\widehat{\lambda}_k} \widehat{\phi}_k(t), \tag{2.2}$$

where $\widehat{\lambda}_k$ and $\widehat{\phi}_k$ are the estimators of the eigenvalues λ_k and the eigenfunctions ϕ_k and the estimates of the FPC scores are $\widehat{\xi}_{ik} = \int_{\mathscr{T}} X_i(t)\widehat{\phi}_k(t)\mathrm{d}t$. We will use this characterization in this paper.

Benatia et al. (2017) propose an estimator similar to (2.2). The sum over k is not truncated, but regularized with the term $\widehat{\lambda}_k + \kappa$

$$\widehat{\theta}^{\ddagger}(t,s) = \frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{n}\frac{\widehat{\xi}_{ik}Y_i(s)}{\widehat{\lambda}_k+\kappa}\widehat{\phi}_k(t).$$

Ramsay and Silverman (2005) obtain the following characterization of the bivariate functional coefficient, for some small or large numbers ι_1 and ι_2, as

$$\theta^{\dagger}(t,s) = \sum_{k=1}^{\iota_1}\sum_{j=1}^{\iota_2}\theta_{k,j}\phi_k(t)\psi_j(s), \quad \text{where } \theta_{k,j} = \frac{\mathbb{E}\big(\xi_k\beta_j\big)}{\lambda_k}.$$

Therefore Imaizumi and Kato (2018) obtain the following characterization

$$\theta^{\dagger\dagger}(t,s) = \sum_{k=1}^{+\infty}\frac{\mathbb{E}\big(\xi_k\beta_j\big)}{\lambda_k}\phi_k(t)\psi_j(s).$$

Notice that $\theta^{\dagger}(t,s)$ and $\theta^{\dagger\dagger}(t,s)$ based on truncating the double series, namely, the double truncation, which will not be discussed here.

2.2 Operatorial point of view

We notice in this subsection that the model (2.1) can be seen from an operatorial point of view. Indeed, we can write the model

$$Y(\boldsymbol{s}) = \Big(\Theta\cdot X\Big)(\boldsymbol{s}) + \epsilon(\boldsymbol{s}),$$

where the slope operator $\Theta : \mathbb{L}^2(\mathscr{T}\times\mathscr{S}) \to \mathbb{L}^2(\mathscr{S})$ is an integral continuous operator defined for all $u \in \mathbb{L}^2(\mathscr{T})$ by $(\Theta\cdot u)(s) = \int_{\mathscr{T}} u(t)\theta(t,s)\mathrm{d}t$ which kernel is $\theta \in \mathbb{L}^2(\mathscr{T}\times\mathscr{S})$.

To close this section, we introduce the operators that will be used through the following theorems in which we describe prediction convergence rates. The covariance operator of X, denoted Γ, is defined by

$$\Gamma(u) = \mathbb{E}\big(X\otimes X(u)\big), \quad \text{for all } u\in\mathbb{L}^2(\mathscr{T}).$$

Note that the covariance operator is a natural extension of the covariance matrix, in the infinite dimensional framework. We also introduce the cross-covariance operator Δ of (X,Y) given by

$$\Delta(u) = \mathbb{E}\big(Y\otimes X(u)\big), \quad \text{for all } u\in\mathbb{L}^2(\mathscr{T}).$$

For any integer k, we define Π_k the orthogonal projection operator on the subspace $\mathrm{Span}(\phi_1,\ldots,\phi_k)$, given by

$$\Pi_k = \sum_{j=1}^{k}\phi_j\otimes\phi_j.$$

Empirical counterparts of Γ, Δ and Π_k, respectively, are denoted $\widehat{\Gamma}_n$, $\widehat{\Delta}_n$ and $\widehat{\Pi}_{k_n}$. These operators are naturally defined by $\widehat{\Gamma}_n = \frac{1}{n}\sum_{i=1}^n X_i \otimes X_i$, $\widehat{\Delta}_n = \frac{1}{n}\sum_{i=1}^n Y_i \otimes X_i$ and $\widehat{\Pi}_{k_n} = \sum_{j=1}^{k_n} \widehat{\phi}_j \otimes \widehat{\phi}_j$. The functional principal component regression estimator $\widehat{\Theta}$ of Θ is defined by

$$\widehat{\Theta} = \widehat{\Pi}_{k_n}\widehat{\Delta}_n(\widehat{\Pi}_{k_n}\widehat{\Gamma}_n\widehat{\Pi}_{k_n})^{-1}.$$

3 The centered Function-on-function linear model with partially observed covariate and response: Reconstructing X and Y

In this section, we are interested in the most general case of missing data in the centered function-on-function linear regression: when both the functional covariate and the functional response are partially observed. We follow the methodology studied in Kneip and Liebl (2020) for reconstructing the missing parts of the curves. Next, once the initial sample is completed, we will present the estimation of the slope operator Θ or its kernel θ and predict new values for the response.

In the following, we denote "$O^{[Y]}$", "$M^{[Y]}$", "$O^{[X]}$" and "$M^{[X]}$" a given production of $O_i^{[Y_i]}$, $M_i^{[Y_i]}$, $O_i^{[X_i]}$ and $M_i^{[X_i]}$, respectively corresponding to the observed domain and the missing domain of Y_i included in $\mathscr{S}$ and the observed domain and the missing domain of X_i included in $\mathscr{T}$. In addition, the corresponding parts of the curve X_i are denoted $X_i^{O^{[X]}}$ and $X_i^{M^{[X]}}$. Similarly, the corresponding parts of Y_i are denoted $Y_i^{O^{[Y]}}$ and $Y_i^{M^{[Y]}}$.

3.1 Curve reconstruction of the covariate and the response

The Karhunen–Loève expansions of the observed sample curves $X_i^{O^{[X]}}$ in $\mathbb{L}^2(O^{[X]})$ and $Y_i^{O^{[Y]}}$ in $\mathbb{L}^2(O^{[Y]})$ are written

$$X_i^{O^{[X]}}(t) = \sum_{k=1}^{+\infty} \xi_{ik}^{O^{[X]}}\phi_k^{O^{[X]}}(t) \quad \text{and} \quad Y_i^{O^{[Y]}}(s) = \sum_{j=1}^{+\infty} \beta_{ij}^{O^{[Y]}}\psi_j^{O^{[Y]}}(s),$$

where $\xi_{ik}^{O^{[X]}} = \int_{\mathscr{T}} X_i^{O^{[X]}}(t)\phi_k^{O^{[X]}}(t)\mathrm{d}t$ and $\beta_{ij}^{O^{[Y]}} = \int_{\mathscr{S}} Y_i^{O^{[X]}}(s)\psi_j^{O^{[X]}}(s)\mathrm{d}s$ are uncorrelated random variables with zero mean and variances $\mathbb{E}\left(\xi_k^{O^{[X]}}\right)^2 = \lambda_k^{O^{[X]}}$ and $\mathbb{E}\left(\beta_j^{O^{[Y]}}\right)^2 = \mu_j^{O^{[Y]}}$ for all $k, j \geqslant 1$.

We consider a reconstruction problem relating the missing part of the curves to the observed part, writing

$$X_i^{M^{[X]}}(t_2) = L^{[X]}(X_i^{O^{[X]}}(t_1)) + \mathscr{Z}_i^{[X]}(t_2), \quad \text{for all } t_1 \in O^{[X]} \text{ and } t_2 \in M^{[X]},$$

and

$$Y_i^{M^{[Y]}}(s_2) = L^{[Y]}(Y_i^{O^{[Y]}}(s_1)) + \mathscr{Z}_i^{[Y]}(s_2), \quad \text{for all } s_1 \in O^{[Y]} \text{ and } s_2 \in M^{[Y]},$$

where $L^{[X]} : \mathbb{L}_2(O^{[X]}) \to \mathbb{L}_2(M^{[X]})$ and $L^{[Y]} : \mathbb{L}_2(O^{[Y]}) \to \mathbb{L}_2(M^{[Y]})$ are linear reconstruction operators and $\mathscr{Z}_i^{[X]} \in \mathbb{L}_2(M^{[X]})$ and $\mathscr{Z}_i^{[Y]} \in \mathbb{L}_2(M^{[Y]})$ are reconstruction errors. We need to minimize the mean square error between the curves with fragmentary data and the linear reconstruction operators $L^{[X]}$ and $L^{[Y]}$ as follows

$$\mathbb{E}\Big(\big\|X_i^{M^{[X]}} - \mathrm{L}^{[X]}(X_i^{O^{[X]}})\big\|^2\Big) \quad \text{and} \quad \mathbb{E}\Big(\big\|Y_i^{M^{[Y]}} - \mathrm{L}^{[Y]}(Y_i^{O^{[Y]}})\big\|^2\Big). \tag{3.1}$$

Thes operator $\mathrm{L}^{[X]}$ $\big($or $\mathrm{L}^{[Y]}\big)$ has been studied by Kraus (2015) and Kneip and Liebl (2020). Kraus (2015) proposed to use the ridge regularization method. The estimator is introduced by $\mathscr{L}^{(\alpha)} = \mathscr{R}_{MO}\mathscr{R}_{OO}^{(\alpha)-1}$ where $\mathscr{R}$ is the covariance operator defined on $\mathbb{L}^2(\mathscr{T})$ and $\mathscr{R}_{OO}^{(\alpha)-1} = \mathscr{R}_{OO} + \alpha\,\mathscr{J}_O$ where α is a positive parameter and $\mathscr{J}_O$ is the identity operator on $\mathbb{L}^2(O^{[X]})$. Besides, Kraus and Stefanucci (2020) prove that the reconstruction operator (considered in Kraus (2015)) can be seen as Hilbert-Schmidt integral operator from $\mathbb{L}^2(O^{[X]})$ to $\mathbb{L}^2(O^{[X]})$ writing

$$\mathrm{L}(X_i^{O^{[X]}}) = \int_{O^{[X]}} a_i(.,t)X_i^{O^{[X]}}(t)\mathrm{d}t,$$

for all $i = 1, \ldots, n$, where a is a square integrable function on $M^{[X]} \times O^{[X]}$.

The optimal reconstruction operators minimizing (3.1) are denoted $\mathscr{L}(X_i^{O^{[X]}})$ and $\mathscr{J}(Y_i^{O^{[Y]}})$ and defined by

$$\mathscr{L}(X_i^{O^{[X]}})(t_2) = \sum_{k=1}^{+\infty} \xi_{ik}^{O^{[X]}} \frac{\langle \phi_k^{O^{[X]}}, \gamma_{t_2}\rangle}{\lambda_k^{O^{[X]}}} \text{ and } \mathscr{J}(Y_i^{O^{[Y]}})(s_2) = \sum_{j=1}^{+\infty} \beta_{ij}^{O^{[Y]}} \frac{\langle \psi_j^{O^{[Y]}}, \gamma_{s_2}\rangle}{\mu_j^{O^{[Y]}}}, \tag{3.2}$$

where $\gamma_{t_2}(t_1) = \mathbb{E}\left(X_i^{M^{[X]}}(t_2)X_i^{O^{[X]}}(t_1)\right)$, for all $t_1 \in O^{[X]}$ and $t_2 \in M^{[X]}$, and $\gamma_{s_2}(s_1) = \mathbb{E}\left(Y_i^{M^{[Y]}}(s_2)Y_i^{O^{[Y]}}(s_1)\right)$, for all $s_1 \in O^{[Y]}$ and $s_2 \in M^{[Y]}$.

The operators in relation (3.2) are estimated as in Kneip and Liebl (2020, Section 2) and Crambes et al. (2023, Subsection 2.1 and Subsection 2.2) by $\widehat{\mathscr{L}_{k_n}}(X_i^{O^{[X]}})$ and $\widehat{\mathscr{J}_{j_n}}(Y_i^{O^{[Y]}})$, where the truncation parameters k_n and j_n are positive integers that can be fixed automatically with a grid search. The solution of (3.2) uses local linear smoothers for unknown quantities in (3.2), considering the following notations. Let κ_1 and κ_1' be kernels and h_X and h_Y be bandwidths of the local linear smoothers of the curves $X_i^{O^{[X]}}$ and $Y_i^{O^{[Y]}}$, respectively. Moreover, let κ_2 and κ_2' be bivariate kernels and $h_{\gamma_{t_2}}$ and $h_{\gamma_{s_2}}$ be bandwidths of the local linear smoothers of the covariance functions γ_{t_2} and γ_{s_2}, respectively.

In the following, we consider the whole sample $\mathscr{D}_n \triangleq \{(X_1^\star, Y_1^\star), \ldots, (X_n^\star, Y_n^\star)\}$, with possibly reconstructed curves on the missing parts, that is

$$X_i^\star(t) = \begin{cases} X_i^{O^{[X]}}(t) & \text{if } t \in O^{[X]}, \\ \widehat{\mathscr{L}_{k_n}}(X_i^{O^{[X]}}(t)) & \text{if } t \in M^{[X]} \end{cases} \quad \text{and} \quad Y_i^\star(s) = \begin{cases} Y_i^{O^{[Y]}}(s) & \text{if } s \in O^{[Y]}, \\ \widehat{\mathscr{J}_{j_n}}(Y_i^{O^{[Y]}})(s) & \text{if } s \in M^{[Y]}. \end{cases}$$

3.2 Estimation of slope operator and its kernel and prediction

We estimate the kernel function θ with

$$\widehat{\theta}^{\star}(t,s) = \frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{k_n}\frac{\widehat{\xi}^{\star}_{ik,rec}Y_i^{\star}(s)}{\widehat{\lambda}^{\star}_{k,rec}}\widehat{\phi}^{\star}_{k,rec}(t), \tag{3.3}$$

where $\widehat{\phi}^{\star}_{1,rec},\ldots,\widehat{\phi}^{\star}_{k_n,rec}$ and $\widehat{\lambda}^{\star}_{1,rec},\ldots,\widehat{\lambda}^{\star}_{k_n,rec}$ represent respectively the k_n first eigenfunctions and eigenvalues of the covariance operator $\widehat{\Gamma}^{\star}_{n,rec}$ and $\widehat{\xi}^{\star}_{ik,rec} = \int_{\mathscr{T}} X_i^{\star}(t)\widehat{\phi}^{\star}_{k,rec}(t)\mathrm{d}t$ are the estimates of the FPC scores.

From an operatorial point of view, the covariance operator $\Gamma^{\star}_{rec}$ of $X^{\star}$ and the cross-covariance operator $\Delta^{\star}_{rec}$ of $(X^{\star},Y^{\star})$ are given by $\Gamma^{\star}_{rec} = \mathbb{E}[X^{\star}\otimes X^{\star}]$ and $\Delta^{\star}_{rec} = \mathbb{E}[Y^{\star}\otimes X^{\star}]$, for all $u\in\mathbb{L}^2(\mathscr{T})$. The orthogonal projection operator $\Pi^{\star}_{k,rec}$ on the subspace $\mathrm{Span}(\phi^{\star}_{1,rec},\ldots,\phi^{\star}_{k,rec})$ is given by $\Pi^{\star}_{k} = \sum_{j=1}^{k}\phi^{\star}_{j,rec}\otimes\phi^{\star}_{j,rec}$. The empirical counterparts of $\Gamma^{\star}_{rec}$, $\Delta^{\star}_{rec}$ and $\Pi^{\star}_{k,rec}$, are denoted respectively $\widehat{\Gamma}^{\star}_{n,rec}$, $\widehat{\Delta}^{\star}_{n,rec}$ and $\widehat{\Pi}^{\star}_{k_n,rec}$. The estimator of the slope operator Θ is given by

$$\widehat{\Theta}^{\star} = \langle\widehat{\theta}^{\star}(\cdot,s),\cdot\rangle = \widehat{\Pi}^{\star}_{k_n,rec}\widehat{\Delta}^{\star}_{n,rec}\left(\widehat{\Pi}^{\star}_{k_n,rec}\widehat{\Gamma}^{\star}_{n,rec}\widehat{\Pi}^{\star}_{k_n,rec}\right)^{-1}.$$

Finally, we obtain the prediction of the response when a new explanatory curve X_{new} is given, by

$$Y^{\star}_{new}(s) = \int_{\mathscr{T}} X^{\star}_{new}(t)\widehat{\theta}^{\star}(t,s)\mathrm{d}t = \frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{k_n}\frac{\widehat{\xi}^{\star}_{ik,rec}\widehat{\xi}^{\star}_{new;k,rec}Y_i^{\star}(s)}{\widehat{\lambda}^{\star}_{k,rec}},$$

for all $s\in\mathscr{S}$, where $\widehat{\xi}^{\star}_{new;k,rec} = \int_{\mathscr{T}} X^{\star}_{new}(t)\widehat{\phi}^{\star}_{k,rec}(t)\mathrm{d}t$. Alternatively, it is also given with the operatorial quantities

$$Y^{\star}_{new} = \widehat{\Pi}^{\star}_{k_n,rec}\widehat{\Delta}^{\star}_{n,rec}\left(\widehat{\Pi}^{\star}_{k_n,rec}\widehat{\Gamma}^{\star}_{n,rec}\widehat{\Pi}^{\star}_{k_n,rec}\right)^{-1}X^{\star}_{new}.$$

3.3 Assumptions

To achieve our theoretical results, we need first to adopt some classical assumptions which have been similarly used in Kneip and Liebl (2020) and Crambes et al. (2023) to control the curve reconstruction for the covariate and the response.

(A.1) X and Y have finite fourth moment order.

(A.2) Let $np\to\infty$ when $n\to\infty$, where $p = p(n)$ is the number of observation points of the covariate. Similarly, $nq\to\infty$ when $n\to\infty$, where $q = q(n)$ is the number of observation points of the response. We assume in the following that $p = n^{\eta_1}$ with $0<\eta_1<\infty$ and $q = n^{\zeta_1}$ with $0<\zeta_1<\infty$.

(A.3) • The bandwidth h_X satisfies $h_X\to 0$ and $(ph_X)\to\infty$ as $p\to\infty$. For instance, we assume that $h_X = \frac{1}{n^{\eta_2}}$ with $0<\eta_2<\eta_1$. The bandwidth $h_{\gamma_{t_2}}$ satisfies $h_{\gamma_{t_2}}\to 0$ and $(n(p^2-p)h_{\gamma_{t_2}})\to\infty$ as $n(p^2-p)\to\infty$. For example, we can take $h_{\gamma_{t_2}} = \frac{1}{n^{\eta_3}}$ with

$0 < \eta_3 < 2\eta_1 + 1$.

• Let κ_1 and κ_2 be nonnegative, second order univariate and bivariate kernel functions with support $[-1, 1]$. For example, we can use univariate and bivariate Epanechnikov kernel functions with compact support $[-1, 1]$, namely $\kappa_1(x) = \frac{3}{4}(1 - x^2)\mathbb{1}_{[-1,1]}(x)$ and $\kappa_2(x, y) = \frac{9}{16}(1 - x^2)(1 - y^2)\mathbb{1}_{[-1,1]}(x)\mathbb{1}_{[-1,1]}(y)$.

(A.4) • For any subinterval $O^{[X]} \subseteq \mathscr{T}$, we assume that the eigenvalues $\lambda_1 > \lambda_2 > \ldots > 0$ have multiplicity one. Moreover, we assume that there exist $a_O > 1$ and $0 < c_O < \infty$ such that `(i)` $\lambda_k^{O^{[X]}} - \lambda_{k+1}^{O^{[X]}} \geqslant c_O k^{-a_O-1}$, `(ii)` $\lambda_k^{O^{[X]}} = \mathscr{O}(k^{-a_O})$, `(iii)` $1/\lambda_k^{O^{[X]}} = \mathscr{O}(k^{a_O})$ as $k \to \infty$.

• $\mathbb{E}\left(\xi_k^4\right) = \mathscr{O}\left(\lambda_k^2\right)$.

• For any subinterval $O^{[X]} \subseteq \mathscr{T}$, we assume that there exists $0 < A_O < \infty$ such that the eigenfunctions satisfy $\sup_{t\in\mathscr{T}} \sup_{k\geqslant 1} \left|\widetilde{\phi}_k^{O^{[X]}}(t)\right| \leqslant A_O$, where $\widetilde{\phi}_k^{O^{[X]}}(t_2) = \langle \phi_k^{O^{[X]}}, \gamma_{t_2}\rangle/\lambda_k^{O^{[X]}}$.

(A.5) • The bandwidth h_Y satisfies $h_Y \to 0$ and $(qh_Y) \to \infty$ as $q \to \infty$. Moreover, we assume that $h_Y = \frac{1}{n^{\zeta_2}}$ with $0 < \zeta_2 < \zeta_1$. The bandwidth $h_{\gamma_{s_2}}$ satisfies $h_{\gamma_{s_2}} \to 0$ and $(n(q^2 - q)h_{\gamma_{s_2}} \to \infty$ as $n(q^2 - q) \to \infty$. For example, we can take $h_{\gamma_{s_2}} = \frac{1}{n^{\zeta_3}}$ with $0 < \zeta_3 < 2\zeta_1 + 1$.

• Let $\kappa_1^{'}$ and $\kappa_2^{'}$ be nonnegative, second order univariate and bivariate kernel functions with support $[-1, 1]$.

(A.6) • For any subinterval $O^{[Y]} \subseteq \mathscr{S}$, we assume that $\mu_1 > \mu_2 > \ldots > 0$ have multiplicity one and we assume that there exist $b_O > 1$ and $0 < d_O < \infty$ such that `(j)` $\mu_j^{O^{[Y]}} - \mu_{j+1}^{O^{[Y]}} \geqslant d_O j^{-b_O-1}$, `(jj)` $\mu_j^{O^{[Y]}} = \mathscr{O}\left(j^{-b_O}\right)$, `(jjj)` $1/\mu_j^{O^{[Y]}} = \mathscr{O}\left(j^{b_O}\right)$ as $j \to \infty$.

• $\mathbb{E}\left(\beta_k^4\right) = \mathscr{O}\left(\mu_k^2\right)$.

• For any subinterval $O^{[Y]} \subseteq \mathscr{S}$, we assume that there exists $0 < B_O < \infty$ such that $\sup_{s\in\mathscr{S}} \sup_{j\geqslant 1} \left|\widetilde{\psi}_j^{O^{[Y]}}(s)\right| \leqslant B_O$, where $\widetilde{\psi}_j^{O^{[Y]}}(s_2) = \langle \psi_j^{O^{[Y]}}, \gamma_{s_2}\rangle/\lambda_j^{O^{[Y]}}$.

Assumption ***(A.1)*** holds for many processes X and Y (Gaussian processes, bounded processes). Assumption ***(A.2)*** is mild and can be satisfied even if the number of observation points p and q do not go fast to infinity. ***(A.3)*** and ***(A.5)*** are classic assumptions in the context of local polynomials smoothers. Assumptions ***(A.4)*** and ***(A.6)*** are similar to the ones in Kneip and Liebl (2020).

3.4 Asymptotic results

Under assumptions ***(A.1)-(A.6)***, it is proved in Kneip and Liebl (2020) that, in the case where $p \sim n^{\eta_1}$ and $q \sim n^{\zeta_1}$ with $\eta_1, \zeta_1 \leqslant 1/2$ we have for any $t \in \mathscr{T}$ and $s \in \mathscr{S}$

$$|X_i^\star(t) - X_i(t)| = \mathscr{O}_p\left(n^{-\eta_1(a_O-1)/(2(a_O+2))}\right) \text{ and } |Y_i^\star(s) - Y_i(s)| = \mathscr{O}_p\left(n^{-\zeta_1(b_O-1)/(2(b_O+2))}\right).$$

The previous result allows to obtain some bounds between quantities related to functional principal components analysis between the reconstructed curves and with the original curves.

We finish this subsection with the main result giving a bound for the prediction error of Y_{new} with a new value of the covariate X_{new}.

Theorem 3.1. *Under assumptions **(A.1)-(A.6)**, taking* $k_n \sim p^{1/(a_O+2)}$, $p \sim n^{\eta_1}$, $j_n \sim q^{1/(b_O+2)}$ *and* $q \sim n^{\zeta_1}$, *with* $\eta_1, \zeta_1 \leqslant 1/2$, *we get*

$$\mathbb{E}\left(\left\|\left(\widehat{\Theta}^{\star}\cdot X^{\star}_{new} - \Theta \cdot X^{\star}_{new}\right)\right\|^2\right) = \mathscr{O}\left(n^{-\eta_1(a_O-1)/(2(a_O+2))} + n^{\eta_1/(a_O+2)-1-\zeta_1(b_O-1)/(b_O+2)}\right).$$

Corollary 3.2. *We make a comparison between the parameters to find the best convergence error. We summarize the error rates in Table 1.*

Table 1: Convergence error rates depending on the observation points and the regularity of the curves X and Y.

(i) $\eta_1 = \zeta_1$	$a_O \leqslant b_O$	$\mathscr{O}\left(n^{-\eta_1(a_O-1)/(2(a_O+2))}\right)$
	$a_O > b_O$	$\mathscr{O}\left(n^{\eta_1/(a_O+2)-1-\eta_1(b_O-1)/(b_O+2)}\right)$
(ii) $\eta_1 < \zeta_1$		$\mathscr{O}\left(n^{\eta_1/(a_O+2)-1-\zeta_1(b_O-1)/(b_O+2)}\right)$
(iii) $\eta_1 > \zeta_1$		$\mathscr{O}\left(n^{-\eta_1(a_O-1)/(2(a_O+2))}\right)$

Comparing the parameters, we remark all the convergence rates depend in particular on the parameter $a_O > 1$, which is directly linked to the smoothness of the stochastic process X. The larger a_O is, the smoother X is. In the case (i) when the number of observation points of the covariate is equivalent to that of the response for $a_O > b_O$ and in the case (ii), the convergence rates depend of the parameter a_O and also of the parameter $b_O > 1$, which is directly linked to the smoothness of the stochastic process Y. In these cases, the final rate of convergence will be linked to the parameter a_O or b_O corresponding to the less smooth process (either X or Y).

4 The centered function-on-function with partially observed covariate and response: Reconstructing X and imputing Y

We have seen in the previous section the methodology for reconstructing the missing parts of the explanatory curves. In this section, we try another strategy to deal with missing data on the response. After reconstructing the missing parts of the covariate X, we apply the regression imputation methodology as presented in Crambes et al. (2023) for a real response. Next, we will present the estimation of the kernel function θ and predict the new response once all sample is completed.

4.1 Regression imputation on the functional response

We consider the following missing data mechanism for the response, through a variable $\delta^{[Y]}$ leading to the sample $(\delta_i^{[Y]})_{i=1,\dots,n}$ such that,

$$\delta_i^{[Y]} = \begin{cases} 0 \text{ if } M_i^{[Y]} \neq \varnothing, \\ 1 \text{ if } O_i^{[Y]} = \mathscr{S}. \end{cases}$$

We assume that the response is missing at random (MAR), which means that the fact that Y contains missing parts does not depend on the response of the model, but can possibly depend on the reconstructed covariate,

$$\mathbb{P}(\delta^{[Y]} = 1 \mid X^\star, Y) = \mathbb{P}(\delta^{[Y]} = 1 \mid X^\star).$$

We denote the number of curves partially observed by

$$m_n^{[Y]} = \sum_{i=1}^{n} \mathbb{1}_{\{\delta_i^{[Y]}=0\}}.$$

Using the exponent notation "obs" to make reference to the units for which the response is completely observed, we define the covariance operator Γ_{rec}^{obs} and the cross-covariance operator Δ_{rec}^{obs} with the reconstructed curves by $\Gamma_{rec}^{obs} = \mathbb{E}\left(\delta^{[Y]} X^\star \otimes X^\star\right)$ and $\Delta_{rec}^{obs} = \mathbb{E}\left(\delta^{[Y]} Y \otimes X^\star\right)$. The empirical counterparts of Γ_{rec}^{obs} and Δ_{rec}^{obs} are denoted respectively $\widehat{\Gamma}_{n,rec}^{obs}$ and $\widehat{\Delta}_{n,rec}^{obs}$.

Let Y_ℓ be a response curve such that $\delta_\ell^{[Y]} = 0$, we define the imputed response $Y_{\ell,imp}$ by

$$Y_{\ell,imp}(s) = \int_{\mathscr{T}} X_\ell^\star(t) \widetilde{\theta}(t,s) \mathrm{d}t,$$

with

$$\widetilde{\theta}(t,s) = \frac{1}{n - m_n^{[Y]}} \sum_{i=1}^{n} \sum_{k=1}^{k_n} \frac{\widehat{\xi}_{ik,rec}^{obs} \delta_i^{[Y]}(s) Y_i^\star(s)}{\widehat{\lambda}_{k,rec}^{obs}} \widehat{\phi}_{k,rec}^{obs}(t), \tag{4.1}$$

for all $(t,s) \in \mathscr{T} \times \mathscr{S}$, where $\widehat{\xi}_{ik,rec}^{obs} = \int_{\mathscr{T}} X_i^\star(t) \widehat{\phi}_{k,rec}^{obs}(t) \mathrm{d}t$ and $\widehat{\phi}_{1,rec}^{obs}, \dots, \widehat{\phi}_{k_n,rec}^{obs}$ and $\widehat{\lambda}_{1,rec}^{obs} \dots, \widehat{\lambda}_{k_n,rec}^{obs}$ represent respectively the k_n first eigenfunctions and eigenvalues of the covariance operator $\widehat{\Gamma}_{n,rec}^{obs}$.

Alternatively, if we denote $\widehat{\Pi}_{k_n,rec}^{obs}$ the projection on the space spanned by the k_n first eigenfunctions of $\widehat{\Gamma}_{n,rec}^{obs}$, the estimation of the slope operator Θ is given by

$$\widetilde{\Theta} = \widehat{\Pi}_{k_n,rec}^{obs} \widehat{\Delta}_{n,rec}^{obs} \left(\widehat{\Pi}_{k_n,rec}^{obs} \widehat{\Gamma}_{k_n,rec}^{obs} \widehat{\Pi}_{k_n,rec}^{obs} \right)^{-1}.$$

and the imputation $Y_{\ell,imp}$ can also be written

$$Y_{\ell,imp}(s) = \widehat{\Pi}_{k_n,rec}^{obs} \widehat{\Delta}_{n,rec}^{obs} \left(\widehat{\Pi}_{k_n,rec}^{obs} \widehat{\Gamma}_{k_n,rec}^{obs} \widehat{\Pi}_{k_n,rec}^{obs} \right)^{-1} X_\ell^\star(t),$$

4.2 Estimation of the slope operator and its kernel and prediction

Once the whole database has been reconstructed, we estimate the bivariate functional coefficient θ with

$$\widehat{\theta}^{\star\star}(t,s) = \frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{k_n} \frac{\widehat{\xi}^{\star}_{ik,rec} Y_i^{\star\star}(s)}{\widehat{\lambda}^{\star}_{k,rec}} \widehat{\phi}^{\star}_{k,rec}(t), \tag{4.2}$$

for all $(t,s) \in \mathscr{T}\times\mathscr{S}$, where $Y_i^{\star\star} = Y_i\delta_i^{[Y]} + Y_{i,imp}(1-\delta_i^{[Y]})$ for all $i=1,\ldots,n$. The estimation of the operator Θ is similarly given by

$$\widehat{\Theta}^{\star\star} = \widehat{\Pi}^{\star}_{k_n,rec}\widehat{\Delta}^{\star\star}_{n,rec}\left(\widehat{\Pi}^{\star}_{k_n,rec}\widehat{\Gamma}^{\star}_{n,rec}\widehat{\Pi}^{\star}_{k_n,rec}\right)^{-1},$$

where $\widehat{\Delta}^{\star\star}_{n,rec}$ is the empirical counterpart of the cross-covariance operator $\Delta^{\star\star}_{rec}$.

We use this estimation to predict a new value of the response Y when a new explanatory curve X_{new} is given by

$$\widehat{Y}_{new}(s) = \int_{\mathscr{T}} X^{\star}_{new}(t)\widehat{\theta}^{\star\star}(t,s)\mathrm{d}(t),$$

which can also be written

$$\widehat{Y}_{new} = \widehat{\Pi}^{\star}_{k_n,rec}\widehat{\Delta}^{\star\star}_{n,rec}\left(\widehat{\Pi}^{\star}_{k_n,rec}\widehat{\Gamma}^{\star}_{k_n,rec}\widehat{\Pi}^{\star}_{k_n,rec}\right)^{-1} X^{\star}_{new}.$$

4.3 Asymptotic results

The proof of these results follows the same lines as the proof of **Theorem (3.1)** and **Theorem (3.3)** in Crambes et al. (2023).

Theorem 4.1. *Under assumptions **(A.1)-(A.4)**, if we take $k_n \sim p^{1/(a_O+2)}$ and $p \sim n^{\eta_1}$ with $\eta_1 \leqslant 1/2$, we obtain*

$$\mathbb{E}\left(\left\|\left(\widetilde{\Theta}\cdot X^{\star}_{\ell} - \Theta\cdot X^{\star}_{\ell}\right)\right\|^2\right) = \mathscr{O}\left(n^{-\eta_1(a_O-1)/(2(a_O+2))} + \frac{n^{\eta_1/(a_O+2)}}{n-m_n^{[Y]}}\right),$$

for $\ell \in \widetilde{\mathscr{D}}_m$, where $\widetilde{\mathscr{D}}_m$ is the set of missing responses of size $m_n^{[Y]}$.

Theorem 4.2. *Under assumptions **(A.1)-(A.4)**, and $k_n \sim p^{1/(a_O+2)}$ and $p \sim n^{\eta_1}$ with $\eta_1 \leqslant 1/2$, the prediction error is*

$$\mathbb{E}\left(\left\|\left(\widehat{\Theta}^{\star\star}\cdot X^{\star}_{new} - \Theta\cdot X^{\star}_{new}\right)\right\|^2\right) = \mathscr{O}\left(n^{-\eta_1(a_O-1)/(2(a_O+2))} + \frac{n^{\eta_1/(a_O+2)}}{n-m_n^{[Y]}}\right).$$

Remark 4.3. *Comparing parameters as $n^{\eta_1/(a_O+2)-1-\zeta_1(b_O-1)/(b_O+2)} \lesssim \frac{n^{\eta_1/(a_O+2)}}{n-m_n^{[Y]}}$, we find that the prediction error with reconstruction (obtained in Section 3) is asymptotically at least the same than the prediction error with imputation.*

Remark 4.4. *Theoretical results are generally obtained under assumptions concerning the rate of convergence of the integer k_n. In practice, this integer is selected by minimizing a certain empirical criterion such as Generalized Cross Validation (GCV) criterion, Cross Validation (CV) criterion, or K-fold Cross Validation (K-fold CV) criterion. We chose in the following simulation section the GCV procedure, known to be computationally fast. The GCV criteria is given as follows for imputation*

$$GCV(k_n) = \frac{(n - m_n^{[Y]}) \sum_{i=1}^{n} \left\| \left(\widetilde{\Theta} \cdot X_i^{\star} - \Theta \cdot X_i^{\star} \right) \right\|^2 \delta_i^{[Y]}}{\left((n - m_n^{[Y]}) - k_n \right)^2},$$

and analogously for prediction.

5 Simulations

5.1 Methodology

In this section, we conduct Monte Carlo experiments to illustrate the finite-sample performance of the proposed methods presented in section 3 and section 4. We set $\mathscr{S} = \mathscr{T} = [0, 1]$. Each response and predictor curve is observed at $q = 90$ and $p = 100$ equally spaced points in their domains, respectively. For computational simplicity we consider equidistant points $s_j = j/(q-1)$, $j = 0, \ldots, q-1$, and $t_k = k/(p-1)$, $k = 0, \ldots, p-1$. Two simulated data mechanisms are generated. Each model is defined by

$$Y^{(w)}(s) = \int_0^1 X^{(w)}(t)\theta^{(w)}(t, s)\mathrm{d}t + \epsilon^{(w)}(s), \tag{5.1}$$

for $t \in \mathscr{T}$ and $s \in \mathscr{S}$, where $w = 1, 2$. We approximate the integral in (5.1) using a Riemann sum over the grid $\boldsymbol{t}$. The analytical expressions of the kernels and processes are given below.

Scenario 1 The kernel is given by

$$\theta^{(1)}(t, s) = \sum_{k=1}^{50} \sum_{j=1}^{50} \theta_{k,j} \phi_k(t) \psi_j(s),$$

where $\phi_k(t) = \sqrt{2}\cos(k\pi t)$, $\psi_j(s) = \sqrt{2}\cos(j\pi s)$ and $\theta_{k,j} = 4(-1)^{k+j} k^{-2.5} j^{-3}$. The input is the random function $X^{(1)}(t) = \sum_{k=1}^{50} k^{-1}\xi_k \phi_k(t)$, where the ξ_k's are independently sampled from the uniform distribution on $[-\sqrt{3}, \sqrt{3}]$. Finally, the noise is given by $\epsilon^{(1)} = \sum_{j=1}^{50} \beta_j' \psi_j(s)$ and the β_j''s are independent mean-zero Gaussian random variables with variances equal to $j^{-1.1}$. Then, the covariance function $\mathrm{C}_{X^{(1)}}$ of $X^{(1)}$ is given by

$$\mathrm{C}_{X^{(1)}}(t_1, t_2) = \sum_{k=1}^{\iota_3} \frac{2}{k^2} \cos(k\pi t_1) \cos(k\pi t_2), \quad \text{where } t_1, t_2 \in \mathscr{T}.$$

SCENARIO 2 The kernel is defined by $\theta^{(2)}(t,s) = s^3 + \sin(2\pi t)^3$ and the noise $\epsilon^{(2)}$ is generated according to a standard Brownian motion divided by 20. In addition, the functional covariate $X^{(2)}$ is generated through its covariance function, defined, for all $t_1, t_2 \in \mathscr{T}$, by

$$\mathrm{C}_{X^{(2)}}(t_1,t_2) = \frac{\sigma_1^2 \exp\left(-|t_1 - t_2|^\alpha\right)}{\varsigma},$$

with $\sigma_1 = 1$, $\alpha = 2$ and $\varsigma = 0.2$. In this setting, even if a polynomial decrease of the eigenvalues of the covariance operator of $X^{(2)}$ is required in our theoretical results (see assumption ***(A.4)***), we want to see how the method works in practice if this assumption is no more satisfied, namely here in the case of an exponential decay.

Figure 1 shows 10 discretized predictor functions $X_i^{(1)}$, the error functions $\epsilon_i^{(1)}$ and the response functions $Y_i^{(1)}$. Figure 2 shows 10 discretized predictor functions $X_i^{(2)}$, the error functions $\epsilon_i^{(2)}$ and the response functions $Y_i^{(2)}$.

Figure 3 shows the covariance functions of the covariate for SCENARIO 1 and 2. Figure 4 shows kernel functions for SCENARIO 1 and 2.

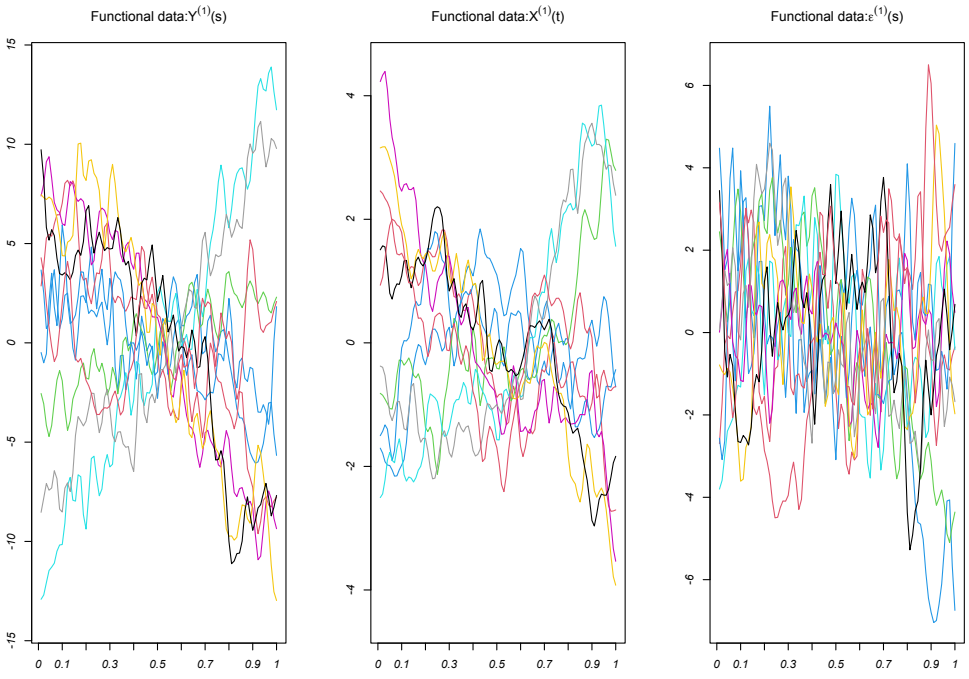

Figure 1: Examples of simulated functions with SCENARIO 1.

To deal with partially observed curves for the covariate and response, we adopted the missing data simulation scenario from Kneip and Liebl (2020) and Crambes et al. (2023) such that

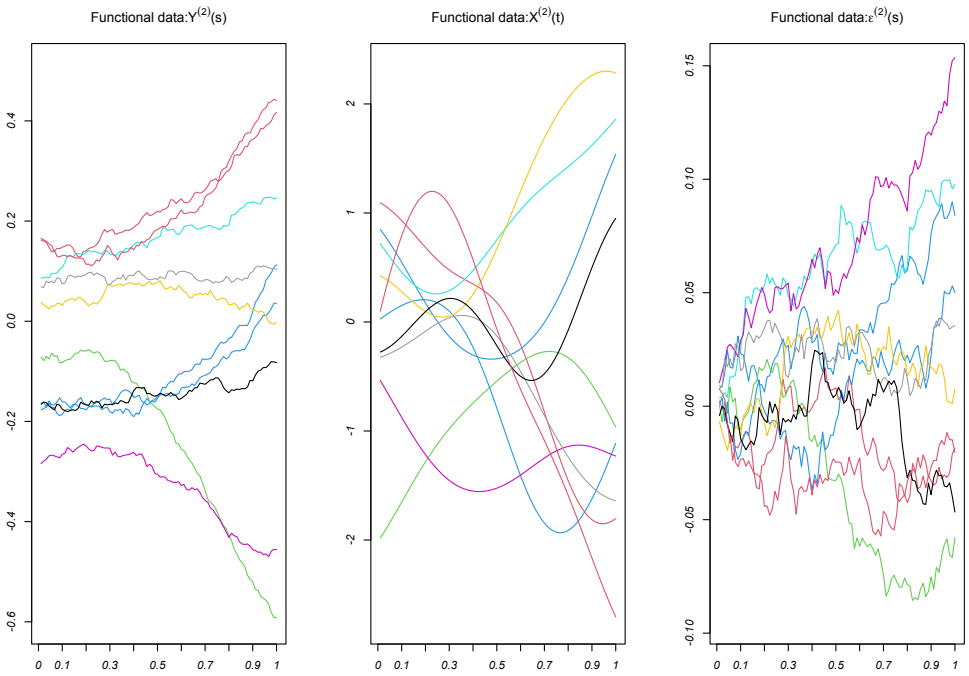

Figure 2: Examples of simulated functions with Scenario 2.

- 70% (respectively 55%) of the curves are fully observed on $[0, 1]$,
- for the 30% (respectively 45%) of partially observed curves, the curves X_i and Y_i are fully observed on $[A_i, B_i] \subset [0, 1]$ with A_i drawn with uniform law on the interval $[0, A]$ and $B_i = A_i + B$, with either $A = 1/50$ and $B = 49/50$ or $A = 3/50$ and $B = 47/50$ for Scenario 1. We take either $A = 1/50$ and $B = 49/50$ or $A = 5/50$ and $B = 45/50$ for Scenario 2.

We simulate the number of missing data on the response Y and the indicator $\delta^{[Y]}$ by the logistic functional regression. The variable δ follows the Bernoulli law with parameter $p(X)$ such that

$$\log\left(\frac{p(X)}{1-p(X)}\right) = \left[\int_S \left(\int_{\mathscr{T}} |s-t|X(t)dt\right)^2 ds\right]^{1/2} + ct,$$

where ct is a constant allowing to take different levels of missing data. For exemple, $ct = 2$ gives around 9.727% of missing data and $ct = 0.2$ gives around 38.801% of missing data.

Let us notice that we use a spline smoothed version of the different estimators (2.2), (3.3), (4.1) and (4.2), according to the so-called Smooth Principal Components Regression (SPCR) from Cardot et al. (2003). Let us remark that, with appropriate conditions on the spline parameters, all the theoretical results obtained in our work will also apply when using the SPCR

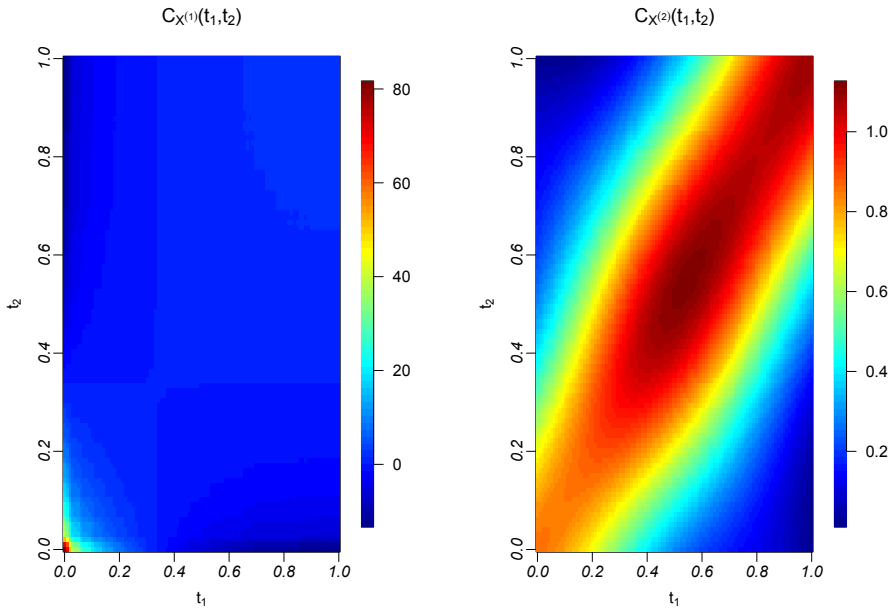

Figure 3: The covariance functions for Scenario 1 and 2.

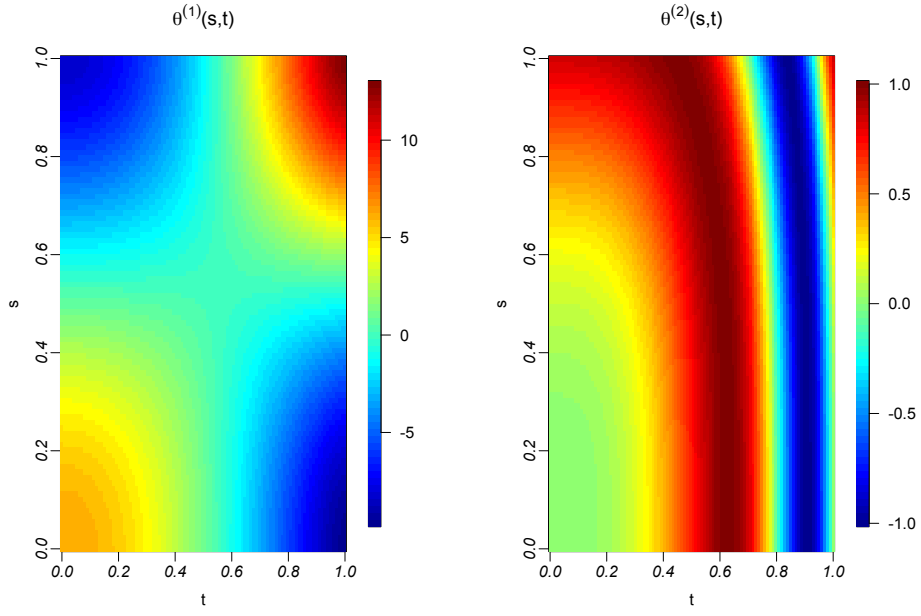

Figure 4: The kernel functions for Scenario 1 and 2.

estimation. We use a regression spline basis with 20 knots, a degree 3 and the order of derivation 2. The choice of these parameters is not crucial in our study, especially in comparison with the choice of the number of principal components. The choice of this optimal tuning parameter is made on a growing sequence of dimension $k_n = 2, \ldots, 22$.

The dataset of size N is randomly splitted into a training set of size $n = \frac{4}{5}N$ and a test set of size $n_1 = \frac{1}{5}N$. We consider sample sizes $N = 500$, 1300. For each scenario, we use 200 Monte Carlo runs for the model assessment. In all numerical experiments, the proposed estimators have been carried out with the free software R.

5.2 Criteria

We use the following criteria to evaluate the performance of the methods.

- Criterion 1: $\overline{MSPE} = \frac{1}{\boldsymbol{Sim}} \sum_{j=1}^{\boldsymbol{Sim}} MSPE(j)$ is the average mean square prediction error. This criterion tends to zero when the sample size tends to infinity, where $MSPE(j) = \frac{1}{n_1} \sum_{\ell=n+1}^{n+n_1} \left\| \left(\widehat{\Theta} \cdot X_\ell^{\star,j} - \Theta \cdot X_\ell^{\star,j} \right) \right\|^2$ is the mean square prediction error computed on the j^{th} simulated sample, $j \in \{1, \ldots, \boldsymbol{Sim}\}$.
- Criterion 2: $\overline{RT} = \frac{1}{\boldsymbol{Sim}} \sum_{j=1}^{\boldsymbol{Sim}} RT(j)$ is the average ratio respect to truth. This criterion tends to one when the sample size tends to infinity, where $RT(j) = \frac{\sum_{\ell=n+1}^{n+n_1} \left\| \left(\widehat{\Theta} \cdot X_\ell^{\star,j} \right) - Y_\ell^j \right\|^2}{\sum_{\ell=n+1}^{n+n_1} \left\| \epsilon_\ell^j \right\|^2}$ is the ratio between the mean square prediction error and the mean square prediction error when the true parameters are known, computed on the j^{th} simulated sample.

We consider another criterion which is the determination coefficient R^2. In this context of functional regression setting, several definitions exist. Given the fitted values $\widehat{Y}_i(s)$, we used the definition as in Harezlak et al. (2007) given by

$$R^2 = \frac{1}{|\mathscr{S}|} \int_{\mathscr{S}} R^2(s) \mathrm{d}s = \frac{1}{|\mathscr{S}|} \int_{\mathscr{S}} \left(1 - \frac{\sum_{i=1}^{n} \left(Y_i(s) - \widehat{Y}_i(s) \right)^2}{\sum_{i=1}^{n} Y_i(s)^2} \right) \mathrm{d}s.$$

5.3 Simulation results

We denote the methods presented in this paper by :

- **Reconst_X_Y** : X and Y are partially observed , the missing parts of X and Y are reconstructed.
- **Reconst_X, Imp_Y** : X and Y are partially observed, the missing parts of X are reconstructed and Y imputed.

Moreover, we compare to other methods :

- **Full_X_Y** : X and Y are fully observed, this corresponds to the complete reference dataset.
- **Reconst_X, Remov_Y** : X and Y are partially observed, the missing parts of X are reconstructed and the missing part of Y are removed from the sample.
- **Reconst_X, Remov_Y** : X and Y are partially observed, the individuals presenting partially observed curves are removed from the sample.

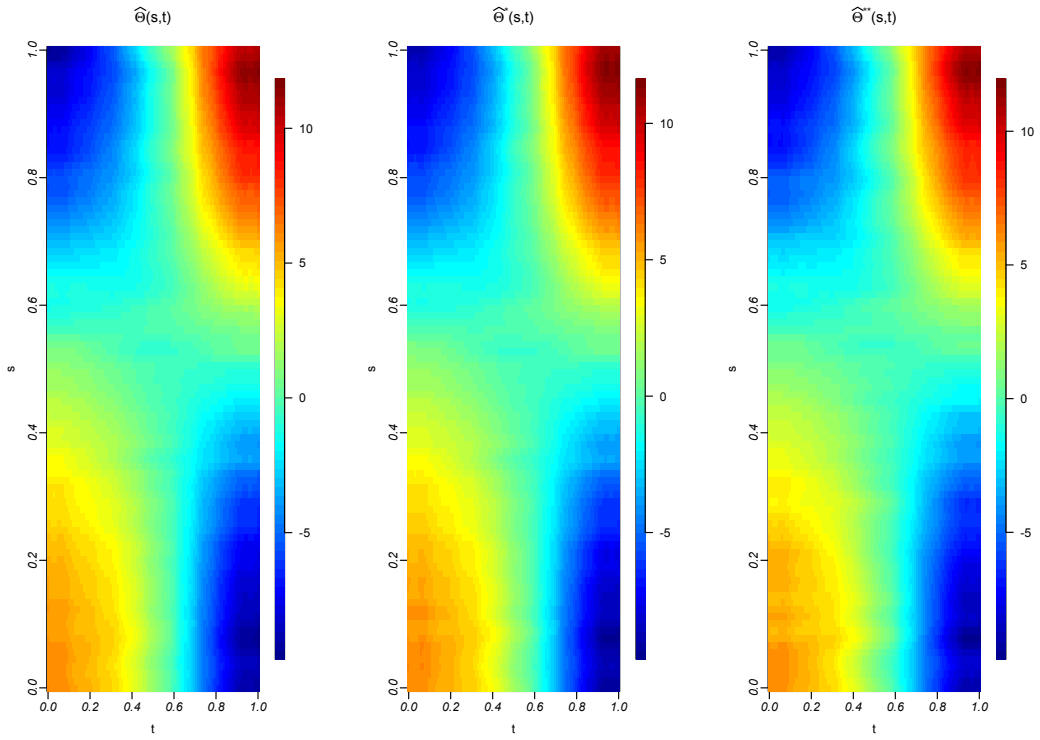

Figure 5: The estimated coefficient functions for Scenario 1.

Even if our main goal is prediction, Figure 5 show estimates of the kernel function in Scenario 1 (with a sample size N = 400) for the dimension $k_n^\star$ chosen by the GCV criterion, respectively with full data ($\widehat{\Theta}$), reconstruction of the missing parts of X and Y ($\widehat{\Theta}^\star$) and reconstruction of the missing parts of X and imputation of Y ($\widehat{\Theta}^{\star\star}$). The missing part is 12% for both curves X and Y, the observed part being [3/50,47/50]. Moreover, 39.375% of curves Y (with ct = 0.1) are affected by missing data and 42.250% of curves X are affected by missing data. We remark that the estimators look graphically quite close, $\widehat{\Theta}^\star$ seems to be a little closer to $\widehat{\Theta}$ than $\widehat{\Theta}^{\star\star}$. A similar plot is obtained for Scenario 2 (Figure 6) with 37.812% of curves Y (with ct = 0.1) affected by missing data and 46.750% of curves X affected by missing data. In this situation, $\widehat{\Theta}^\star$ seems much closer to $\widehat{\Theta}$ than $\widehat{\Theta}^{\star\star}$.

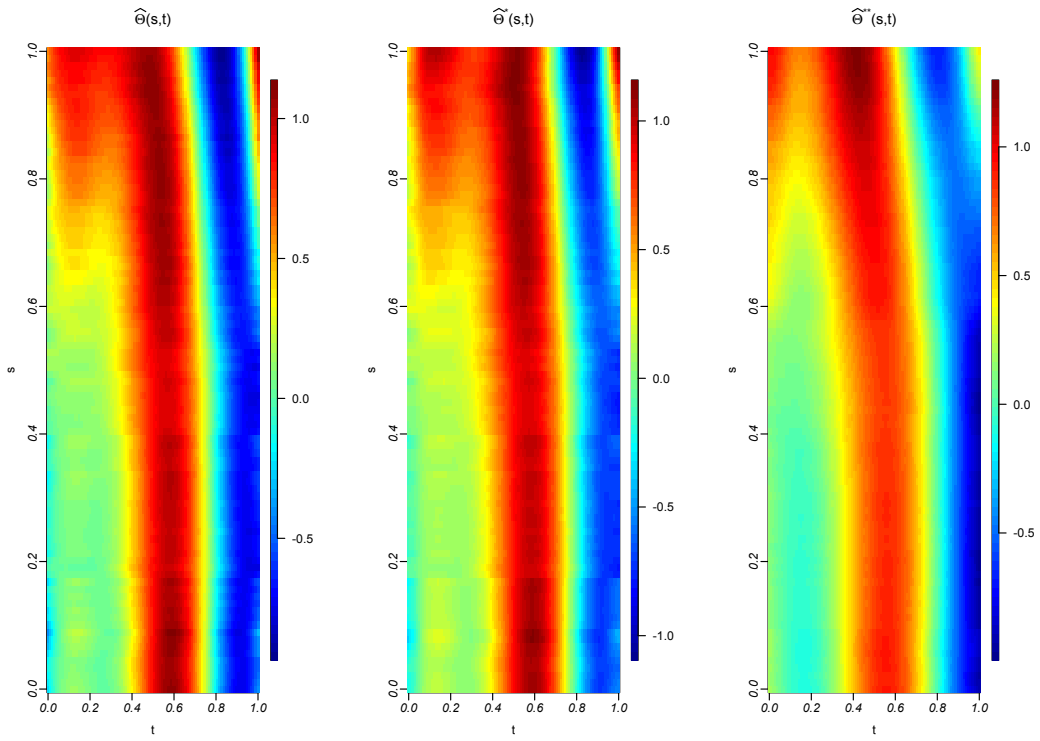

Figure 6: The estimated coefficient functions for Scenario 2.

We give in Table 7 the values of the determination coefficient R^2 and the value $k_n^\star$ chosen by the GCV criterion both scenarios 1 and 2. In scenario 1, we get a worse R^2 coefficient, maybe due to the fact that the curves X are not so smooth and do not seem easy to reconstruct.

Table 2: R^2 and $k_n^\star$ for scenarios 1 and 2.

Methods		Scenario 1	Scenario 2
Full_X_Y	R^2	68.996 %	98.652 %
	$k_n^\star$	2	6
Reconst_X_Y	R^2	68.859 %	98.641 %
	$k_n^\star$	2	6
Reconst_X, Imp_Y	R^2	68.853 %	98.630 %
	$k_n^\star$	2	5

Tables 3, 4, 5 and 6 give the values of the criteria $\overline{MSPE}$ and $\overline{RT}$ for scenarios 1 and 2 with different values of sample size, and different levels of missing data. The first conclusion is the fact that the errors decrease as the sample size increases. Secondly, these errors increase with the percentage of missing data on X or on Y. The rate of missing data on Y seems to

have a more important impact on the errors, whatever the scenario we consider. In all cases, the method **Reconst_X_Y** reconstructing both curves X and Y has a better behaviour than the method **Reconst_X, Imp_Y** reconstructing X and imputing Y, which is quite in accordance to our theoretical results. The part of the observed curve is an important parameter in the curve reconstruction: as it can be expected, the results are better when the curve reconstruction is easier (for example when the observed part is $[1/50, 49/50]$, corresponding to 4% of missing information on the curves). Results tend to deteriorate when the curve reconstruction is harder (for example when the observed part is $[3/50, 47/50]$, corresponding to 12% of missing information on the curves). Finally, these two methods behave better than the other more naive methods (**Reconst_X, Remov_Y** and **Reconst_X, Remov_Y**) that partially or completely ignore missing individuals affected by missing data.

Table 3: Mean and standard deviation errors for the predicted values based on 200 simulation replications with different levels of missing data and a sample size 500 (left panel) and a sample size 1300 (right panel). Partially observed curves are fully observed on $[1/50, 49/50]$ with SCENARIO 1.

Rate of missing	10.150	10.181	40.599	40.542
data in Y in %	(1.407)	(1.481)	(2.742)	(2.596)
Rate of missing	30.114	44.814	29.778	44.958
data in X in %	(1.998)	(2.498)	(2.354)	(2.180)
Full_X_Y : $\overline{MSPE} \times 10^3$	27.618	27.472	27.620	27.312
	(7.073)	(7.162)	(7.584)	(7.077)
$\overline{RT} \times 10$	10.073	10.067	10.071	10.081
	(0.066)	(0.058)	(0.057)	(0.066)
Reconst_X_Y : $\overline{MSPE} \times 10^3$	28.159	28.234	28.843	28.804
	(7.149)	(7.270)	(8.222)	(7.654)
$\overline{RT} \times 10$	10.075	10.069	10.072	10.084
	(0.068)	(0.059)	(0.059)	(0.070)
Reconst_X, Imp_Y : $\overline{MSPE} \times 10^3$	30.388	30.547	43.315	41.853
	(7.754)	(8.388)	(13.043)	(10.940)
$\overline{RT} \times 10$	10.080	10.074	10.113	10.120
	(0.070)	(0.061)	(0.073)	(0.082)
Reconst_X, Remov_Y : $\overline{MSPE} \times 10^3$	30.588	30.609	43.981	42.239
	(8.032)	(8.464)	(13.870)	(11.563)
$\overline{RT} \times 10$	10.080	10.074	10.114	10.120
	(0.071)	(0.061)	(0.075)	(0.084)
Remov_X_Y : $\overline{MSPE} \times 10^3$	40.575	48.537	58.408	70.458
	(11.757)	(14.961)	(19.486)	(21.764)
$\overline{RT} \times 10$	10.105	10.112	10.147	10.198
	(0.101)	(0.108)	(0.113)	(0.138)

Rate of missing	10.174	10.236	40.345	40.293
data in Y in %	(0.991)	(0.933)	(1.694)	(1.659)
Rate of missing	30.085	44.919	30.030	44.911
data in X in %	(1.238)	(1.352)	(1.326)	(1.391)
Full_X_Y : $\overline{MSPE} \times 10^3$	12.418	12.473	12.686	12.164
	(2.692)	(3.016)	(3.498)	(2.636)
$\overline{RT} \times 10$	10.029	10.031	10.030	10.033
	(0.026)	(0.027)	(0.026)	(0.025)
Reconst_X_Y : $\overline{MSPE} \times 10^3$	12.909	13.231	13.906	13.399
	(2.726)	(3.013)	(3.982)	(3.101)
$\overline{RT} \times 10$	10.030	10.034	10.033	10.036
	(0.027)	(0.028)	(0.027)	(0.028)
Reconst_X, Imp_Y : $\overline{MSPE} \times 10^3$	14.110	14.111	19.008	19.135
	(3.114)	(3.178)	(4.963)	(4.669)
$\overline{RT} \times 10$	10.032	10.036	10.047	10.052
	(0.026)	(0.029)	(0.031)	(0.035)
Reconst_X, Remov_Y : $\overline{MSPE} \times 10^3$	14.138	14.152	19.538	19.401
	(3.103)	(3.227)	(5.113)	(4.769)
$\overline{RT} \times 10$	10.032	10.036	10.049	10.052
	(0.026)	(0.029)	(0.032)	(0.035)
Remov_X_Y : $\overline{MSPE} \times 10^3$	18.444	22.823	25.462	30.748
	(4.846)	(5.951)	(6.564)	(8.185)
$\overline{RT} \times 10$	10.047	10.058	10.061	10.077
	(0.036)	(0.045)	(0.044)	(0.068)

Table 4: Mean and standard deviation errors for the predicted values based on 200 simulation replications with different levels of missing data and a sample size 500 (left panel) and a sample size 1300 (right panel). Partially observed curves are fully observed on $[3/50, 47/50]$ with SCENARIO 1.

Rate of missing	10.250	10.193	40.324	40.250
data in Y in %	(1.502)	(1.607)	(2.412)	(2.415)
Rate of missing	29.915	44.870	29.830	44.879
data in X in %	(2.010)	(2.301)	(2.067)	(2.156)
Full_X_Y : $\overline{MSPE} \times 10^3$	26.989	27.259	28.020	27.176
	(7.726)	(7.507)	(7.759)	(7.059)
$\overline{RT} \times 10$	10.072	10.069	10.078	10.068
	(0.063)	(0.060)	(0.063)	(0.062)
Reconst_X_Y : $\overline{MSPE} \times 10^3$	34.408	38.372	36.244	38.411
	(8.040)	(8.323)	(8.611)	(7.747)
$\overline{RT} \times 10$	10.089	10.098	10.098	10.095
	(0.078)	(0.081)	(0.077)	(0.083)
Reconst_X, Imp_Y : $\overline{MSPE} \times 10^3$	36.752	40.348	49.018	51.103
	(8.608)	(8.839)	(10.995)	(11.675)
$\overline{RT} \times 10$	10.094	10.105	10.123	10.127
	(0.081)	(0.082)	(0.088)	(0.094)
Reconst_X, Remov_Y : $\overline{MSPE} \times 10^3$	36.891	40.403	49.436	51.639
	(8.717)	(8.860)	(11.610)	(12.131)
$\overline{RT} \times 10$	10.095	10.105	10.124	10.127
	(0.082)	(0.083)	(0.087)	(0.094)
Remov_X_Y : $\overline{MSPE} \times 10^3$	39.619	47.940	56.390	69.935
	(10.117)	(14.026)	(16.332)	(21.684)
$\overline{RT} \times 10$	10.100	10.125	10.147	10.174
	(0.087)	(0.111)	(0.111)	(0.118)

Rate of missing	10.174	10.033	40.423	40.514
data in Y in %	(0.965)	(0.898)	(1.426)	(1.548)
Rate of missing	30.182	45.107	30.108	44.896
data in X in %	(1.232)	(1.459)	(1.285)	(1.290)
Full_X_Y : $\overline{MSPE} \times 10^3$	12.573	12.141	12.608	12.303
	(3.071)	(2.684)	(3.125)	(2.797)
$\overline{RT} \times 10$	10.034	10.031	10.033	10.034
	(0.026)	(0.027)	(0.027)	(0.029)
Reconst_X_Y : $\overline{MSPE} \times 10^3$	19.898	22.796	20.396	23.408
	(3.654)	(3.328)	(3.815)	(3.604)
$\overline{RT} \times 10$	10.052	10.060	10.053	10.057
	(0.038)	(0.044)	(0.041)	(0.045)
Reconst_X, Imp_Y : $\overline{MSPE} \times 10^3$	21.153	23.828	26.195	28.732
	(3.999)	(3.659)	(5.075)	(5.266)
$\overline{RT} \times 10$	10.053	10.063	10.067	10.068
	(0.040)	(0.044)	(0.045)	(0.049)
Reconst_X, Remov_Y : $\overline{MSPE} \times 10^3$	21.180	23.847	26.477	29.101
	(3.969)	(3.694)	(5.229)	(5.365)
$\overline{RT} \times 10$	10.054	10.063	10.068	10.069
	(0.040)	(0.044)	(0.046)	(0.049)
Remov_X_Y : $\overline{MSPE} \times 10^3$	18.892	22.105	25.893	31.192
	(4.792)	(5.681)	(6.873)	(8.474)
$\overline{RT} \times 10$	10.047	10.058	10.067	10.073
	(0.042)	(0.051)	(0.050)	(0.063)

Table 5: Mean and standard deviation errors for the predicted values based on 200 simulation replications with different levels of missing data and a sample size 500 (left panel) and a sample size 1300 (right panel). Partially observed curves are fully observed on $[1/50, 49/50]$ with SCENARIO 2.

Rate of missing	9.727	9.757	38.801	38.865
data in Y in %	(1.575)	(1.456)	(2.297)	(2.383)
Rate of missing	29.826	44.816	30.032	45.152
data in X in %	(2.069)	(2.192)	(2.253)	(2.288)
Full_X_Y : $\overline{MSPE} \times 10^6$	20.654	19.593	20.815	19.787
	(10.200)	(8.722)	(9.723)	(9.110)
$\overline{RT} \times 10$	10.158	10.143	10.194	10.153
	(0.219)	(0.210)	(0.234)	(0.202)
Reconst_X_Y : $\overline{MSPE} \times 10^6$	20.706	19.614	22.114	20.838
	(10.140)	(8.743)	(10.572)	(9.366)
$\overline{RT} \times 10$	10.158	10.142	10.208	10.162
	(0.219)	(0.211)	(0.241)	(0.209)
Reconst_X, Imp_Y : $\overline{MSPE} \times 10^6$	22.473	21.336	32.038	31.248
	(11.216)	(9.245)	(16.620)	(15.914)
$\overline{RT} \times 10$	10.170	10.159	10.291	10.240
	(0.229)	(0.228)	(0.298)	(0.254)
Reconst_X, Remov_Y : $\overline{MSPE} \times 10^6$	22.473	21.340	32.052	31.312
	(11.213)	(9.239)	(16.651)	(15.948)
$\overline{RT} \times 10$	10.171	10.159	10.291	10.241
	(0.229)	(0.228)	(0.298)	(0.256)
Remov_X_Y : $\overline{MSPE} \times 10^6$	30.380	38.660	46.638	58.802
	(13.412)	(19.485)	(21.421)	(30.753)
$\overline{RT} \times 10$	10.206	10.293	10.403	10.518
	(0.322)	(0.438)	(0.451)	(0.577)

Rate of missing	9.726	9.678	39.232	39.113
data in Y in %	(0.947)	(0.974)	(1.551)	(1.417)
Rate of missing	29.847	45.090	30.013	45.147
data in X in %	(1.290)	(1.365)	(1.203)	(1.425)
Full_X_Y : $\overline{MSPE} \times 10^6$	7.646	7.919	8.067	7.931
	(3.042)	(3.148)	(3.528)	(3.736)
$\overline{RT} \times 10$	10.062	10.065	10.065	10.055
	(0.085)	(0.086)	(0.079)	(0.082)
Reconst_X_Y : $\overline{MSPE} \times 10^6$	7.740	8.009	8.917	8.860
	(3.070)	(3.224)	(3.640)	(3.916)
$\overline{RT} \times 10$	10.062	10.066	10.072	10.063
	(0.085)	(0.088)	(0.082)	(0.087)
Reconst_X, Imp_Y : $\overline{MSPE} \times 10^6$	8.441	8.809	12.872	12.691
	(3.599)	(3.633)	(5.396)	(6.048)
$\overline{RT} \times 10$	10.067	10.072	10.103	10.098
	(0.091)	(0.096)	(0.101)	(0.121)
Reconst_X, Remov_Y : $\overline{MSPE} \times 10^6$	8.442	8.809	12.938	12.702
	(3.599)	(3.632)	(5.390)	(6.052)
$\overline{RT} \times 10$	10.067	10.072	10.104	10.098
	(0.009)	(0.010)	(0.010)	(0.012)
Remov_X_Y : $\overline{MSPE} \times 10^6$	12.175	15.342	17.406	23.095
	(5.109)	(6.675)	(7.836)	(10.587)
$\overline{RT} \times 10$	10.106	10.107	10.146	10.175
	(0.138)	(0.154)	(0.144)	(0.189)

Table 6: Mean and standard deviation errors for the predicted values based on 200 simulation replications with different levels of missing data and a sample size 500 (left panel) and a sample size 1300 (right panel). Partially observed curves are fully observed on $[5/50, 45/50]$ with SCENARIO 2.

Rate of missing	9.520	9.684	39.322	38.958
data in Y in %	(1.629)	(1.552)	(2.519)	(2.277)
Rate of missing	30.069	45.063	30.075	45.245
data in X in %	(2.096)	(2.232)	(2.106)	(2.213)
Full_X_Y : $\overline{MSPE} \times 10^6$	19.781	20.747	20.282	19.100
	(9.412)	(9.748)	(10.321)	(8.074)
$\overline{RT} \times 10$	10.169	10.170	10.176	10.160
	(0.217)	(0.229)	(0.222)	(0.217)
Reconst_X_Y : $\overline{MSPE} \times 10^6$	24.236	26.528	32.528	32.342
	(10.292)	(10.650)	(12.805)	(11.271)
$\overline{RT} \times 10$	10.198	10.204	10.274	10.271
	(0.222)	(0.240)	(0.261)	(0.258)
Reconst_X, Imp_Y : $\overline{MSPE} \times 10^6$	24.832	27.026	34.701	35.151
	(10.676)	(11.140)	(14.806)	(14.391)
$\overline{RT} \times 10$	10.209	10.203	10.288	10.281
	(0.235)	(0.249)	(0.295)	(0.281)
Reconst_X, Remov_Y : $\overline{MSPE} \times 10^6$	24.845	27.032	34.788	35.305
	(10.685)	(11.154)	(14.898)	(14.574)
$\overline{RT} \times 10$	10.209	10.205	10.289	10.281
	(0.235)	(0.249)	(0.295)	(0.282)
Remov_X_Y : $\overline{MSPE} \times 10^6$	29.948	39.318	44.115	56.488
	(13.803)	(18.900)	(20.894)	(31.158)
$\overline{RT} \times 10$	10.259	10.345	10.390	10.458
	(0.346)	(0.387)	(0.430)	(0.548)

Rate of missing	9.629	9.673	39.176	39.195
data in Y in %	(0.963)	(0.946)	(1.550)	(1.502)
Rate of missing	29.976	44.939	29.942	44.967
data in X in %	(1.284)	(1.488)	(1.325)	(1.392)
Full_X_Y : $\overline{MSPE} \times 10^6$	7.923	7.563	7.556	7.916
	(3.677)	(3.144)	(3.433)	(3.391)
$\overline{RT} \times 10$	10.056	10.065	10.063	10.067
	(0.082)	(0.079)	(0.087)	(0.078)
Reconst_X_Y : $\overline{MSPE} \times 10^6$	10.834	11.297	16.881	17.037
	(3.977)	(3.315)	(5.605)	(4.891)
$\overline{RT} \times 10$	10.082	10.096	10.132	10.139
	(0.088)	(0.086)	(0.117)	(0.109)
Reconst_X, Imp_Y : $\overline{MSPE} \times 10^6$	11.062	11.409	14.049	15.743
	(4.226)	(3.421)	(5.293)	(5.268)
$\overline{RT} \times 10$	10.082	10.097	10.118	10.127
	(0.094)	(0.091)	(0.107)	(0.117)
Reconst_X, Remov_Y : $\overline{MSPE} \times 10^6$	11.067	11.412	14.086	15.773
	(4.222)	(3.423)	(5.289)	(5.290)
$\overline{RT} \times 10$	10.082	10.098	10.119	10.128
	(0.094)	(0.091)	(0.108)	(0.118)
Remov_X_Y : $\overline{MSPE} \times 10^6$	12.702	15.481	16.638	22.266
	(5.567)	(7.721)	(7.140)	(11.556)
$\overline{RT} \times 10$	10.092	10.125	10.151	10.195
	(0.120)	(0.156)	(0.150)	(0.224)

6 Real dataset study: Hawaii Ocean data

The Hawaii ocean time-series program has been making repeated observations of various hydrographic, chemical and biological properties of the water column at a station north of Oahu, Hawaii since October, 1988. In the CTD dataset[1] of this program, various variables were measured every two meters between 0 and 200 meters below the sea surface. These variables are viewed as functions of depth. The layer between 0 and 200 meters below the sea surface is called the epipelagic zone (or sunlight zone), where enough light is available for photosynthesis. Therefore, in this zone, the primary production in the ocean occurs, and plants and animals are largely concentrated. The measurements were repeated at different dates, which we consider as different sample curves. We use three functional variables: *Temperature*, *Salinity* and *Oxygen*. Each curve being observed at 101 equally spaced points in [0, 200], we use the Oxygen as the response curve, expressed as a function of temperature, and we use the Salinity as the explanatory variable, also expressed as a function of temperature. We study the relationship between Salinity and Oxygen with the following model, written in the operatorial point of view:

$$Oxygen_i(Temperature) = \Big(\Theta \cdot Salinity_i\Big)(Temperature) + \varepsilon_i(Temperature),\ i = 1, \ldots, n, \text{ with } n = 191.$$

The graphical display of the initial sample of 191 pairs curves $\{Oxygen_i, Salinity_i\}_{i=1}^{191}$ (raw curves and smooth curves) can be observed in Figure 7. We can see that, on these data, all the curves are partially observed, which is a different situation compared to the simulations realized in the previous section, where only a percentage of curves were partially observed. In particular, the imputation method presented in Section 4 cannot be directly applied. We want to explore in this situation a possibility of a hybrid use of both reconstruction and imputation of the response curves.

We consider two scenarios for the response curves and for each scenario we use two samples: a training sample of size $\ell_1 = 153$ equal to 4/5 of the initial sample from which the estimates are computed and a testing sample of size $\ell_2 = 38$ equal to 1/5 of the initial sample on which the prediction errors are calculated. First, we reconstruct the missing part of all covariates using the observed parts. Then, as all the response curves are partially observed, we cannot directly use the regression imputation method from Section 4. As a consequence, we choose to reconstruct only a certain percentage of the response curves.

- 'Scenario 1': we reconstruct more response curves (87.958% of initial response curves which is equivalent to 168 response curves) using the observed parts while leaving curves with a low percentage of missing part.
- 'Scenario 2': we reconstruct less response curves (47.644% of initial response curves which is equivalent to 91 response curves) using the observed parts while leaving curves with a high percentage of missing part.

[1] https://hahana.soest.hawaii.edu/hot/hot-dogs/cextraction.html

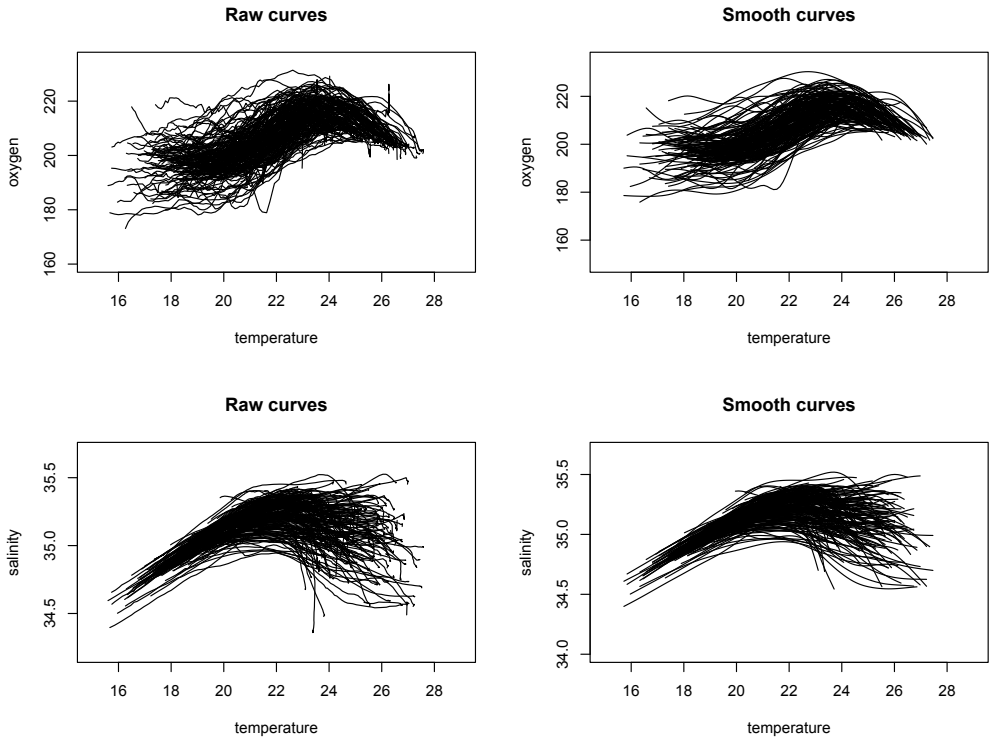

Figure 7: Partially Observed Functions: Raw curves of the oxygen variable (top left) and the salinity variable (bottom left). Smooth curves of the oxygen variable (top right) and the salinity variable (bottom right).

For 'Scenario 1', the training sample contains 23 partially observed response curves and 130 observed response curves and for 'Scenario 2' the training sample contains 100 partially observed response curves and 53 observed response curves. Moreover, for 'Scenario 1', 12.042% of response curves are affected by missing data. For 'Scenario 2', 52.356% of response curves are affected by missing data.

The graphical display of the training (resp. testing) samples of 153 (resp. 38) pairs of curves $\left\{Oxygen_i, Salinity_i\right\}_{i=1}^{153(38)}$ can be observed in Figures 8 and 9.

We consider the MSPE criterion as in Section 5 to evaluate the quality of prediction error. We also consider a MSPA (Mean Square Prediction Absolute error) criterion, defined in the same way, replacing the 2-norm with the 1-norm in the definition of MSPE. Table 7 gives MSPE and MSPA criteria for both 'Scenario 1' and 'Scenario 2'. We can observe that the prediction errors are much better in 'Scenario 1'. In other words, in such situations where all the response curves are partially observed, it seems better to reconstruct a more important part of them, and then use the imputation method on the remaining part.

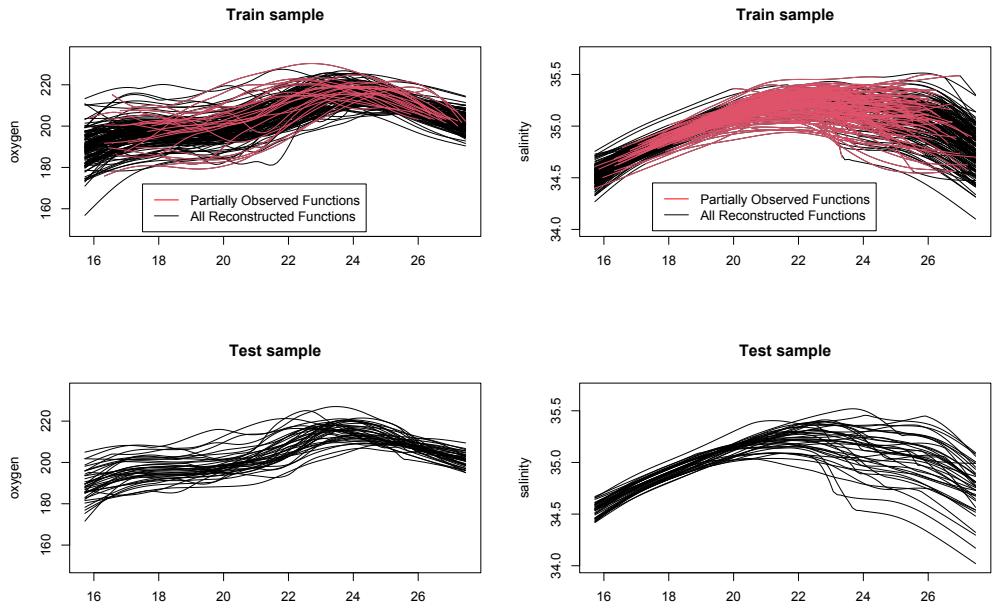

Figure 8: Training (resp. testing) samples of 153 (resp. 38) pairs of curves $\{Oxygen_i, Salinity_i\}_{i=1}^{153(38)}$ in 'Scenario 1'.

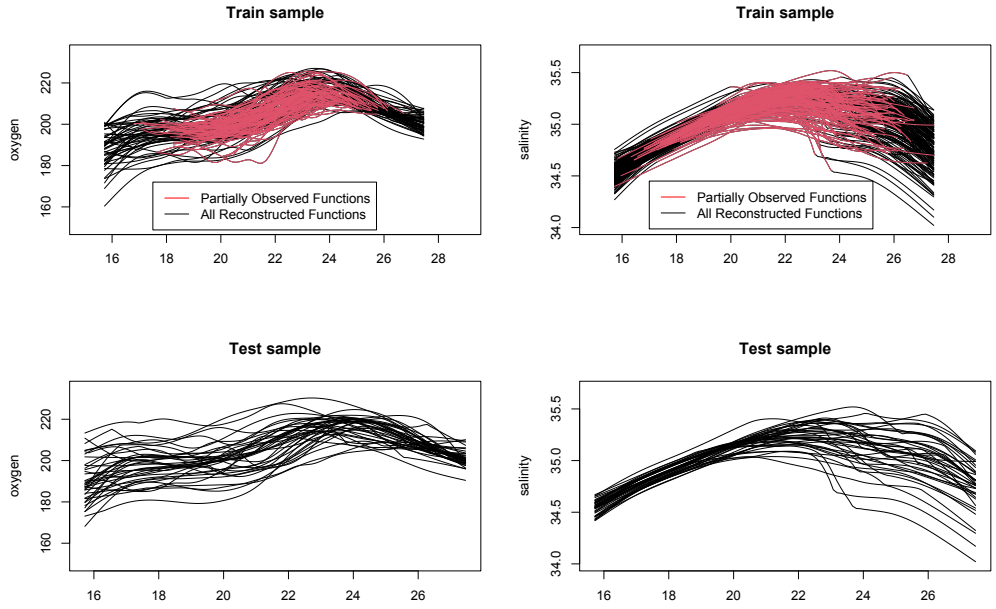

Figure 9: Training (resp. testing) samples of 153 (resp. 38) pairs of curves $\{Oxygen_i, Salinity_i\}_{i=1}^{153(38)}$ in 'Scenario 2'.

Table 7: Real dataset: prediction errors over one hazard sample. MSPE, MSPA and $k_n^\star$ for 'Scenario 1' and 'Scenario 2'.

Methods		'Scenario 1'	'Scenario 2'
Rate of partially observed response curves (%)		12.042	52.356
Rate of missing data in response curve (%)		Less than 27	More than 45
Reconst_X_Y	MSPE	23.566	46.351
	MSPA	3.855	5.144
	$k_n^\star$	13	8
Reconst_X, Imp_Y	MSPE	24.338	54.199
	MSPA	3.893	5.839
	$k_n^\star$	9	12

7 Proofs

7.1 Proof of Theorem 3.1

Starting with the reconstruction cross covariance operator,

$$\begin{aligned}
\widehat{\Delta}^\star_{n,rec} =& \frac{1}{n}\sum_{i=1}^{n} Y_i^\star \otimes X_i^\star \\
=& \frac{1}{n}\sum_{i=1}^{n} \Big(Y_i + (Y_i^\star - Y_i)\Big) \otimes X_i^\star, \\
=& \frac{1}{n}\sum_{i=1}^{n} Y_i \otimes X_i^\star + \frac{1}{n}\sum_{i=1}^{n} (Y_i^\star - Y_i) \otimes X_i^\star, \\
=& \Theta\widehat{\Gamma}_{n,rec} + \frac{1}{n}\sum_{i=1}^{n} \epsilon_i \otimes X_i^\star + \frac{1}{n}\sum_{i=1}^{n} (Y_i^\star - Y_i) \otimes X_i^\star.
\end{aligned}$$

Next, we obtain

$$
\begin{aligned}
&\mathbb{E}\left(\left\|\widehat{\Theta}^{\star}\cdot X^{\star}_{new} \;-\; \Theta\cdot X^{\star}_{new}\right\|^2\right)\\
&=\mathbb{E}\left(\left\|\widehat{\Pi}_{k_n,rec}\widehat{\Delta}^{\star}_{n,rec}\left(\widehat{\Pi}_{k_n,rec}\widehat{\Gamma}_{n,rec}\widehat{\Pi}_{k_n,rec}\right)^{-1}X^{\star}_{new}-\Theta\cdot X^{\star}_{new}\right\|^2\right)\\
&\leqslant 4\mathbb{E}\left(\left\|\widehat{\Pi}_{k_n,rec}\Theta\widehat{\Gamma}_{n,rec}\left(\widehat{\Pi}_{k_n,rec}\widehat{\Gamma}_{n,rec}\widehat{\Pi}_{k_n,rec}\right)^{-1}X^{\star}_{new}-\Theta\cdot X^{\star}_{new}\right\|^2\right)\\
&+4\mathbb{E}\left(\left\|\widehat{\Pi}_{k_n,rec}\left(\frac{1}{n}\sum_{i=1}^{n}\epsilon_i\otimes X_i^{\star}\right)\left(\widehat{\Pi}_{k_n,rec}\widehat{\Gamma}_{n,rec}\widehat{\Pi}_{k_n,rec}\right)^{-1}X^{\star}_{new}\right\|^2\right)\\
&+2\mathbb{E}\left(\left\|\widehat{\Pi}_{k_n,rec}\left(\frac{1}{n}\sum_{i=1}^{n}(Y_i^{\star}-Y_i)\otimes X_i^{\star}\right)\left(\widehat{\Pi}_{k_n,rec}\widehat{\Gamma}_{n,rec}\widehat{\Pi}_{k_n,rec}\right)^{-1}X^{\star}_{new}\right\|^2\right).
\end{aligned}
$$

Applying several times the identity $(a+b)^2 \leqslant 2a^2+2b^2$ for any $a,b\in\mathbb{R}$, we get

$$
\begin{aligned}
\mathbb{E}\left(\left\|\widehat{\Theta}^{\star}\cdot X^{\star}_{new} \;-\; \Theta\cdot X^{\star}_{new}\right\|^2\right) &\leqslant 64\mathbb{E}\left(\left\|\Theta\widehat{\Pi}_{k_n,rec}X^{\star}_{new}-\Theta\widehat{\Pi}_{k_n}X^{\star}_{new}\right\|^2\right)\\
&\quad+64\mathbb{E}\left(\left\|\Theta\widehat{\Pi}_{k_n}X^{\star}_{new}-\Theta\widehat{\Pi}_{k_n}X_{new}\right\|^2\right)\\
&\quad+32\mathbb{E}\left(\left\|\Theta\widehat{\Pi}_{k_n}X_{new}-\Theta\Pi_{k_n}X_{new}\right\|^2\right)\\
&\quad+16\mathbb{E}\left(\left\|\Theta\Pi_{k_n}X_{new}-\Theta X_{new}\right\|^2\right)\\
&\quad+8\mathbb{E}\left(\left\|\Theta X_{new}-\Theta X^{\star}_{new}\right\|^2\right)\\
&\quad+4\mathbb{E}\left(\left\|\frac{1}{n}\sum_{i=1}^{n}\langle X_i^{\star},\left(\widehat{\Pi}_{k_n,rec}\widehat{\Gamma}_{n,rec}\widehat{\Pi}_{k_n,rec}\right)^{-1}X^{\star}_{new}\rangle\epsilon_i\right\|^2\right)\\
&\quad+2\mathbb{E}\left(\left\|\frac{1}{n}\sum_{i=1}^{n}\langle X_i^{\star},\left(\widehat{\Pi}_{k_n,rec}\widehat{\Gamma}_{n,rec}\widehat{\Pi}_{k_n,rec}\right)^{-1}X^{\star}_{new}\rangle(Y_i^{\star}-Y_i)\right\|^2\right).
\end{aligned}
$$

Results of terms in the above decomposition are in Crambes et al. (2023), exceptionally the last term, let be noted by

$$
P_n=\frac{1}{n}\sum_{i=1}^{n}\langle X_i^{\star},\left(\widehat{\Pi}_{k_n,rec}\widehat{\Gamma}_{n,rec}\widehat{\Pi}_{k_n,rec}\right)^{-1}X^{\star}_{new}\rangle(Y_i^{\star}-Y_i)
$$

. Hence, using the Cauchy-Schwarz inequality, we have

$$
\mathbb{E}(\|P_n\|^2)\leqslant\sqrt{\mathbb{E}\left(\left\|\frac{1}{n}\sum_{i=1}^{n}\langle X_i^{\star},\left(\widehat{\Pi}_{k_n,rec}\widehat{\Gamma}_{n,rec}\widehat{\Pi}_{k_n,rec}\right)^{-1}X^{\star}_{new}\rangle\right\|^4\right)\mathbb{E}\left(\|Y_i^{\star}-Y_i\|^4\right)}.
$$

The result comes from Lemma 5.2 in Crambes and Henchiri (2019) and the result (3.4) that gives us

$$\begin{aligned}\mathbb{E}(\|P_n\|^2) &= \mathscr{o}\left(\frac{k_n}{n}\right) + \mathscr{O}\left(n^{-\zeta_1(b_O-1)/(b_O+2)}\right) \\ &= \mathscr{O}\left(n^{\eta_1/(a_O+2)-1-\zeta_1(b_O-1)/(b_O+2)}\right).\end{aligned}$$

Summarizing, we get

$$\mathbb{E}\left(\left\|\widehat{\Theta}^\star \cdot X^\star_{new} \;-\; \Theta \cdot X^\star_{new}\right\|^2\right) = \mathscr{O}\left(n^{-\eta_1(a_O-1)/(2(a_O+2))} + n^{\eta_1/(a_O+2)-1-\zeta_1(b_O-1)/(b_O+2)}\right).$$

References

Aguilera, A., F. Ocaña, and M. Valderrama (2008). Estimation of functional regression models for functional responses by wavelet approximation. In S. Dabo-Niang and F. Ferraty (Eds.), *In Functional and Operatorial Statistics: Contributions to Statistics*, Chapter 3, pp. 15–21. Physica-Verlag/Springer, Heidelberg.

Benatia, D., M. Carrasco, and J.-P. Florens (2017). Functional linear regression with functional response. *Journal of Econometrics 201*(2), 269–291.

Bosq, D. (2000). *Linear processes in function spaces: Theory and applications.* New York: Springer Verlag.

Brunel, E., A. Mas, and A. Roche (2016). Non-asymptotic adaptive prediction in functional linear models. *Journal of Multivariate Analysis 143*, 208–232.

Cai, T. and P. Hall (2006). Prediction in functional linear regression. *Annals of Statistics 34*, 2159–2179.

Cai, T. T. and M. Yuan (2012). Minimax and adaptive prediction for functional linear regression. *Journal of the American Statistical Association 107*(499), 1201–1216.

Cardot, H., F. Ferraty, and P. Sarda (2003). Spline estimators for the functional linear model. *Statistica Sinica 13*, 571–591.

Comte, F. and J. Johannes (2012). Adaptive functional linear regression. *The Annals of Statistics 40*(6), 2765–2797.

Crambes, C., C. Daayeb, A. Gannoun, and Y. Henchiri (2023). Functional linear model with partially observed covariate and missing values in the response. *Journal of Nonparametric Statistics 35*, 172–197.

Crambes, C. and Y. Henchiri (2019). Regression imputation in the functional linear model with missing values in the response. *Journal of Statistical Planning and Inference 201*, 103–119.

Crambes, C., N. Hilgert, and T. Manrique (2016). Estimation of the noise covariance operator in functional linear regression with functional outputs. *Statistics and Probability Letters 113*, 7–15.

Crambes, C., A. Kneip, and P. Sarda (2009). Smoothing splines estimators for functional linear regression. *The Annals of statistics 37*, 35–72.

Crambes, C. and A. Mas (2013). Asymptotics of prediction in functional linear regression with functional outputs. *Bernoulli 19*, 2627–2651.

Delaigle, A., P. Hall, W. Huang, and A. Kneip (2020). Estimating the covariance of fragmented and other related types of functional data. *Journal of the American Statistical Association 116*, 35–72.

Ferraty, F., I. Van Keilegom, and P. Vieu (2012). Regression when both response and predictor are functions. *Journal of Multivariate Analysis 109*, 10–28.

Ferraty, F. and P. Vieu (2006). *Nonparametric functional data analysis: Theory and practice*. New York: Springer Verlag.

Gellar, J. E., E. Colantuoni, D. M. Needham, and C. M. Crainiceanu (2014). Variable-domain functional regression for modeling ICU data. *Journal of the American Statistical Association 109*, 1425–1439.

Goldberg, Y., Y. Ritov, and A. Mandelbaum (2014). Predicting the continuation of a function with applications to call center data. *Journal of Statistical Planning and Inference 147*, 53–65.

Hall, P. and J. Horowitz (2007). Methodology and convergence rates for functional linear regression. *Annals of Statistics 35*, 70–91.

Harezlak, J., B. A. Coull, N. M. Laird, S. R. Magari, and D. C. . Christiani (2007). Penalized solutions to functional regression problems. *Computational Statistics and Data Analysis 51*, 4911–4925.

Hsing, T. and R. Eubank (2015). *Theoretical foundations of functional data analysis, with an introduction to linear operators*. John Wiley and Sons.

Imaizumi, M. and K. Kato (2018). PCA-based estimation for functional linear regression with functional responses. *Journal of Multivariate Analysis 163*, 15–36.

Kneip, A. and D. Liebl (2020). On the optimal reconstruction of partially observed functional data. *The Annals of Statistics 48*, 1692–1717.

Kokoszka, P. and M. Reimherr (2018). *Introduction to functional data analysis*. New York: Chapman and Hall.

Kraus, D. (2015). Components and completion of partially observed functional data. *Journal of the Royal Statistical Society: Series B 77*, 777–801.

Kraus, D. (2019). Inferential procedures for partially observed functional data. *Journal of Multivariate Analysis 173*, 583–603.

Kraus, D. and M. Stefanucci (2018). Classification of functional fragments by regularized linear classifiers with domain selection. *Biometrika 106*, 161–180.

Kraus, D. and M. Stefanucci (2020). Ridge reconstruction of partially observed functional data is asymptotically optimal. *Statistics and Probability Letters 165*, 1–5.

Li, Y. and T. Hsing (2007). On rates of convergence in functional linear regression. *Journal of Multivariate Analysis 98*, 1782–1804.

Lian, H. (2011). Convergence of functional k-nearest neighbor regression estimate with functional responses. *Electronic Journal of Statistics 5*, 31–40.

Lin, Z. and J.-L. Wang (2022). Mean and covariance estimation for functional snippets. *Journal of the American Statistical Association 117*, 348–360.

Lin, Z., J.-L. Wang, and Q. Zhong (2021). Basis expansions for functional snippets. *Biometrika 108*, 709–726.

Luo, R. and X. Qi (2017). Function-on-function linear regression by signal compression. *Journal of the American Statistical Association 112*(518), 690–705.

Park, J. and J. Qian (2012). Functional regression of continuous state distribution. *Journal of Econometrics 167*, 397–412.

Park, Y., X. Chen, and D. S. Simpson (2022). Robust inference for partially observed functional response data. *Statistica Sinica 32*, 1–29.

Park, Y. and D. G. Simpson (2019). Robust probabilistic classification applicable to irregularly sampled functional data. *Computational Statistics and Data Analysis 131*, 37–49.

Prchal, L. and P. Sarda. (2007). Spline estimator for functional linear regression with functional response. Unpublished, Preprint at https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.583.1816&rep=rep1&type=pdf.

Ramsay, J. O. and B. W. Silverman (2005). *Functional data analysis (Second edition)*. New York: Springer Verlag.

Sun, X., P. Du, X. Wang, and P. Ma (2018). Optimal penalized function-on-function regression under a Reproducing Kernel Hilbert Space framework. *Journal of the American Statistical Association 113*(524), 1601–1611.

Yao, F., H.-G. Müller, and J.-L. Wang (2005). Functional linear regression analysis for longitudinal data. *The Annals of Statistics 33*(6), 2873–2903.