Components and completion of partially observed functional data

David Kraus

Institute of Social and Preventive Medicine, University Hospital Lausanne, Switzerland

Summary. Functional data are traditionally assumed to be observed on the same domain. Motivated by a data set of heart rate temporal profiles, we develop methodology for the analysis of incomplete functional samples where each curve may be observed on a subset of the domain and unobserved elsewhere. We formalise this observation regime and develop the fundamental procedures of functional data analysis for this framework: estimation of parameters (mean, covariance operator) and principal component analysis. Principal scores of a partially observed function cannot be computed directly and we solve this challenging issue by estimating their best predictions as linear functionals of the observed part of the trajectory. Next, we propose a functional completion procedure that recovers the missing part using the observed part of the curve. We construct prediction intervals for principal scores and bands for missing parts of trajectories. The prediction problems are seen to be ill-posed inverse problems; regularisation techniques are used to obtain a stable solution. A simulation study shows the good performance of our methods. We illustrate the methods on the heart rate data, and provide practical computational algorithms and theoretical arguments and proofs of all results.

Keywords: Functional data analysis; Incomplete observation; Inverse problem; Prediction; Principal component analysis; Regularisation

1. Introduction

Contemporary data sets often consist of data units that are complex objects, such as functions, curves or images; see, for example, Ramsay and Silverman (2005), Ferraty and Vieu (2006), Ferraty and Romain (2011), Horváth and Kokoszka (2012). It is standard in the field of functional data analysis to assume that all functions are observed on the same domain. In this paper, we develop methods of analysis for functional data that are observed incompletely in the sense that each function might be observed only on a subset of the domain, whereas no information about the curve is available on the complement of this subset.

Our work is motivated by an ambulatory blood pressure monitoring data set that is part of the SKIPOGH project (Swiss Kidney Project on Genes in Hypertension, Pruijm et al., 2013) which is a multi-center cross-sectional study focussing on the role of kidney function and genes in blood pressure regulation and hypertension. In ambulatory blood pressure monitoring, participants wear a calibrated automatic device that is programmed to record systolic and diastolic blood pressure and heart rate at frequent intervals during 24 hours (every 15 minutes during the day and every 30 minutes during the night). Ideally, this design should provide enough information for each continuous temporal profile to be reconstructed by standard smoothing techniques; the resulting sample of curves would then be analysed by traditional methods of functional data analysis. In reality, however, some values have not been measured and the time points corresponding to unobserved values form series (intervals) of non-negligible length. There are two main reasons why there are

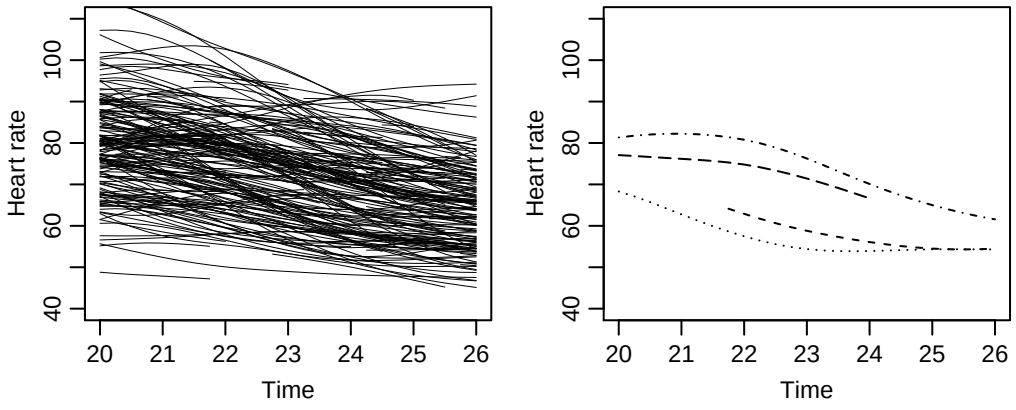

Fig. 1. A subset of the sample of heart rate profiles (left) and several curves in detail (right)

no available measurements for certain periods: first, it is the participant's discomfort (the participants can remove the device when they feel uncomfortable), and second, it is the failure of the device to take measurements. On the other hand, there are series of frequent, properly recorded measurements. It is therefore possible to reconstruct the underlying profiles in continuous time on these periods. The left panel in Figure 1 displays a subset of 685 heart rate profiles (values in beats per minute); we focus on the time interval $[20, 26]$ (i.e., from 8 pm of one day to 2 am next day) that is of particular medical interest because it is the transition period between the day and night regime. On the right panel in Figure 1, we plot separately four profiles to illustrate the type of available data: while some curves (dotted and dash-dot lines) are observed completely (on the entire domain $[20, 26]$), other curves (short-dashed and long-dashed lines) have unobserved periods. The percentage of incomplete functions is $31\,\%$ for blood pressure profiles and $44\,\%$ for heart rate profiles. This is a considerable fraction of the data, and we therefore wish to avoid removing the incomplete curves from the analysis.

The partial observation regime that we encounter in this data set is of general interest in applications as it is often the case that despite the failure to observe the curves in some regions, there is enough observed information in the rest of the domain. The mechanism that causes the absence of data can be random, like in our data, but the curves may also be partially observed by design. Moreover, data need not necessarily be curves indexed by time; methods that we develop can be extended to more general object data subject to incomplete observation, such as partially observed images, spatial curves or surfaces. Hence this kind of functional data is worth systematic investigation. Interestingly and surprisingly, this observation pattern, however natural and likely to occur in many applications it is, has received relatively little attention in the literature. James et al. (2000) and James and Hastie (2001) use parametric mixed effects models for principal components analysis and classification of partially observed curves. Bugni (2012) develops a goodness-of-fit test under circumstances similar to those of our paper. Delaigle and Hall (2013) deal with classification of functional data when only fragments of curves are available. Liebl (2013) studies low-rank extensions of curves observed on sub-domains. Goldberg et al. (2014) propose a prediction procedure for the continuation of a low-rank functional observation.

In this paper we introduce a formal framework for analysing incompletely observed functional

data and develop basic nonparametric, fully functional (infinite-dimensional) inferential procedures. When exploring functional data, one often finds interesting information in their covariance structure; see Ramsay and Silverman (2005) for a number of examples, and, e.g., Benko et al. (2009), Sangalli et al. (2009) or Panaretos et al. (2010) for other illustrations. Therefore, we first focus on the main building blocks of the analysis of the second-order properties: estimation of the covariance operator and principal component analysis. We propose an estimator of the covariance operator and its eigenvalues and eigenfunctions for partially observed functions and derive their properties. We deal with the estimation of projections (principal scores) of individual incomplete functions which is especially challenging. We develop a procedure that enables to predict the value of a principal score of a function when only a fragment of the function is available and direct computation is thus impossible. Next, we propose a method that can recover the unobserved part of the function from the observed part, using the information about the distribution of the data that it learns from the sample. We develop automatic procedures for the selection of the tuning parameter of the method that is based on generalised cross-validation for incompletely observed functions. We quantify the uncertainty of the predictions of unobserved quantities and provide approximate prediction regions (intervals and bands) covering the unobserved random quantity with high probability. Simulations confirm the usefulness and good performance of the proposed methodology.

Both the prediction of principal scores and reconstruction of an incomplete function or its derivatives are important problems. Principal scores are key elements in the exploration of complex data and can be used as input quantities in many inferential procedures. Their usefulness in the multivariate setting is well described, e.g., in the books by Krzanowski (2000) and Jolliffe (2002). In the functional context Ramsay and Silverman (2005) provide a number of real data examples illustrating how principal scores help to understand the properties of the data. Further applications can be found in Ramsay and Silverman (2002) and Ramsay et al. (2009). Horváth and Kokoszka (2012) give a comprehensive account of the utility of principal scores in procedures like two-sample tests, linear and non-linear regression, clustering and classification, time series analysis or change-point analysis. In this paper, we will see in Section 6 that the first three principal components of the heart rate profiles and their derivatives explain a large proportion of the total variability and are flexible enough to describe interesting features of the curves. Hence the corresponding scores provide an effectively reduced representation of the complex individual heart rate profiles. To perform graphical or formal analyses of the scores, one needs to be able to compute them which is not straightforward in the partial observation regime. Also, when an individual curve, surface or image is observed incompletely, one is interested in visualising and studying the shape of the missing part, for instance in order to forecast the continuation of the natural or social process described by the functional variable. Our paper provides solutions to these problems by developing methods that predict unobserved quantities via their conditional expectation given the observed data. In addition to their direct application to data, these methods will be an important tool in future research: for instance, advanced techniques of missing data analysis in the multivariate setting involve conditional expectations in some form, and our results will be helpful in extending them to the functional case.

To our knowledge, no results of the kind we provide here exist for functional data that are fully (densely in practice) observed on subsets of the domain. A related but different (in terms of applicability, used methods and achievable results) type of imperfectly observed functional data was studied by Yao et al. (2005a) who consider sparsely observed functions, i.e., situations where only a few observed values are available for each function, making it impossible to reconstruct each curve from these values. Our approach is novel in that it enables, under the assumed observation regime, to investigate some genuinely functional aspects of the data. From the theoretical point

of view, exploiting the continuous-time nature of the observed data, we are able to obtain stronger results than in the sparse regime. For example, the rates of convergence of estimators of parameters (covariance operator, eigenelements) are parametric, unlike with sparsely observed data (see also Hall et al., 2006). Also, the consistency result for our functional completion procedure is fully functional, while the restrictions of the sparse regime enabled Yao et al. (2005a) to achieve pointwise or finite dimensional convergence of the reconstructed trajectory. From the practical perspective, an important advantage of our method is that derivatives can be readily analysed in our setting while with methods for sparsely observed functions it is complicated. The method of Liu and Müller (2009) is a variant of that of Yao et al. (2005a) that can deal with derivatives in the sparse regime to some extent. While the method of Liu and Müller (2009) can reconstruct derivatives, it does not provide insight into their covariance structure because it neither estimates the covariance operator of the derivatives nor performs principal component analysis of the derivatives (it is based on derivatives of eigenfunctions rather than on eigenfunctions of derivatives). Since derivatives describe the dynamics of the underlying real world process, the analysis of derivatives, and especially of the principal sources of their variability, is often revealing in many applications, including the one we consider in this paper.

Mathematically, the problem that we need to solve for the computation of unobserved quantities (prediction of principal scores or reconstruction of missing parts of trajectories) is seen to be an ill-posed inverse problem (e.g., Groetsch, 1993), and regularisation techniques need to be applied. Such problems previously appeared in the literature on complete functional data mainly in the area of functional regression modelling; see, for example, Cardot et al. (1999), Müller and Stadtmüller (2005), Cai and Hall (2006), Cardot et al. (2007), Hall and Horowitz (2007) or He et al. (2010). Inverse problems similar to those we encounter here also arise in connection with functional canonical correlations (e.g., He et al., 2003) or with tests of hypotheses on parameters of functional data (e.g., Mas, 2007; Horváth et al., 2010; Aston and Kirch, 2012; Kraus and Panaretos, 2012; Horváth et al., 2013; Jarušková, 2013). Our problem is related to the task of prediction that was previously studied in the literature on functional time series; see, for example, Bosq (2000), Antoniadis and Sapatinas (2003) or Kargin and Onatski (2008). None of these papers, however, assumes the partial observation pattern that we consider in this paper.

The paper is organised as follows. In Section 2 we formalise the mechanism of partial observation of functional data and deal with the estimation of the mean function and covariance operator. Section 3 develops principal component analysis for incompletely observed functions. In Section 4, a method is proposed to reconstruct the missing part of a partially observed curve. Sections 5 and 6 present a simulation study and a data example. Appendix A contains proofs of the main theoretical results (Theorems 1 and 2). A supplementary document available online contains proofs of Propositions 1, 2, 3 and 4 and a detailed description of computational procedures. Computer programs can be obtained from

`http://wileyonlinelibrary.com/journal/rss-datasets`

2. Partially observed functional data

Functional data $X_1, \dots, X_n$ are seen as independent identically distributed random variables in the separable Hilbert space of square integrable functions on a bounded domain. Without loss of generality, we consider the space $L^2([0,1])$ with inner product $\langle f, g\rangle = \int_0^1 f(t)g(t)dt$, $f, g \in L^2([0,1])$ and norm $\|f\| = \langle f, f\rangle^{1/2}$. It is possible to extend our results to vector-valued functions or more general domains for applications with spatial curves, surfaces, images etc.

In traditional functional data analysis, it is assumed that the functions $X_1, \ldots, X_n$ are observed on the whole interval $[0, 1]$. We consider situations where each curve X_i is observed only on a subset of $[0, 1]$. Specifically, let the observation periods be $O_i \subset [0, 1]$, $i = 1, \ldots, n$. Then the observed data for the ith curve are $X_i(t)$, $t \in O_i$. (In practice, the raw data are most often in the form of possibly noisy observations on a dense grid of points in O_i, which enables us to assume that the curves are observed fully in O_i, as is explained by Hall et al. (2006).) We collectively denote the observed part of the curve as X_{iO_i}, which can be seen as a random element of the space $L^2(O_i)$. The values of X_i on the complement of O_i, $M_i = [0, 1] \setminus O_i$, are not observed; the missing part of the trajectory is denoted as X_{iM_i}. The observation periods O_i, $i = 1, \ldots, n$ are modelled as random subsets of $[0, 1]$. We assume that each realisation of O_i is the union of a finite number of intervals. This assumption is not restrictive for practical applications, although some generalisations are probably possible. We assume that the observation periods are independent of the functions $X_1, \ldots, X_n$, that is, the data are missing completely at random. (Under this assumption, the observation periods can also be seen as fixed when inference is made about the curves.)

The main characteristics of the distribution that generates the data are the mean function and the covariance operator. Let the mean function be $\mu = \mathsf{E}\, X_1$. The covariance operator $\mathscr{R} : L^2([0, 1]) \to L^2([0, 1])$ is defined as $\mathscr{R}f = \mathsf{E}\{\langle f, X_1 - \mu\rangle(X_1 - \mu)\} = \int_0^1 \rho(\cdot, t)f(t)dt$, where $\rho(s, t) = \operatorname{cov}\{X_1(s), X_1(t)\}$ is the covariance kernel of the stochastic process X_1.

Like in the multivariate case, the mean function μ at point $t \in [0, 1]$ can be estimated by the sample mean of observed values at this point. Formally, the estimator can be written as

$$\hat{\mu}(t) = \frac{J(t)}{\sum_{i=1}^n O_i(t)} \sum_{i=1}^n O_i(t)X_i(t),$$

where the notation $O_i(t)$ is used for the indicator $1_{O_i}(t)$ and $J(t) = 1_{[\sum_{i=1}^n O_i(t)>0]}$. The values of $X_i(t)$ are available only if $O_i(t) = 1$; otherwise, the contribution $O_i(t)X_i(t)$ in the sum above is zero. The term $J(t)$ is included to avoid division by zero: if $J(t) = 0$, the estimate of the mean is zero (or arbitrary, as such situations vanish asymptotically).

The estimator $\hat{\mathscr{R}}$ of the covariance operator $\mathscr{R}$ is defined through an estimator of its covariance kernel ρ. We estimate $\rho(s, t)$ by the sample covariance computed from all complete pairs of functional values at s and t. The estimator equals

$$\hat{\rho}(s, t) = \frac{I(s, t)}{\sum_{i=1}^n U_i(s, t)} \sum_{i=1}^n U_i(s, t)(X_i(s) - \hat{\mu}_{st}(s))(X_i(t) - \hat{\mu}_{st}(t)), \tag{1}$$

where $U_i(s, t) = O_i(s)O_i(t)$ and $I(s, t) = 1_{[\sum_{i=1}^n U_i(s,t)>0]}$. The estimator of the mean function used here is

$$\hat{\mu}_{st}(s) = \frac{I(s, t)}{\sum_{i=1}^n U_i(s, t)} \sum_{i=1}^n U_i(s, t)X_i(s),$$

i.e., for the computation of the covariance at s, t, functional values are centred at the sample mean computed from complete pairs. (It is also possible to centre by the estimator $\hat{\mu}$ introduced before; all results remain valid when $\hat{\mu}$ is used in place of $\hat{\mu}_{st}$.)

The sample covariance operator computed from incomplete functions may be indefinite. This phenomenon is similar to the multivariate setting. However, unlike with multivariate data, our experience in the functional context is that this problem is practically unimportant because negative eigenvalues occur far in the tail of the spectrum and are of small magnitude in comparison with the

leading eigenvalues. The corresponding high-frequency features of the data are practically never of interest. If needed, the estimate $\hat{\mathscr{R}}$ can be modified by setting negative eigenvalues equal to zero.

It is seen that $\hat{\mu}(t)$ is an unbiased estimator of $\mu(t)$. Similarly, if we subtract 1 in the denominator of $\hat{\rho}(s,t)$, the estimator becomes unbiased for $\rho(s,t)$. For the estimators $\hat{\mu}$ and $\hat{\mathscr{R}}$ to be consistent, we need to assume that the observation pattern asymptotically provides enough information. For the mean function, the right assumption is that

$$\text{there exists } \delta_1 > 0 \text{ such that } \sup_{t\in[0,1]} P\bigg(n^{-1}\sum_{i=1}^{n} O_i(t) \le \delta_1\bigg) = O(n^{-2}) \text{ as } n \to \infty. \tag{2}$$

Similarly, for the covariance operator, we need the stronger assumption that

$$\text{there exists } \delta_2 > 0 \text{ such that } \sup_{(s,t)\in[0,1]^2} P\bigg(n^{-1}\sum_{i=1}^{n} U_i(s,t) \le \delta_2\bigg) = O(n^{-2}) \text{ as } n \to \infty. \tag{3}$$

Assumption (2) is satisfied, for example, when the observation sets $O_1, \dots, O_n$ are independent and identically distributed and $\pi_0 = \inf_{t\in[0,1]} P(O_1(t) = 1) > 0$. To see this, set $\delta_1 = \pi_0/2$ and use Hoeffding's inequality to show that

$$\sup_{t\in[0,1]} P\bigg(n^{-1}\sum_{i=1}^{n} O_i(t) \le \delta_1\bigg) \le \exp(-\pi_0^2 n/2).$$

Analogously, assumption (3) is satisfied when we further assume that $\inf_{(s,t)\in[0,1]^2} P(U_1(s,t) = 1) > 0$. Under these weak assumptions, we obtain a consistency result as follows.

PROPOSITION 1.

(a) Let $\mathsf{E}\,\|X_1\|^2 < \infty$ *and assumption* (2) *be satisfied. Then* $\mathsf{E}\,\|\hat{\mu} - \mu\|^2 = O(n^{-1})$ *for* $n \to \infty$.

(b) Let $\mathsf{E}\,\|X_1\|^4 < \infty$ *and assumption* (3) *be satisfied. Then* $\mathsf{E}\,\|\hat{\mathscr{R}} - \mathscr{R}\|_2^2 = O(n^{-1})$ *for* $n \to \infty$ *(here* $\|\cdot\|_2$ *denotes the Hilbert–Schmidt norm).*

Notice that the properties of the estimators are unaffected by the fact that the functions are observed only partially. The full (dense) observation regime, albeit only on subsets of the domain, preserves the convergence rates known for complete functional data (see Bosq (2000) or Horváth and Kokoszka (2012) for results in the traditional setting).

3. Principal component analysis

3.1. Estimation of eigenfunctions and eigenvalues

Probably the most fundamental method for functional data is functional principal component analysis. It provides insight into the complex covariance structure of functional data, is used to identify main sources of variability and quantify their importance and to reduce the dimension of the data.

The theoretical foundation of functional principal component analysis is the Karhunen–Loève theorem (e.g., Bosq, 2000, Theorem 1.5) stating that there exist random variables β_{ij} and nonrandom functions φ_j such that the stochastic process X_i admits the decomposition

$$X_i(t) = \mu(t) + \sum_{j=1}^{\infty} \beta_{ij}\varphi_j(t), \quad t \in [0,1],$$

where the series converges in mean square, uniformly in t. Here φ_j, $j = 1, 2, \dots$ are the orthonormal eigenfunctions of the operator $\mathscr{R}$ and β_{ij}, $j = 1, 2, \dots$ are uncorrelated mean zero variables with variances λ_j, where $\lambda_1 \geq \lambda_2 \geq \cdots > 0$ are the eigenvalues of $\mathscr{R}$. Functional principal component analysis is the empirical version of the Karhunen–Loève expansion that aims to estimate the elements involved in the expansion. For background information on this classical topic, we refer to Ramsay and Silverman (2005, Chapter 8) for an introduction from an applied perspective, and to Dauxois et al. (1982), Bosq (2000) or Hall and Hosseini-Nasab (2006) for theoretical studies.

In the case of completely observed functional data, to estimate the eigenvalues λ_j and eigenfunctions φ_j, one performs eigen-decomposition of the usual sample covariance operator. When the functions are observed partially, one can proceed similarly and define the estimators $\hat{\lambda}_j$ and $\hat{\varphi}_j$ as the eigenvalues and eigenfunctions of the operator $\hat{\mathscr{R}}$ given by the kernel $\hat{\rho}$ in (1).

It turns out that the asymptotic properties of the empirical eigenvalues and eigenfunctions remain unchanged by the incompleteness of the observed functions. The following proposition shows that first, the empirical eigenvalues are consistent estimators of the true eigenvalues and this consistency is uniform over all indices, and second, the empirical eigenfunctions are consistent estimators of the true eigenfunctions, up to the usual sign ambiguity.

PROPOSITION 2. *Let* $\mathsf{E}\,\|X_1\|^4 < \infty$ *and assumption* (3) *be satisfied. Then* $\mathsf{E}\sup_{j\in\mathbb{N}}\{|\hat{\lambda}_j - \lambda_j|^2\} = O(n^{-1})$. *If moreover all eigenvalues of* $\mathscr{R}$ *have multiplicity* 1, *then* $\mathsf{E}\,\|\hat{\varphi}_j - \hat{s}_j\varphi_j\|^2 = O(n^{-1})$ *for all* $j \in \mathbb{N}$, *where* $\hat{s}_j = \operatorname{sign}\langle\hat{\varphi}_j, \varphi_j\rangle$.

The rates of convergence are parametric due to the full observation regime on subsets; the situation is different from that of sparsely observed functions, where the estimators of eigenelements (constructed differently) converge at nonparametric rates (Yao et al., 2005a; Hall et al., 2006).

3.2. Estimation of principal component scores

In principal component analysis, one is usually interested not only in estimating the eigenfunctions and eigenvalues but also in the estimation of the principal component scores

$$\beta_{ij} = \langle X_i - \mu, \varphi_j\rangle, \quad i = 1, \dots, n,\ j = 1, 2, \dots$$

representing the individual coordinates of each curve with respect to the eigenbasis (the expression of the feature φ_j for the ith observation). The leading principal scores provide the optimal finite dimensional representation of each curve and can be further analysed by traditional techniques.

In the standard situation of complete functional data, the scores are easily estimated by $\hat{\beta}_{ij} = \langle X_i - \hat{\mu}, \hat{\varphi}_j\rangle$. When the functional observations are incomplete, the direct computation of $\langle X_i - \hat{\mu}, \hat{\varphi}_j\rangle$ is impossible because the last term in the expression

$$\langle X_i - \hat{\mu}, \hat{\varphi}_j\rangle = \langle X_{iO_i} - \hat{\mu}_{O_i}, \hat{\varphi}_{jO_i}\rangle + \langle X_{iM_i} - \hat{\mu}_{M_i}, \hat{\varphi}_{jM_i}\rangle$$

is not available. In the equation above the subscript O_i or M_i denotes the restriction of the corresponding function to the ith observed or missing period, respectively. We develop a procedure to estimate the missing quantity $\langle X_{iM_i} - \hat{\mu}_{M_i}, \hat{\varphi}_{jM_i}\rangle$ from the observed data.

First, we consider the population version of the problem. Let the function X with mean zero and covariance operator $\mathscr{R}$ be observed on the set O and missing on M. For the purpose of the following considerations, the sets O and M, which are independent of X, can be regarded as nonrandom, equivalently, derivations can be made conditionally on them. The goal is to predict

$\beta_{jM} = \langle X_M, \varphi_{jM} \rangle$ from the observed part X_O. It is a standard fact that in terms of the mean squared prediction error, the best approximation of β_{jM} by a functional of X_O is the conditional expectation $\mathsf{E}(\beta_{jM}|X_O)$. The conditional expectation may be a nonlinear functional of the condition and thus difficult to estimate. Therefore, we propose to look for the best linear prediction corresponding to a continuous linear functional of the observed curve. This is equivalent to the best linear approximation of the conditional expectation. By the Riesz representation theorem, a continuous linear functional takes the form $\langle a_j, X_O \rangle$, where a_j is an element of $L^2(O)$. The best continuous linear prediction of β_{jM} equals $\tilde{\beta}_{jM} = \langle \tilde{a}_j, X_O \rangle$, where $\tilde{a}_j$ solves the infinite-dimensional optimisation problem

$$\min_{a_j \in L^2(O)} \mathsf{E}\{(\beta_{jM} - \langle a_j, X_O \rangle)^2\}. \tag{4}$$

The objective functional can be rewritten as

$$\begin{aligned} \mathsf{E}\{(\beta_{jM} - \langle a_j, X_O \rangle)^2\} &= \mathsf{E}\{\langle \varphi_{jM}, X_M \rangle^2 - 2\langle \varphi_{jM}, X_M \rangle \langle a_j, X_O \rangle + \langle a_j, X_O \rangle^2\} \\ &= \langle \varphi_{jM}, \mathscr{R}_{MM} \varphi_{jM} \rangle - 2\langle \varphi_{jM}, \mathscr{R}_{MO} a_j \rangle + \langle a_j, \mathscr{R}_{OO} a_j \rangle, \end{aligned}$$

where $\mathscr{R}_{OO}$ is the covariance operator of X_O and $\mathscr{R}_{MO}$ is the cross-covariance operator of X_M and X_O. It is obvious that the objective functional is convex. If a minimiser exists, it can be found by setting the derivative equal to zero. The derivatives in this context are in the Fréchet sense. In particular, we see that $\frac{\partial}{\partial a_j} \mathsf{E}\{(\beta_{jM} - \langle a_j, X_O \rangle)^2\} = -2r_j + 2\mathscr{R}_{OO} a_j$, where $r_j = \mathscr{R}_{OM} \varphi_{jM}$ with $\mathscr{R}_{OM} = \mathscr{R}_{MO}^*$ (the star * denotes the adjoint operator). Thus we need to solve the equation

$$\mathscr{R}_{OO} a_j = r_j. \tag{5}$$

We recognize that this is a linear inverse problem where we need to recover the function $a_j \in L^2(O)$ from its image through the linear operator $\mathscr{R}_{OO}$.

Let λ_{OOk}, $k = 1, 2, \ldots$ be the decreasing positive eigenvalues and φ_{OOk} the corresponding orthonormal eigenfunctions of the operator $\mathscr{R}_{OO}$. By comparing the coefficients of the left and right hand side of (5) with respect to the basis φ_{OOk}, we arrive at the system of equations $\lambda_{OOk} \langle a_j, \varphi_{OOk} \rangle = \langle r_j, \varphi_{OOk} \rangle$, $k = 1, 2, \ldots$ This suggests that a candidate for the solution be

$$\tilde{a}_j = \sum_{k=1}^{\infty} \frac{\langle r_j, \varphi_{OOk} \rangle}{\lambda_{OOk}} \varphi_{OOk}, \tag{6}$$

that is, $\tilde{a}_j = \mathscr{R}_{OO}^{-1} r_j$. This is a valid solution, if it is an element of $L^2(O)$, that is, if

$$\sum_{k=1}^{\infty} \frac{\langle r_j, \varphi_{OOk} \rangle^2}{\lambda_{OOk}^2} < \infty. \tag{7}$$

This condition is known in the theory of inverse problems as Picard's condition. A solution to the inverse problem (5) exists if and only if condition (7) is satisfied.

Condition (7) is equivalent to the condition

$$\sum_{k=1}^{\infty} \frac{\operatorname{cor}(\beta_{jM}, \langle X_O, \varphi_{OOk} \rangle)^2}{\operatorname{var}(\langle X_O, \varphi_{OOk} \rangle)} < \infty, \tag{8}$$

which has a clear interpretation. It states that the missing variable β_{jM} must not be strongly correlated with complicated, high-frequency components of the observed function. The variability of these components must be sufficiently large to provide enough information for the prediction of β_{jM}. The precise balance between the complexity of the correlation of the unobserved score with the predictor components and the variability of the predictor components is quantified by the requirement on the series above to converge.

In the Gaussian case, the conditional expectation of β_{jM} given the principal scores $\langle X_O, \varphi_{OOk}\rangle$, $k = 1, 2, \dots$ is an infinite linear combination of these scores (an almost surely convergent infinite series). One can show this by conditioning on finitely many components (this multivariate conditional expectation is linear) and applying Lévy's zero-one law (Kallenberg, 2002, Theorem 7.23) to obtain the limit. The infinite sum of variances of terms in this series converges, which is equivalent to the convergence of $\sum_{k=1}^{\infty} \frac{\langle r_j, \varphi_{OOk}\rangle^2}{\lambda_{OOk}}$ or $\sum_{k=1}^{\infty} \operatorname{cor}(\beta_{jM}, \langle X_O, \varphi_{OOk}\rangle)^2$. If, moreover, condition (7) or (8) is satisfied, then the coefficients in the infinite linear combination for the conditional expectation form an ℓ^2 sequence, hence the conditional expectation is continuous in the condition.

From now on, to guarantee the existence of a continuous solution to (5), we assume that condition (7) holds. If it is a priori known that the conditional expectation $\mathsf{E}(\beta_{jM}|X_O)$ is a continuous linear functional of X_O, then condition (7) is automatically satisfied.

The operator $\mathscr{R}_{OO}$ is a compact operator with infinite-dimensional range, therefore, its inverse $\mathscr{R}_{OO}^{-1}$ is not bounded (i.e., not continuous). Consequently, small perturbations of r_j may lead to large perturbations of $\tilde{a}_j = \mathscr{R}_{OO}^{-1} r_j$. It is seen from equation (6) that an overall small change of r_j may result in an arbitrarily large change of $\tilde{a}_j$, if the change of r_j occurs on a coefficient with a sufficiently high index k; the division by a sufficiently small eigenvalue may enormously magnify the perturbation. In other words, the solution $\tilde{a}_j = \mathscr{R}_{OO}^{-1} r_j$ is extremely unstable and the inverse problem in (5) is ill-posed. It is important for a solution to be stable with respect to perturbations of the right-hand side r_j because r_j is unknown and needs to be estimated. With estimated right-hand side, the solution to the inverse problem may be arbitrarily far from the true one no matter how accurate the estimate is. This is true even when $\mathscr{R}_{OO}$ is known. Moreover, the operator $\mathscr{R}_{OO}$ is not known either; its estimate has finite rank and therefore is not invertible in $L^2(O)$.

To obtain a stable solution, one needs to use regularisation, that is, to modify the ill-posed inverse problem in such a way that it becomes well-posed with a stable solution. We use ridge regularisation. Instead of (5), we solve the problem $\mathscr{R}_{OO}^{(\alpha)} a_j = r_j$ with $\mathscr{R}_{OO}^{(\alpha)} = \mathscr{R}_{OO} + \alpha \mathscr{I}_O$, where $\alpha > 0$ and $\mathscr{I}_O$ is the identity operator on $L^2(O)$. The inverse $\mathscr{R}_{OO}^{(\alpha)-1}$ of the bounded operator $\mathscr{R}_{OO}^{(\alpha)}$ is bounded and therefore the solution $\tilde{a}_j^{(\alpha)} = \mathscr{R}_{OO}^{(\alpha)-1} r_j$ is stable. Denote the regularised best linear prediction of β_{jM} by $\tilde{\beta}_{jM}^{(\alpha)} = \langle \tilde{a}_j^{(\alpha)}, X_O\rangle$. The stability of the solution increases with α but the bias of the solution increases too because the problem becomes more different from the original problem; conversely, with α decreasing, the solution gets closer to the exact but unstable solution of the original problem.

We now turn to the practical, empirical version of the problem of computation of principal scores from partially observed functional data. We have a sample of n functions $X_{1O_1}, \dots, X_{nO_n}$ observed on the sets $O_1, \dots, O_n$. The mean function μ and the covariance operator $\mathscr{R}$ are estimated by $\hat{\mu}$ and $\hat{\mathscr{R}}$ introduced in Section 2. The principal score of the ith curve with respect to the jth eigenfunction is estimated by $\hat{\beta}_{ij}^{(\alpha)} = \hat{\beta}_{ijO_i} + \hat{\beta}_{ijM_i}^{(\alpha)}$, where $\hat{\beta}_{ijO_i} = \langle X_{iO_i} - \hat{\mu}_{O_i}, \hat{\varphi}_{jO_i}\rangle$ and $\hat{\beta}_{ijM_i}^{(\alpha)} = \langle \hat{a}_{ij}^{(\alpha)}, X_{iO_i} - \hat{\mu}_{O_i}\rangle$. Here the function $\hat{a}_{ij}^{(\alpha)} = \hat{\mathscr{R}}_{O_iO_i}^{(\alpha)-1} \hat{r}_{ij}$ solves the empirical regularised inverse problem $\hat{\mathscr{R}}_{O_iO_i}^{(\alpha)} a_{ij} = \hat{r}_{ij}$, where $\hat{\mathscr{R}}_{O_iO_i}^{(\alpha)} = \hat{\mathscr{R}}_{O_iO_i} + \alpha \mathscr{I}_{O_i}$ with $\hat{\mathscr{R}}_{O_iO_i}$ being an integral

operator on $L^2(O_i)$ with kernel equal to the restriction of the kernel $\hat{\rho}$ of $\hat{\mathscr{R}}$ (see (1)) to $O_i \times O_i$, and $\hat{r}_{ij} = \hat{\mathscr{R}}_{O_i M_i} \hat{\varphi}_{jM_i}$ with $\hat{\mathscr{R}}_{O_i M_i}$ defined analogously by restriction of $\hat{\rho}$ to $O_i \times M_i$.

We are ready to state the main convergence result that justifies this method. The difference between the regularised estimator $\hat{\beta}^{(\alpha)}_{ijM_i}$ and the best linear prediction $\tilde{\beta}_{ijM_i}$ can be decomposed into the sum of the estimation error for the regularised prediction and the approximation error due to regularisation, i.e., $\hat{\beta}^{(\alpha)}_{ijM_i} - \tilde{\beta}_{ijM_i} = (\hat{\beta}^{(\alpha)}_{ijM_i} - \tilde{\beta}^{(\alpha)}_{ijM_i}) + (\tilde{\beta}^{(\alpha)}_{ijM_i} - \tilde{\beta}_{ijM_i})$. We show that when the amount of regularisation decreases at a suitable rate as the sample size increases, both terms converge to zero in $L^2(P)$ and thus the regularised estimator of the prediction is consistent.

THEOREM 1. *Let* $\mathsf{E}\,\|X_1\|^4 < \infty$*, assumption* (3) *be satisfied, all eigenvalues of* $\mathscr{R}$ *have multiplicity* 1 *and condition* (7) *be satisfied for* O_i *and* M_i *in place of* O *and* M*, respectively. Then*

$$\mathsf{E}\{(\hat{\beta}^{(\alpha)}_{ijM_i} - \tilde{\beta}_{ijM_i})^2\} \le O(\alpha^{-3})O(n^{-1}) + O(\alpha)$$

as $\alpha \to 0$ *and* $n \to \infty$*. Hence, if* $\alpha = \alpha_n$ *such that* $\alpha_n \to 0$ *and* $\alpha_n n^{1/3} \to \infty$ *as* $n \to \infty$*, then* $\hat{\beta}^{(\alpha_n)}_{ijM_i}$ *is a consistent estimator of the best linear prediction* $\tilde{\beta}_{ijM_i}$ *of* β_{ijM_i}*.*

Sometimes one is interested in estimating other linear functionals than the principal score $\langle X_i - \mu, \varphi_j \rangle$. Our consistency results remain valid when $\hat{\varphi}_{jO_i}$ is replaced by an arbitrary random or fixed function $\hat{f}_{O_i} \in L^2(O_i)$ such that $\mathsf{E}\,\|\hat{f}_{O_i} - f_{O_i}\|^2 = O(n^{-1})$ for some deterministic $f_{O_i} \in L^2(O_i)$.

Notice that the theorem has no strong assumptions. Picard's condition (7) is a basic assumption that is required in all inverse problems to guarantee the existence of a solution. Except this standard requirement, no other condition on the rate of decrease of the eigenvalues $\lambda_{O_i O_i k}$ is needed. This is due to the fact that we estimate the prediction $\langle \tilde{a}_{ij}, X_{iO_i} \rangle$ rather than the prediction functional $\tilde{a}_{ij}$ itself. Intuitively, the integration in $\langle \tilde{a}_{ij}, X_{iO_i} \rangle$ brings additional smoothness; the exact way this happens is seen in the proof of the theorem. In a related context of prediction in functional linear regression, it was observed by Cai and Hall (2006) and Cardot et al. (2007) that weaker assumptions are needed and stronger results can be obtained when the focus is on prediction rather than on the estimation of the regression functional. The inverse problem is similar to the one solved in the functional linear model (Cardot et al., 1999, 2007; Hall and Horowitz, 2007). However, the way we arrive at it differs from the functional linear model because, for instance, due to the incompleteness of observations there is no collection of response–covariate pairs in the present situation.

As an alternative to ridge regularisation, one may consider the spectral truncation approach. Both methods have their advantages and disadvantages. For instance, it is known that the behaviour of spectral cut-off methods depends on the spacings between the eigenvalues of the operator to be inverted which makes them less robust with respect to situations with similar or even identical eigenvalues (see Hall and Horowitz, 2007). Indeed, in a preliminary analysis of our motivating dataset we observed some very similar estimated eigenvalues. There is also an important computational advantage of the ridge method. For this method, one only needs to solve a linear equation with $\hat{\mathscr{R}}^{(\alpha)}_{O_i O_i}$ which is very easy and fast. On the other hand, the spectral truncation approach requires computing the eigen-decomposition of $\hat{\mathscr{R}}_{O_i O_i}$ and projecting on the corresponding subspace. This is computationally more demanding, especially since it must be done repeatedly for each function because different suboperators $\hat{\mathscr{R}}_{O_i O_i}$ of $\hat{\mathscr{R}}$ corresponding to different functions have different spectral decompositions. Yet another approach may be based on smoothing, for instance, by penalising the roughness of the solution of the inverse problem.

3.3. Regularisation parameter selection

Theorem 1 shows that for an appropriate choice of α_n, the estimator $\hat{\beta}_{ijM_i}^{(\alpha_n)}$ is consistent for the best prediction $\tilde{\beta}_{ijM_i}$. The theorem however does not give a practical recommendation on how to select the regularisation parameter. It is desirable to have an automatic, data-driven selection procedure.

Since the parameter α is difficult to understand, we first translate it into more comprehensible values. By analogy with ridge regression or various standard smoothing techniques, we define the number of effective degrees of freedom as the trace of the covariance of the predictors composed with its regularised inverse, i.e.,

$$\mathrm{df}_i(\alpha) = \mathrm{trace}(\hat{\mathscr{R}}_{O_iO_i}^{(\alpha)-1}\hat{\mathscr{R}}_{O_iO_i}) = \sum_{k=1}^{\infty} \frac{\hat{\lambda}_{O_iO_ik}}{\hat{\lambda}_{O_iO_ik}+\alpha}, \tag{9}$$

which is a decreasing function of α. Unlike in standard situations the covariance operator here is computed from partially observed data. Another way to measure the amount of regularisation is the proportion of retained variability like in classical principal component analysis using, e.g.,

$$\frac{\mathrm{trace}(\hat{\mathscr{R}}_{O_iO_i}\hat{\mathscr{R}}_{O_iO_i}^{(\alpha)-1}\hat{\mathscr{R}}_{O_iO_i}\hat{\mathscr{R}}_{O_iO_i}^{(\alpha)-1}\hat{\mathscr{R}}_{O_iO_i})}{\mathrm{trace}\,\hat{\mathscr{R}}_{O_iO_i}} = \frac{\sum_{k=1}^{\infty}\frac{\hat{\lambda}_{O_iO_ik}^3}{(\hat{\lambda}_{O_iO_ik}+\alpha)^2}}{\sum_{k=1}^{\infty}\hat{\lambda}_{O_iO_ik}} \tag{10}$$

or a similar quantity. One can determine α such that the effective degrees of freedom equal some value or the proportion of retained variability exceeds some threshold. These quantities however do not measure the predictive performance of the regularised solution.

A universal recipe for situations of this type is to use generalised cross-validation. In traditional settings, the generalised cross-validation score is the residual sum of squares (a measure of goodness of fit) divided by a decreasing function of the effective degrees of freedom (a penalty included to avoid under-regularisation). The residual sum of squares is the sum of squared differences of the response variables and their predictions, which in our case are $\hat{\beta}_{kjM_i} = \langle X_{kM_i} - \hat{\mu}_{M_i}, \hat{\varphi}_{jM_i}\rangle$ and $\hat{\beta}_{kjM_i}^{(\alpha)} = \langle \hat{a}_{ij}^{(\alpha)}, X_{kO_i} - \hat{\mu}_{O_i}\rangle$, $k = 1, \ldots, n$, respectively. In the situation of partially observed functions, the pair of the response variable $\hat{\beta}_{kjM_i}$ and the explanatory variable X_{kO_i} is not available for all individuals $k = 1, \ldots, n$. The idea is, therefore, to consider the set of completely observed functions with indices $C = \{k : 1 \le k \le n,\ \int_0^1 O_k(t)dt = 1\}$. If this set is reasonably large, one can compute the residual sum of squares over the complete functions

$$\mathrm{rss}_{ij}(\alpha) = \sum_{k\in C}(\hat{\beta}_{kjM_i} - \hat{\beta}_{kjM_i}^{(\alpha)})^2.$$

The cross-validation score for the regularised estimation of the jth score of the ith function is

$$\mathrm{gcv}_{ij}(\alpha) = \frac{\mathrm{rss}_{ij}(\alpha)}{\{1 - \frac{1}{|C|}\mathrm{df}_i(\alpha)\}^2},$$

where $|C|$ is the number of complete functions. One selects the value of α that minimises this quantity. Separate values of the regularisation parameter are used for each function and each score.

3.4. Prediction uncertainty

For a statistical procedure to be useful, it is important to quantify its uncertainty, i.e., to assess how far $\hat{\beta}_{ijM_i}^{(\alpha_n)}$ can be from β_{ijM_i}. The following proposition answers these questions.

PROPOSITION 3. *Let the assumptions of Theorem 1 be satisfied and let* $\alpha_n \to 0$ *and* $\alpha_n n^{1/4} \to \infty$ *as* $n \to \infty$. *Then* $\hat{\beta}_{ijM_i}^{(\alpha_n)} - \beta_{ijM_i}$ *is asymptotically distributed as* $\tilde{\beta}_{ijM_i} - \beta_{ijM_i}$, *which is a zero-mean random variable with variance that can be consistently estimated by*

$$\hat{v}_{ij}^2 = \langle \hat{\varphi}_{jM_i}, (\hat{\mathscr{R}}_{M_iM_i} - \hat{\mathscr{R}}_{M_iO_i}\hat{\mathscr{R}}_{O_iO_i}^{(\alpha_n)-1}\hat{\mathscr{R}}_{O_iO_i}\hat{\mathscr{R}}_{O_iO_i}^{(\alpha_n)-1}\hat{\mathscr{R}}_{O_iM_i})\hat{\varphi}_{jM_i} \rangle.$$

If the distribution of the data is Gaussian, then the limiting variable is Gaussian.

The assumptions of this proposition are similar to those of the consistency result of Theorem 1, except that a slower rate of convergence of the regularisation parameter to zero is needed to consistently estimate the limiting variance.

The prediction uncertainty, as expressed by the variance $\hat{v}_{ij}^2$, does not converge to zero as the sample size converges to infinity. This is due to the fact that the situation is a prediction problem rather than an estimation problem in the sense that we try to recover a random variable rather than a non-random parameter. Thus, while increasing the sample size eventually removes the uncertainty due to unknown, estimated quantities (mean function and covariance operator) and regularisation, there is a fundamental uncertainty that cannot be removed asymptotically. In other words, the knowledge of the principal score will never be precise, if the functional observation is incomplete, and the limits of accuracy of the prediction are given by the asymptotic variance v_{ij}^2. We refer to Didericksen et al. (2012) for an interesting discussion of similar questions in somewhat related prediction problems in the context of functional time series.

Proposition 3 immediately enables us to construct a prediction interval for the score. Assume that a Gaussian distribution is a good approximation for the distribution of the data. Then

$$I_{ij;\eta} = (\hat{\beta}_{ij}^{(\alpha_n)} - z_{1-\eta/2}\hat{v}_{ij}, \hat{\beta}_{ij}^{(\alpha_n)} + z_{1-\eta/2}\hat{v}_{ij}), \tag{11}$$

where $z_{1-\eta/2}$ is the $(1-\eta/2)$-quantile of the standard normal distribution, is a prediction interval for β_{ij} with asymptotic coverage probability $1-\eta$, i.e., $P(\beta_{ij} \in I_{ij;\eta}) \to 1-\eta$ as $n \to \infty$.

Since principal component analysis is often used as a dimension reduction procedure and the resulting principal scores are subsequently analysed by traditional techniques, it is useful to have a measure of reliability of the computed scores. The true score β_{ij} is a random variable with variance estimated by $\hat{\lambda}_j$. The predicted score $\hat{\beta}_{ij}^{(\alpha_n)}$ can be seen as the true score contaminated by error with variance estimated by $\hat{v}_{ij}^2$. One can define the relative error

$$\hat{v}_{ij}/\hat{\lambda}_j^{1/2}, \tag{12}$$

which is the ratio of the error variability and the natural, intrinsic variability of the score. This value, lying between 0 and 1, can be used as an indicator of observations that are too uncertain, and the scores whose relative error exceeds a certain threshold (for example, 0.2) can be excluded from the subsequent analysis. The uncertainty will be high when the association between the missing part of the score and the observed fragment is weak.

The high uncertainty of predictions due to a small amount of observed information is one example of situations where one must be cautious. Another such case could be when missingness is very frequent in certain regions or the overlap of observation periods is not frequent enough because then the precision of the estimation of the covariance function will be locally reduced, and consequently the prediction procedure may be less accurate. The performance of generalised cross-validation may also be negatively influenced. Yet another problem could arise when the data are

not missing at random (e.g., when missingness is more likely to occur when functional values are high). In such cases, missing functional chunks may be indeed very insidious because important features of the data distribution may be lost. Furthermore, the presence of functional outliers can be a complication as they may be more difficult to detect when only fragments are available.

4. Functional completion

4.1. Reconstruction of incomplete functions

It is natural to ask whether it is possible to recover not only the missing part of a principal score (and thus compute the score of an incomplete function) like in Section 3 but also the whole missing part of the trajectory (and thus reconstruct the whole functional variable). The answer is positive.

In the population version of the problem, the best prediction of X_M by a function of X_O in the sense of the mean integrated prediction squared error is the conditional expectation $\mathsf{E}(X_M|X_O)$. It is in general a nonlinear operator from $L^2(O)$ to $L^2(M)$ and similarly to the case of principal scores, we consider its best continuous linear approximation. Assuming for simplicity that the functional variable has mean zero, the minimisation problem to be solved is

$$\min_{\mathscr{A}:\|\mathscr{A}\|_\infty<\infty} \mathsf{E}\,\|X_M - \mathscr{A}X_O\|^2,$$

where the solution is looked for in the class of continuous (bounded) linear operators from $L^2(O)$ to $L^2(M)$ (by $\|\cdot\|_\infty$ we denote the operator norm). We see (by Fréchet differentiation or direct computation) that solving this minimisation is equivalent to solving the (normal) equation $\mathscr{A}\mathscr{R}_{OO} = \mathscr{R}_{MO}$. This suggests the solution $\tilde{\mathscr{A}} = \mathscr{R}_{MO}\mathscr{R}_{OO}^{-1}$ and the best linear prediction of X_M in the form $\tilde{X}_M = \tilde{\mathscr{A}}X_O$. From now on, we assume the existence of a bounded solution, that is, we assume that $\|\mathscr{R}_{MO}\mathscr{R}_{OO}^{-1}\|_\infty < \infty$. Similarly to the case of principal scores, the inverse problem to be solved is ill-posed. Using ridge regularisation we obtain the solution $\tilde{\mathscr{A}}^{(\alpha)} = \mathscr{R}_{MO}\mathscr{R}_{OO}^{(\alpha)-1}$. The regularised best linear prediction equals $\tilde{X}_M^{(\alpha)} = \tilde{\mathscr{A}}^{(\alpha)}X_O$.

Practically, when the sample $X_{1O_1}, \dots, X_{nO_n}$ is observed on the subsets $O_1, \dots, O_n$, we replace the covariance operator by its estimate and set $\hat{\mathscr{A}}_i^{(\alpha)} = \hat{\mathscr{R}}_{M_iO_i}\hat{\mathscr{R}}_{O_iO_i}^{(\alpha)-1}$. The mean function needs to be estimated as well. For the ith curve, the best linear prediction of X_{iM_i} is estimated by

$$\hat{X}_{iM_i}^{(\alpha)} = \hat{\mu}_{M_i} + \hat{\mathscr{A}}_i^{(\alpha)}(X_{iO_i} - \hat{\mu}_{O_i}).$$

To prove the consistency, we assume not only that the solution to the inverse problem (the prediction operator) is bounded but that it is Hilbert–Schmidt. We have a result as follows.

THEOREM 2. *Let* $\mathsf{E}\,\|X_1\|^4 < \infty$, *assumption* (3) *be satisfied and* $\|\mathscr{R}_{M_iO_i}\mathscr{R}_{O_iO_i}^{-1}\|_2 < \infty$. *Then*

$$\mathsf{E}\{\|\hat{X}_{iM_i}^{(\alpha)} - \tilde{X}_{iM_i}\|^2\} \le O(\alpha^{-3})O(n^{-1}) + O(\alpha)$$

as $\alpha \to 0$ *and* $n \to \infty$. *Hence, if* $\alpha = \alpha_n$ *such that* $\alpha_n \to 0$ *and* $\alpha_n n^{1/3} \to \infty$ *as* $n \to \infty$, *then* $\hat{X}_{iM_i}^{(\alpha_n)}$ *is a consistent estimator of the best linear prediction* $\tilde{X}_{iM_i}$ *of* X_{iM_i}.

Notice that our consistency result is genuinely functional. It is different from Theorem 3 of Yao et al. (2005a) where it was only possible to obtain a pointwise consistent estimator of the functional variable. The reason is that we assume that the functions are observed fully (or densely in practice)

on subsets of the domain while Yao et al. (2005a) work in a sparse observation regime. In other words, we are able to achieve stronger results because our data contain more information.

The assumption that the prediction operator $\tilde{\mathscr{A}_i} = \mathscr{R}_{M_iO_i}\mathscr{R}_{O_iO_i}^{-1}$ is Hilbert–Schmidt ($\|\tilde{\mathscr{A}_i}\|_2 < \infty$) needed for the proof is a strengthening of the basic assumption on the continuity of $\tilde{\mathscr{A}_i}$ ($\|\tilde{\mathscr{A}_i}\|_\infty < \infty$). Assumptions of this type were used in related contexts of, e.g., prediction in functional time series (Bosq (2000, Chapter 8), Kargin and Onatski (2008)) and functional linear model (Yao et al., 2005b; He et al., 2010). It seems possible to replace this assumption by a combination of the condition $\|\tilde{\mathscr{A}_i}\|_\infty < \infty$ and a condition on the eigenvalue sequence $\lambda_{O_iO_ik}$ such that the regularisation error can be controlled.

The condition $\|\tilde{\mathscr{A}_i}\|_2 < \infty$ can be written explicitly in terms of the covariance structure of the principal scores of the observed and unobserved part of the function. If the eigen-decompositions of $\mathscr{R}_{O_iO_i}$ and $\mathscr{R}_{M_iM_i}$ are

$$\mathscr{R}_{O_iO_i} = \sum_{k=1}^{\infty} \lambda_{O_iO_ik}\varphi_{O_iO_ik} \otimes \varphi_{O_iO_ik}, \qquad \mathscr{R}_{M_iM_i} = \sum_{k=1}^{\infty} \lambda_{M_iM_ik}\varphi_{M_iM_ik} \otimes \varphi_{M_iM_ik}$$

(where $\otimes$ stands for the tensor product, $(f \otimes g)u = \langle g, u\rangle f$), then we can write

$$\mathscr{R}_{M_iO_i} = \sum_{j=1}^{\infty}\sum_{k=1}^{\infty} \gamma_{M_iO_ijk}\varphi_{M_iM_ij} \otimes \varphi_{O_iO_ik},$$

where $\gamma_{M_iO_ijk} = \langle \varphi_{M_iM_ij}, \mathscr{R}_{M_iO_i}\varphi_{O_iO_ik}\rangle = \mathrm{cov}(\langle X_{M_i} - \mu_{M_i}, \varphi_{M_iM_ij}\rangle, \langle X_{O_i} - \mu_{O_i}, \varphi_{O_iO_ik}\rangle)$. Then the operator $\tilde{\mathscr{A}_i}$ is Hilbert–Schmidt whenever

$$\sum_{j=1}^{\infty}\sum_{k=1}^{\infty} \frac{\gamma^2_{M_iO_ijk}}{\lambda^2_{O_iO_ik}} < \infty,$$

which is equivalent to

$$\sum_{j=1}^{\infty} \lambda_{M_iM_ij} \sum_{k=1}^{\infty} \frac{\mathrm{cor}(\langle X_{M_i} - \mu_{M_i}, \varphi_{M_iM_ij}\rangle, \langle X_{O_i} - \mu_{O_i}, \varphi_{O_iO_ik}\rangle)^2}{\lambda_{O_iO_ik}} < \infty.$$

It is seen that this condition combines conditions for the prediction of $\langle X_{M_i} - \mu_{M_i}, \varphi_{M_iM_ij}\rangle$, $j = 1, 2, \ldots$ (compare the inner series above with condition (8)).

4.2. Selection of the regularisation parameter

To understand the amount of regularisation corresponding to α, one can use the effective degrees of freedom or the proportion of retained variability as defined in (9) and (10), respectively. For the selection of α automatically balancing the stability and accuracy of the prediction of X_{iM_i}, we propose a similar cross-validation procedure as in Subsection 3.3 for principal scores. The residual sum of squares for the prediction of trajectories on M_i computed for the completely observed curves in the sample is

$$\mathrm{rss}_i(\alpha) = \sum_{k\in C} \|X_{kM_i} - \hat{X}^{(\alpha)}_{kM_i}\|^2.$$

The value of α used for the prediction of a function on M_i from its observation on O_i is minimises

$$\mathrm{gcv}_i(\alpha) = \frac{\mathrm{rss}_i(\alpha)}{\{1 - \frac{1}{|C|}\,\mathrm{df}_i(\alpha)\}^2}.$$

4.3. Uncertainty and prediction bands

Theorem 2 shows that $\hat{X}_{iM_i}^{(\alpha_n)}$ consistently estimates the best linear prediction $\tilde{X}_{iM_i}$. We are now interested in the variation of $\hat{X}_{iM_i}^{(\alpha_n)}$ around the target quantity, the unobserved function X_{iM_i}.

PROPOSITION 4. *Let the assumptions of Theorem 2 be satisfied and let $\alpha_n \to 0$ and $\alpha_n n^{1/4} \to \infty$ as $n \to \infty$. Then $\hat{X}_{iM_i}^{(\alpha_n)} - X_{iM_i}$ is asymptotically distributed (in the sense of weak convergence of probability measures on $L^2([0,1])$) as the mean zero stochastic process $\tilde{X}_{iM_i} - X_{iM_i}$. The limiting covariance operator is consistently estimated (with respect to the Hilbert–Schmidt norm) by*

$$\hat{\mathscr{V}}_i = \hat{\mathscr{R}}_{M_iM_i} - \hat{\mathscr{R}}_{M_iO_i}\hat{\mathscr{R}}_{O_iO_i}^{(\alpha_n)-1}\hat{\mathscr{R}}_{O_iO_i}\hat{\mathscr{R}}_{O_iO_i}^{(\alpha_n)-1}\hat{\mathscr{R}}_{O_iM_i}.$$

If the data are Gaussian, then the limiting stochastic process is Gaussian.

The trace of $\hat{\mathscr{V}}_i$ quantifies the total amount of uncertainty of the linear prediction of X_{iM_i}. It approaches zero as the Lebesgue measure of the missing region M_i approaches zero, i.e., as we approach a completely observed function. When the measure of the observation period O_i converges to zero, the total prediction uncertainty converges to the trace of $\hat{\mathscr{R}}$, which corresponds to the situation of no information about the ith curve. The scale invariant ratio

$$(\operatorname{trace}\hat{\mathscr{V}}_i)^{1/2}/(\operatorname{trace}\hat{\mathscr{R}})^{1/2} \tag{13}$$

measures the relative prediction error, i.e., the amount of uncertainty about the ith curve as a proportion of the total spread of the distribution of the functional random variable. One minus this value corresponds to the reduction of uncertainty achieved by the best linear prediction and can be seen as a measure of performance of the completion procedure. Alternatively, we can use $\hat{\mathscr{R}}_{M_iM_i}$ instead of $\hat{\mathscr{R}}$ in the denominator in the relative prediction error, leading to the ratio of the uncertainty about the missing trajectory when the prediction method is used versus the uncertainty there would be about X_{iM_i} if we ignored the observed part.

We use the asymptotic distribution of $\hat{X}_{iM_i}^{(\alpha_n)} - X_{iM_i}$ for the construction of prediction bands for the unobserved part of the trajectory, that is, regions containing the curve X_{iM_i} with high probability. We consider bands of the form

$$\{(t,x) : \hat{X}_{iM_i}^{(\alpha_n)}(t) - c_{1-\eta}\hat{h}(t) \le x \le \hat{X}_{iM_i}^{(\alpha_n)}(t) + c_{1-\eta}\hat{h}(t),\ t \in M_i\}, \tag{14}$$

where $\hat{h}$ is a function that consistently estimates some limiting function h that is bounded away from zero, and $c_{1-\eta}$ is the $(1-\eta)$-quantile of the random variable $\sup_{t\in M_i}|\tilde{X}_{iM_i}(t) - X_{iM_i}(t)|/h(t)$. This band has asymptotic coverage $1-\eta$. One can choose $\hat{h} = 1$, leading to a band with constant width, but typically one prefers a band whose width at time t reflects the uncertainty of the prediction of the missing function at t. We use $\hat{h}(t) = \max(\hat{h}_0, \hat{v}_i(t))$ where $\hat{v}_i(t)$ is the estimated standard deviation of the limiting predictive distribution at time t, i.e., the square root of the diagonal of the kernel of $\hat{\mathscr{V}}_i$, and $\hat{h}_0$ is a threshold guaranteeing that the limiting function h is bounded away from 0. For example, the choice $\hat{h}_0 = 0.2\sup_{t\in M_i}\hat{v}_i(t)$ works well in practice. If the distribution of the data can be considered as Gaussian, the quantile $c_{1-\eta}$ can be computed by simulation as follows. One generates a large number of independent realisations of the Gaussian process with mean zero and covariance operator $\hat{\mathscr{V}}_i$, divides them by $\hat{h}(t)$, computes the maxima of their absolute values and determines the $(1-\eta)$-quantile of this sample. The simulation of the trajectories and the computation of the maxima is performed on a fine grid of points. Note that the

width of the band does not converge to zero because it is a prediction band, i.e., it must contain, with high probability, a random function.

We conlude this section by a theoretical remark. While the proposed prediction bands work well in practice, as is documented in the simulation study in Section 5, for a strictly rigorous justification arguments based on Proposition 4 (which is a consequence of Theorem 2) need to be extended. Proposition 4 guarantees the convergence in distribution in the sense of the topology of the L^2-norm of the Hilbert space $L^2([0,1])$. This justifies the construction of prediction regions in the form of balls in $L^2([0,1])$ which, however, are not practical because they cannot be plotted. For prediction bands, the convergence is needed in the sense of the uniform topology. To this end, one needs to leave the geometric world of $L^2([0,1])$ and switch to the space of continuous functions $C([0,1])$. Under modified assumptions (which would include conditions on sample paths, such as Hölder continuity), it seems possible to prove the convergence in the uniform topology. We do not pursue this theoretical study but give arguments indirectly justifying the use of the bands. Suppose that the asymptotic approximation suggested by Theorem 2 and Proposition 4 is considered applicable if the L^2-distance from the limiting variable is sufficiently small. The probability of this L^2-distance exceeding some $\varepsilon > 0$ is, in light of Chebyshev's inequality, bounded as follows, $P(\|\hat{X}_{iM_i}^{(\alpha_n)} - \tilde{X}_{iM_i}\|_2^2 > \varepsilon) \leq \varepsilon^{-2}\,\mathsf{E}\,\|\hat{X}_{iM_i}^{(\alpha_n)} - \tilde{X}_{iM_i}\|_2^2$. On the other hand, the convergence in the L^2-norm does not imply the uniform convergence because large deviations may occur on a small set of arguments. Let us compute the Lebesgue measure γ of the set where $|\hat{X}_{iM_i}^{(\alpha_n)} - \tilde{X}_{iM_i}|$ deviates more than ε from zero. We compute $\gamma(\{t : |\hat{X}_{iM_i}^{(\alpha_n)}(t) - \tilde{X}_{iM_i}(t)| > \varepsilon\}) \leq \varepsilon^{-2}\|\hat{X}_{iM_i}^{(\alpha_n)} - \tilde{X}_{iM_i}\|_2^2$ using Chebyshev's inequality. Taking expectations on both sides, we obtain on the right side the same bound as before. Hence, if the bound is considered to be sufficiently small for the asymptotic approximation in the L^2-norm to be applicable, then also the expected Lebesgue measure of the set of large pointwise deviations will be negligible.

5. Simulations

A simulation study was designed to address the following goals: to investigate the performance of generalised cross-validation as a selector of the regularisation parameter, to verify the validity and accuracy of the prediction intervals and bands and to explore the effect of the observation pattern.

We generate random samples of curves of the form

$$X(t) = \sum_{k=1}^{100} \nu_k^{1/2} \xi_k 2^{1/2} \cos(2\pi k t) + \sum_{k=1}^{100} \omega_k^{1/2} \eta_k 2^{1/2} \sin(2\pi k t), \quad t \in [0,1],$$

where ξ_k, η_k are independent standard normal variables and the eigenvalues are of the form $\nu_k = 3^{-(2k-1)}$, $\omega_k = 3^{-2k}$. The three most important components represent 67, 22 and 7 percent of the total variability. For each curve we generate independently a random period on which this curve is not observed. The functional values on this period are removed. For the ith function, the missing period M_i is simulated in the form $M_i = [C_i - E_i, C_i + E_i] \cap [0,1]$ with $C_i = dU_{i,1}^{1/2}$, $E_i = fU_{i,2}$, where d, f are parameters and $U_{i,1}$, $U_{i,2}$ are independent variables uniformly distributed on $[0,1]$. The performance of our procedures is measured on one curve in the sample, say X_1. For this curve, we use a fixed (nonrandom) missing period to guarantee that values computed from different simulation runs have the same meaning. In all simulations, we use $L = 1000$ repetitions.

For the first two sets of simulations, we set $d = 1.4$, $f = 0.2$. This leads to an observation pattern with similar characteristics as in our motivating data set. The cross-sectional probability of

Table 1. Performance of the generalised cross-validation selection procedure (MSPE and the variability of the target quantity are multiplied by 1000)

Target quantity (and its variability)	n	*MSPE for* $\alpha = c\alpha_{\mathrm{gcv}}$ c					*Median df for* $\alpha = \alpha_{\mathrm{gcv}}$
		0.04	0.2	1	5	25	
Score 1	100	1.91	1.55	1.32	1.61	3.78	7.68
(333)	500	0.60	0.44	0.36	0.42	1.07	12.73
Score 2	100	0.46	0.37	0.35	0.44	0.80	8.61
(111)	500	0.16	0.13	0.12	0.15	0.27	13.71
Score 3	100	1.45	1.13	0.95	1.08	2.00	8.62
(37)	500	0.48	0.34	0.28	0.29	0.53	13.71
Missing trajectory	100	10.07	7.90	6.95	8.24	15.16	7.98
(500)	500	4.04	2.79	2.24	2.30	3.48	15.02

observation ranges from $99\,\%$ at time 0 to $79\,\%$ at time 1. The percentage of complete curves is $39\,\%$. The median length of the missing period (given the curve has a missing period) is 0.15. For the curve X_1, on which the performance is measured, we set $M_1 = (0.4, 0.7)$.

First, we investigate the performance of generalised cross-validation based on complete curves. As a measure of quality of the prediction of a missing quantity, we use the mean squared prediction error (MSPE) which is the average over all simulation runs of the squared distances of the predicted value and the true value, i.e., $L^{-1}\sum_{l=1}^{L}(\hat{\beta}_{1jM_1}^{(\alpha)[l]} - \hat{\beta}_{1jM_1}^{[l]})^2$ for the jth score and $L^{-1}\sum_{l=1}^{L}\|\hat{X}_{1M_1}^{(\alpha)[l]} - X_{1M_1}^{[l]}\|^2$ for the missing part of the trajectory, where the superscript $^{[l]}$ indicates that the value pertains to the lth generated sample. Table 1 shows values of the mean squared prediction error for the first three principal scores and for the missing part of the trajectory. The table also includes the variability of the target quantities (i.e., the true eigenvalues for the scores and the trace of the true covariance operator $\mathscr{R}$ for the trajectory) to put the values into context. The mean squared prediction error is reported for α set to the value selected by generalised cross-validation and to values slightly smaller or bigger in the form of multiples of the selected value. We see that the method successfully approximates the best value of α and can be recommended as the tuning parameter selector. The accuracy increases with increasing sample size n; however, it should be noted that the mean squared prediction error cannot converge to zero because there is always some uncertainty due to the randomness of the target quantity, as discussed in Subsections 3.4 and 4.3. The last column of the table reports the median of the effective degrees of freedom corresponding to the selected value of α. It is seen that in all cases the typical number of degrees of freedom is in a reasonable relation to the sample size.

The second set of simulations explores the properties of the approximate distribution of the deviation of the prediction from the predicted quantity established in Propositions 3 and 4. We simulate from the same distribution and observation pattern as before. The regularisation parameter is selected by generalised cross-validation. We consider prediction intervals and bands of the form (11) and (14), respectively, with nominal coverage $95\,\%$. We compute bands with both constant and variable width, as discussed in Subsection 4.3. Empirical coverage probabilities (i.e., the percentage of cases when the unobserved quantity was covered by the constructed region) are reported in Table 2. We see that the proposed intervals and bands have coverage that is close to the nominal coverage and, therefore, provide useful information on the probable values of the scores or the missing trajectory. Table 2 also reports the median of relative error measures (12) and (13). For instance, we can see that the approximate distribution is relatively more spread for less variable

Table 2. Empirical coverage (in %) of prediction regions (intervals for scores, bands with constant and variable width for curves) and the median relative error measure

n	*Score 1*		*Score 2*		*Score 3*	
	Coverage	*Med. rel. err.*	*Coverage*	*Med. rel. err.*	*Coverage*	*Med. rel. err.*
100	97.2	0.073	95.2	0.056	94.5	0.143
500	97.4	0.042	95.0	0.036	96.3	0.092

	Missing trajectory		
	Coverage (const. wid.)	*Coverage (var. wid.)*	*Med. rel. err.*
100	94.3	96.7	0.123
500	94.2	98.4	0.071

Table 3. Standardised mean squared prediction error for different observation patterns

n	*Observ. pattern (X_1)*	*Score 1*		*Score 2*		*Score 3*		*Missing trajectory*	
		Observ. pattern (sample)		*Observ. pattern (sample)*		*Observ. pattern (sample)*		*Observ. pattern (sample)*	
		A	*B*	*A*	*B*	*A*	*B*	*A*	*B*
100	*I*	0.022	0.045	0.052	0.093	0.035	0.067	0.040	0.075
	II	0.039	0.073	0.078	0.128	0.107	0.155	0.076	0.136
500	*I*	0.006	0.013	0.018	0.031	0.010	0.023	0.013	0.024
	II	0.019	0.027	0.037	0.051	0.060	0.076	0.035	0.051

(higher index) scores. This is in line with conclusions from Table 1 where we observed a similar relation between MSPE and the variability of the target quantity. Hence the relative error measures of (12) and (13), which can be computed from the data, seem to be valuable indicators of the accuracy of the reconstruction procedure.

In the last set of simulations, we study the effect of the observation pattern on the the accuracy of our methods. We vary the amount of observed information both for X_1 (whose characteristics are to be reconstructed) and for the whole sample (which is used to learn the reconstruction procedure). Two settings are used for the missing period of X_1: (I) $M_1 = (0.4, 0.7)$, (II) $M_1 = (0.4, 0.9)$. For the simulation of the missing periods of other curves in the sample, we simulate M_i of the form given earlier in this section, with parameters (A) $d = 1.4$, $f = 0.2$, (B) $d = 1.4$, $f = 0.5$. Basic characteristics of the observation pattern for (A) were discussed before; for (B), the cross-sectional observation probability varies from $95\,\%$ at $t = 0$ to $50\,\%$ at $t = 1$, $21\,\%$ of curves are complete, the average length of missing periods (among incomplete curves) is 0.29. Configuration (IA) was used in the first two sets of simulations, other combinations contain less observed information. Results are reported in Table 3 where mean squared prediction errors are presented after standardisation by the true variance of the predicted quantity, i.e., by the variance of the missing part of the score, $\operatorname{var} \beta_{1jM_1}$, or by the trace of the covariance operator of the missing part of the trajectory, $\operatorname{trace} \mathscr{R}_{M_1M_1}$; after this standardisation it is possible to compare values under pattern (I) with their counterparts computed under (II). We see that the precision of estimation decreases as the amount of observed information (either on the curve of interest or on the sample) decreases.

6. An illustration: ambulatory blood pressure monitoring data

Heart rate profiles displayed in Figure 1 and their first derivative plotted in Figure 2 were obtained from raw observations by penalised spline smoothing described in the supplementary file available

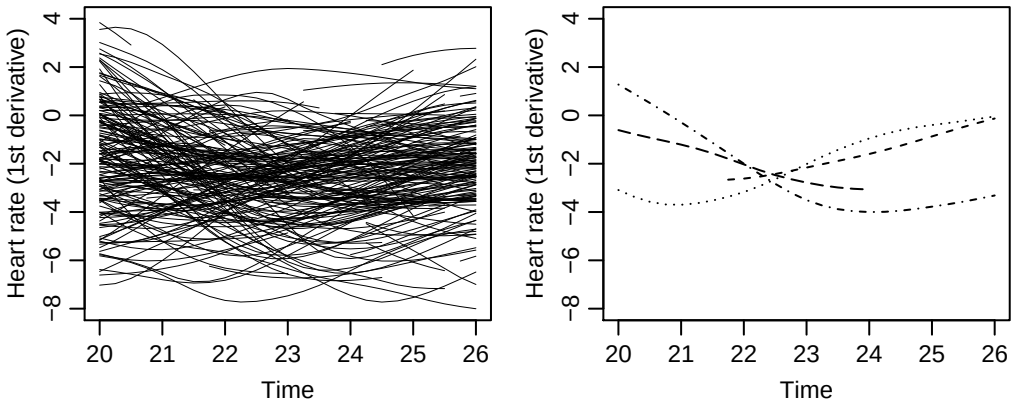

Fig. 2. A subset of the sample of the first derivatives of heart rate profiles (left) and several curves in detail (right)

online. The curves were registered by shifting the individual time scales so that every person's bedtime is 23 (i.e., 11 pm); individual bedtimes were available from a questionnaire. The methodology developed in this paper requires that the observation periods be independent of the curves. The expert opinion is that this is a realistic assumption; in addition, we performed exploratory graphical checks that did not indicate any problem with regard to this assumption.

From the shape of the mean functions of the profiles and their first derivatives it is obvious that on average heart rate profiles have a decreasing shape in this part of day and they decrease fastest around the bedtime. We wish to understand the main sources of variability between individual heart rate profiles. In Figure 3 we plot the first three eigenfunctions of the profiles and of their derivatives as perturbations of the mean shape (see Ramsay and Silverman, 2005, Subsection 8.3.1), i.e., we plot the mean profile plus and minus a suitable multiple of each eigenfunction (the eigenfunctions are multiplied by $0.9\hat{\lambda}_j^{1/2}$). For the profiles, we see that the most important component is the global level of heart rate, followed by a component describing the difference between the day and night values and a component that can be interpreted as a time shift. In terms of the first derivative, the first component quantifies the global level of the speed of decrease, the second component captures a shift in time and the third one characterises whether the individual's heart rate decreases rather suddenly or more gradually. The first three components explain a large proportion of the total variability and provide enough flexibility to capture individual shape features, e.g., the increasing trend of some curves in regions where the mean and most curves decrease.

Let us now focus on the individual level. To illustrate our prediction method for principal scores, we first consider the curve plotted in short dashed line in the right panels in Figures 1 and 2. The functional values are missing on a subset of the time interval and hence the principal scores cannot be computed directly. They can, however, be predicted. We give the results for the profile only (one can proceed analogously for the first derivative). The predicted values for the first three components are $(-28.7, 2.9, -1.9)$. Their prediction standard deviations quantifying the uncertainty are $(1.7, 2.3, 1.8)$. Mainly for the first two components they are relatively small compared with the standard deviations of the intrinsic variability $(24.0, 7.8, 3.7)$ (the square root of the eigenvalues); the corresponding relative errors are $(0.07, 0.29, 0.48)$. It is not surprising that the best precision is achieved for the first component: this component dominates the spectrum and is

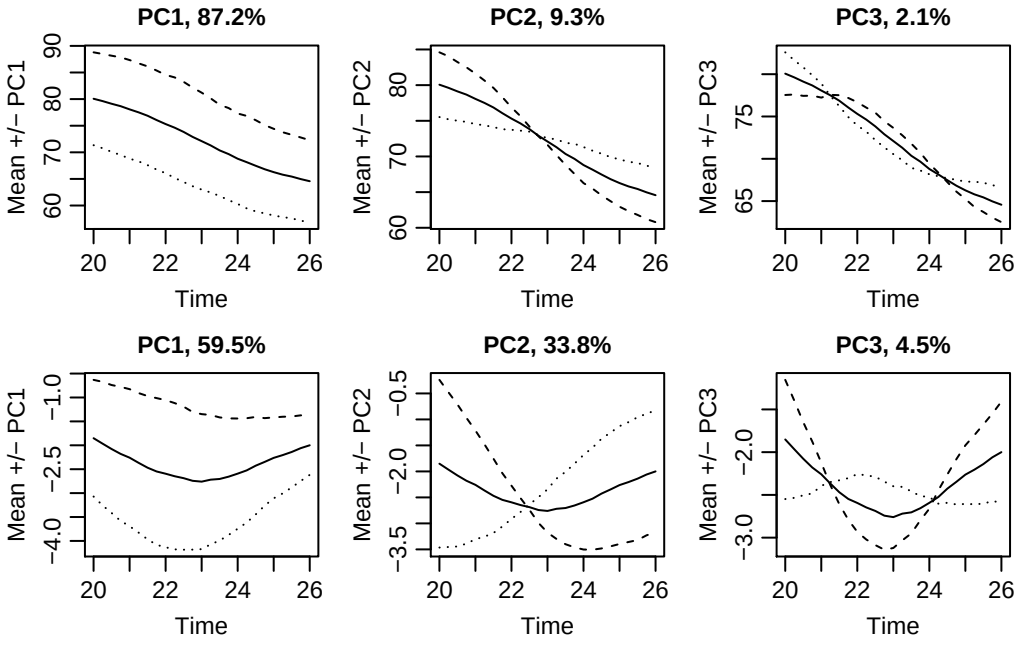

Fig. 3. The first three eigenfunctions of heart rate profiles (top row) and of their first derivative (bottom row) plotted as perturbations (dashed and dotted lines) of the mean (solid line)

rather simple (constant), so even a fraction of the curve provides relatively much information about the score. Next, we illustrate the method on the completely observed function plotted as dash-dot line in the right panels of Figures 1 and 2 from which we artificially remove observations on the time interval $[23.75, 26]$. Using the remaining part for the prediction, we estimate the scores by $(5.84, 4.43, 4.18)$ (with prediction standard deviations $(2.12, 2.68, 2.01)$) which is quite close to the true values $(5.76, 4.55, 4.32)$ computed from the complete curve (recall however that there will always be some random, non-vanishing discrepancy between the predicted and true values because we predict random variables by their conditional expectations).

Finally, we illustrate the functional reconstruction procedure. In Figure 4 we plot the two curves (and their derivatives) considered before and the reconstructed missing parts along with $95\,\%$ prediction bands. For the originally complete function (bottom row in Figure 4), we chose a difficult scenario: the missing period is relatively large (2.25 hours) and it contains a non-trivial change of shape of the curve mainly in terms of the first derivative which is decreasing in the observed region and increasing in the missing period. However, it is seen that the completion procedure is able to recover the missing part of information as the predicted curve (thick) approximates very well the true function (thin). It is interesting to notice that our method captures to some extent the presence of a local minimum in the first derivative. This illustrates the usefulness of the reconstruction procedure: without it important shape features like this would be concealed from the analyst. At first glance, some of the bands may seem to be wide but one needs to keep in mind that they are prediction (not confidence) bands and, therefore, must cover the random trajectory (rather than a nonrandom function) with a high probability. The uncertainty of the completion is in fact not big

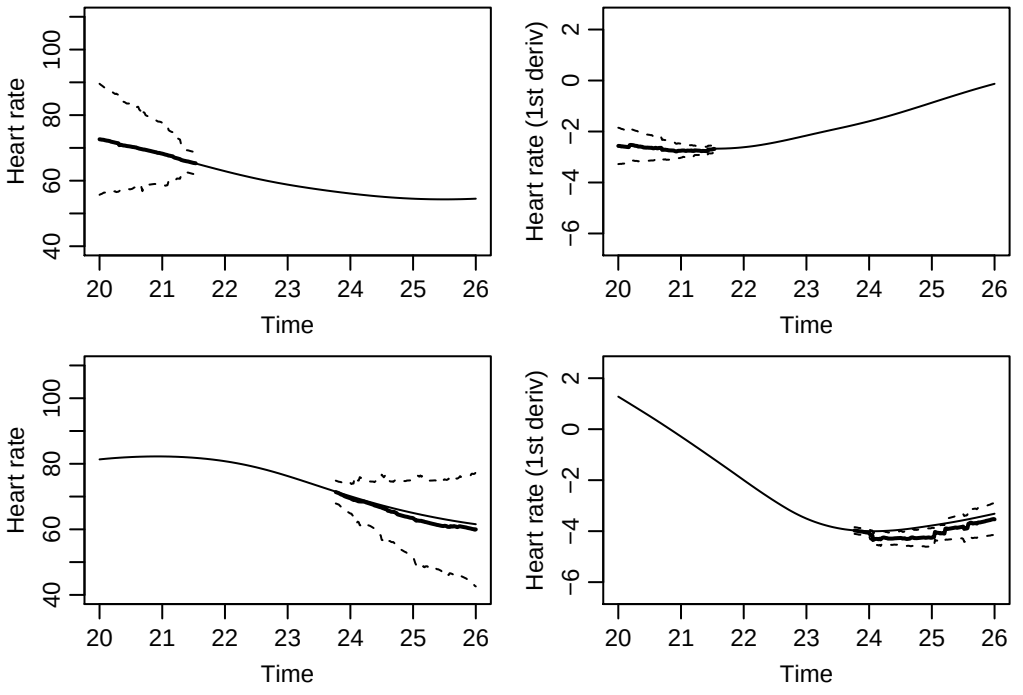

Fig. 4. Observed (thin) and reconstructed (thick) heart rate profiles (left) and derivatives (right) along with $95\,\%$ prediction bands for an incompletely observed curve (top row) and for a complete curve with an artificially introduced missing period (bottom row)

in proportion to the intrinsic variability of the stochastic process: the relative error is 0.10 and 0.11 for the curves on the top and bottom rows, respectively. A referee pointed out that the prediction bands for the derivatives are narrower than those for the curves. This is not a general phenomenon: it is possible to construct simple examples with prediction bands for derivatives wider than those for curves or examples with no such inequality. Differentiation is an operation that changes the covariance structure of functional data in a complex manner.

We compared our method with that of Yao et al. (2005a) applied to the raw heart rate values (not preprocessed by smoothing). Although their method was primarily developed for sparsely observed curves, it can be also used in our situation. Main results regarding the covariance structure of the profiles were similar for both methods. The proportion of variance explained by the first three principal components was $82.9\,\%$, $10.8\,\%$, $3.4\,\%$. The first three eigenfunctions had a similar shape and interpretation with both methods. There was a high degree of agreement between principal scores obtained by the two methods. The method of Liu and Müller (2009) can reconstruct derivatives. However, our method seems to be the only currently available method that can perform principal component analysis of derivatives under incompleteness. This is an important asset of our method over the other approach provided the data are sufficiently dense on subsets of the domain.

Acknowledgements

This work was done within the SKIPOGH project, a collaboration of Murielle Bochud (principal investigator), M. Burnier, O. Devuyst, P.-Y. Martin, M. Mohaupt, F. Paccaud, A. Péchère-Bertschi, B. Vogt, D. Ackermann, H. Alwan, Y. Bouatou, N. Dhayat, G. Ehret, I. Guessous, P. Monney, M.-E. Mueller, B. Ponte, M. Pruijm, S. Reverdin, P. Vuistiner, Z. Kutalik, S. Estoppey; the project was funded by the Swiss National Science Foundation. Special thanks are given to Murielle Bochud for her support and interest, and for her understanding of the importance of methodological developments in statistics. The hospitality of the Institute of Social and Preventive Medicine Lausanne is gratefully acknowledged. I am also grateful to the editor, an associate editor and two anonymous referees for their interesting comments and positive attitude.

A. Main proofs

Here we prove Theorems 1 and 2. Propositions 1, 2, 3 and 4 are proven in the supplementary document available online. Recall that we denote by $\|\cdot\|$ the L^2-norm of square integrable functions on a domain S that is obvious from the context (S will be $[0,1]$ or O_i or M_i). For linear operators, the symbols $\|\cdot\|_\infty$ and $\|\cdot\|_2$ are used for the operator norm and the Hilbert–Schmidt norm, respectively, where the operator will be a mapping between $L^2(S_1)$ and $L^2(S_2)$ with S_1, S_2 obvious from the context. For definitions of basic notions from operator theory, we refer to Bosq (2000).

A.1. Proof of Theorem 1

We neglect the fact that the data are centred by the estimated mean function and assume that the mean is known and equal to zero. The result remains valid when the curves are centred empirically, as the additional terms are negligible. It is enough to prove the inequality in the statement of the theorem; the remaining assertions follow easily. We write $|\hat\beta^{(\alpha)}_{ijM_i} - \tilde\beta_{ijM_i}| \le |\hat\beta^{(\alpha)}_{ijM_i} - \tilde\beta^{(\alpha)}_{ijM_i}| + |\tilde\beta^{(\alpha)}_{ijM_i} - \tilde\beta_{ijM_i}|$, which is a decomposition into the estimation error and approximation error. If we show that both errors converge in $L^2(P)$ to zero, the result will follow.

We denote the approximation error $A_1 = |\tilde\beta^{(\alpha)}_{ijM_i} - \tilde\beta_{ijM_i}|$ and compute

$$\begin{aligned}
\mathsf{E}(A_1^2) &= \mathsf{E}\{\langle X_{iO_i}, \tilde a^{(\alpha)}_{ij} - \tilde a_{ij}\rangle^2\} \\
&= \|\mathscr{R}^{1/2}_{O_iO_i}(\tilde a^{(\alpha)}_{ij} - \tilde a_{ij})\|^2 \\
&= \|\mathscr{R}^{1/2}_{O_iO_i}(\mathscr{R}^{(\alpha)-1}_{O_iO_i} - \mathscr{R}^{-1}_{O_iO_i})r_{ij}\|^2 \\
&= \sum_{k=1}^{\infty} \lambda_{O_iO_ik}\left(\frac{1}{\lambda_{O_iO_ik}+\alpha} - \frac{1}{\lambda_{O_iO_ik}}\right)^2 \langle r_{ij}, \varphi_{O_iO_ik}\rangle^2 \\
&= \alpha \sum_{k=1}^{\infty} \frac{\alpha\lambda_{O_iO_ik}}{(\lambda_{O_iO_ik}+\alpha)^2} \frac{\langle r_{ij}, \varphi_{O_iO_ik}\rangle^2}{\lambda^2_{O_iO_ik}} \\
&= O(\alpha),
\end{aligned}$$

where $\lambda_{O_iO_ik}$ and $\varphi_{O_iO_ik}$ are the eigenvalues and eigenfunctions of $\mathscr{R}_{O_iO_i}$ and the result follows from the fact that $\alpha\lambda_{O_iO_ik}/(\lambda_{O_iO_ik}+\alpha)^2 \le 1$ and Picard's condition (7).

Let us turn to the estimation error $|\hat{\beta}^{(\alpha)}_{ijM_i} - \tilde{\beta}^{(\alpha)}_{ijM_i}|$. The computation of expectations is complicated by the fact that the quantities $\hat{\mathscr{R}}_{O_iO_i}$ and $\hat{r}_{ij}$ are obtained from the whole sample including the ith function and thus are dependent with the ith function. We overcome this complication by first considering a modified problem with estimates of $\mathscr{R}_{O_iO_i}$ and r_{ij} independent of the ith function and then showing that this modification is asymptotically negligible. Specifically, we introduce $\hat{\beta}^{(\alpha)}_{ijM_i(-i)} = \hat{\mathscr{R}}^{(\alpha)-1}_{O_iO_i(-i)}\hat{r}_{ij(-i)}$ with $\hat{\mathscr{R}}^{(\alpha)}_{O_iO_i(-i)} = \hat{\mathscr{R}}_{O_iO_i(-i)} + \alpha\mathscr{I}_{O_i}$ and $\hat{r}_{ij(-i)} = \hat{\mathscr{R}}_{O_iM_i(-1)}\hat{\varphi}_{jM_i(-i)}$. Here $\hat{\mathscr{R}}_{O_iO_i(-i)}$ and $\hat{\mathscr{R}}_{O_iM_i(-i)}$ are suboperators of the estimated covariance operator $\hat{\mathscr{R}}_{(-i)}$ that is computed from all functions except the ith one, and $\hat{\varphi}_{jM_i(-i)}$ is a subfunction of the jth eigenfunction $\hat{\varphi}_{j(-i)}$ of $\hat{\mathscr{R}}_{(-i)}$. We decompose $|\hat{\beta}^{(\alpha)}_{ijM_i} - \tilde{\beta}^{(\alpha)}_{ijM_i}|$ as follows

$$|\hat{\beta}^{(\alpha)}_{ijM_i} - \tilde{\beta}^{(\alpha)}_{ijM_i}| \le |\hat{\beta}^{(\alpha)}_{ijM_i} - \hat{\beta}^{(\alpha)}_{ijM_i(-i)}| + |\hat{\beta}^{(\alpha)}_{ijM_i(-i)} - \tilde{\beta}^{(\alpha)}_{ijM_i}|, \tag{15}$$

and show that both terms converge in $L^2(P)$ to zero.

For the second term on the right-hand side in (15), $A_2 = |\hat{\beta}^{(\alpha)}_{ijM_i(-i)} - \tilde{\beta}^{(\alpha)}_{ijM_i}|$, we have

$$\begin{aligned}
\mathsf{E}(A_2^2) &= \mathsf{E}(\mathsf{E}[|\hat{\beta}^{(\alpha)}_{ijM_i(-i)} - \tilde{\beta}^{(\alpha)}_{ijM_i}|^2|\{X_{kO_k} : k \neq i\}]) \\
&= \mathsf{E}(\mathsf{E}[|\langle X_{iO_i}, \hat{a}^{(\alpha)}_{ij(-i)} - \tilde{a}^{(\alpha)}_{ij}\rangle^2|\{X_{kO_k} : k \neq i\}]) \\
&= \mathsf{E}\{\|\mathscr{R}^{1/2}_{O_iO_i}(\hat{a}^{(\alpha)}_{ij(-i)} - \tilde{a}^{(\alpha)}_{ij})\|^2\}.
\end{aligned}$$

Using the definitions of $\hat{a}^{(\alpha)}_{ij(-i)}$ and $\tilde{a}^{(\alpha)}_{ij}$ and the triangle inequality, we obtain

$$\begin{aligned}
\|\mathscr{R}^{1/2}_{O_iO_i}(\hat{a}^{(\alpha)}_{ij(-i)} - \tilde{a}^{(\alpha)}_{ij})\| \le{}& \|\mathscr{R}^{1/2}_{O_iO_i}\hat{\mathscr{R}}^{(\alpha)-1}_{O_iO_i(-i)}(\hat{\mathscr{R}}_{O_iM_i(-i)} - \mathscr{R}_{O_iM_i})\hat{\varphi}_{jM_i(-i)}\| \\
&+ \|\mathscr{R}^{1/2}_{O_iO_i}\hat{\mathscr{R}}^{(\alpha)-1}_{O_iO_i(-i)}\mathscr{R}_{O_iM_i}(\hat{\varphi}_{jM_i(-i)} - \hat{s}_j\varphi_{jM_i})\| \\
&+ \|\mathscr{R}^{1/2}_{O_iO_i}(\hat{\mathscr{R}}^{(\alpha)-1}_{O_iO_i(-i)} - \mathscr{R}^{(\alpha)-1}_{O_iO_i})\mathscr{R}_{O_iM_i(-i)}\varphi_{jM_i}\|
\end{aligned}$$

with $\hat{s}_j = \operatorname{sign}\langle\hat{\varphi}_{j(-i)}, \varphi_j\rangle$. Denote these three terms A_{21}, A_{22}, A_{23}, respectively. We see that

$$A_{21} \le \|\mathscr{R}^{1/2}_{O_iO_i}\|_\infty\|\hat{\mathscr{R}}^{(\alpha)-1}_{O_iO_i(-i)}\|_\infty\|\hat{\mathscr{R}}_{O_iM_i(-i)} - \mathscr{R}_{O_iM_i}\|_\infty\|\hat{\varphi}_{jM_i(-i)}\|.$$

Here, $\|\mathscr{R}^{1/2}_{O_iO_i}\|_\infty$ is a finite constant, $\|\hat{\mathscr{R}}^{(\alpha)-1}_{O_iO_i(-i)}\|_\infty \le \alpha^{-1}$ and $\|\hat{\varphi}_{jM_i(-i)}\| \le \|\hat{\varphi}_{j(-i)}\| = 1$. Using Proposition 1 we obtain $\mathsf{E}(A_{21}^2) \le \alpha^{-2}O(n^{-1})$. For the term A_{22} we have the bound

$$A_{22} \le \|\mathscr{R}^{1/2}_{O_iO_i}\|_\infty\|\hat{\mathscr{R}}^{(\alpha)-1}_{O_iO_i(-i)}\|_\infty\|\mathscr{R}_{O_iM_i}\|_\infty\|\hat{\varphi}_{jM_i(-i)} - \hat{s}_j\varphi_{jM_i}\|.$$

In light of Proposition 2, we see that $\mathsf{E}\,\|\hat{\varphi}_{jM_i(-i)} - \hat{s}_j\varphi_{jM_i}\|^2 \le \mathsf{E}\,\|\hat{\varphi}_{j(-i)} - \hat{s}_j\varphi_j\|^2 = O(n^{-1})$. This implies that $\mathsf{E}(A_{22}^2) \le \alpha^{-2}O(n^{-1})$. For the term A_{23}, first notice that

$$\begin{aligned}
\hat{\mathscr{R}}^{(\alpha)-1}_{O_iO_i(-i)} - \mathscr{R}^{(\alpha)-1}_{O_iO_i} &= \mathscr{R}^{(\alpha)-1}_{O_iO_i}(\mathscr{R}^{(\alpha)}_{O_iO_i} - \hat{\mathscr{R}}^{(\alpha)}_{O_iO_i(-i)})\hat{\mathscr{R}}^{(\alpha)-1}_{O_iO_i(-i)} \\
&= \mathscr{R}^{(\alpha)-1}_{O_iO_i}(\mathscr{R}_{O_iO_i} - \hat{\mathscr{R}}_{O_iO_i(-i)})\hat{\mathscr{R}}^{(\alpha)-1}_{O_iO_i(-i)}.
\end{aligned}$$

Therefore, we see that

$$A_{23} \leq \|\mathscr{R}_{O_iO_i}^{1/2}\mathscr{R}_{O_iO_i}^{(\alpha)-1}\|_\infty \|\hat{\mathscr{R}}_{O_iO_i(-i)} - \mathscr{R}_{O_iO_i}\|_\infty \|\hat{\mathscr{R}}_{O_iO_i(-i)}^{(\alpha)-1}\|_\infty \|\mathscr{R}_{O_iM_i}\|_\infty \|\hat{\varphi}_{jM_i(-i)}\|.$$

The first, third and fifth term are dominated by $\alpha^{-1/2}$, α^{-1} and 1, respectively. The fourth term is a finite constant. Using these bounds and Proposition 1 we get $\mathsf{E}(A_{23}^2) \leq \alpha^{-3}O(n^{-1})$. Hence with the help of the Cauchy–Schwarz inequality we finally obtain that $\mathsf{E}(A_2^2) \leq \alpha^{-3}O(n^{-1})$.

It remains to analyse the first term on the right-hand side of (15). It reflects the effect of omitting the ith observation in the estimation. As this effect is of order $O(n^{-2})$ in terms of mean squared difference, this term is negligible compared with the second term. In particular, it can be shown that $\mathsf{E}\{(\hat{\beta}_{ijM_i}^{(\alpha)} - \hat{\beta}_{ijM_i(-i)}^{(\alpha)})^2\} \leq \alpha^{-3}O(n^{-2})$. We omit technical details.

A.2. Proof of Theorem 2

To simplify the proof we assume that the mean is known to be zero and no centring is performed. The difference due to the estimation of the mean is of negligible order in comparison with other terms. Similarly to the proof of Theorem 1, we split the prediction error into the estimation error and regularisation error as follows, $\|\hat{X}_{iM_i}^{(\alpha)} - \tilde{X}_{iM_i}\| \leq \|\hat{X}_{iM_i}^{(\alpha)} - \tilde{X}_{iM_i}^{(\alpha)}\| + \|\tilde{X}_{iM_i}^{(\alpha)} - \tilde{X}_{iM_i}\|$.

For the regularisation error we compute

$$\begin{aligned}
\mathsf{E}\,\|\tilde{X}_{iM_i}^{(\alpha)} - \tilde{X}_{iM_i}\|^2 &= \|(\tilde{\mathscr{A}}_i^{(\alpha)} - \tilde{\mathscr{A}}_i)\mathscr{R}_{O_iO_i}^{1/2}\|_2^2 \\
&= \|\alpha\mathscr{R}_{M_iO_i}\mathscr{R}_{O_iO_i}^{-1}\mathscr{R}_{O_iO_i}^{(\alpha)-1}\mathscr{R}_{O_iO_i}^{1/2}\|_2^2 \\
&\leq \alpha\|\mathscr{R}_{M_iO_i}\mathscr{R}_{O_iO_i}^{-1}\|_2^2 \|\alpha^{1/2}\mathscr{R}_{O_iO_i}^{(\alpha)-1}\mathscr{R}_{O_iO_i}^{1/2}\|_\infty^2 \\
&= \alpha\|\tilde{\mathscr{A}}_i\|_2^2 \left(\sup_{k\in\mathbb{N}} \frac{\alpha^{1/2}\lambda_{O_iO_ik}^{1/2}}{\lambda_{O_iO_ik} + \alpha}\right)^2 \\
&\leq O(\alpha).
\end{aligned}$$

We turn to the estimation error. Similarly to the proof of Theorem 1 we avoid the dependence between $\hat{\mathscr{A}}_i^{(\alpha)}$ and X_{iO_i} in $\hat{X}_{iM_i}^{(\alpha)} = \hat{\mathscr{A}}_i^{(\alpha)}X_{iO_i}$ by considering $\hat{X}_{iM_i(-i)}^{(\alpha)} = \hat{\mathscr{A}}_{i(-i)}^{(\alpha)}X_{iO_i}$, where the estimator of the covariance operator in the prediction operator is replaced by its analogue based on all curves except the ith one. The difference is negligible in comparison with the remaining terms; for an analogous discussion see the proof of Theorem 1. The modified estimation error equals

$$\begin{aligned}
\mathsf{E}\,\|\hat{X}_{iM_i(-i)}^{(\alpha)} - \tilde{X}_{iM_i}^{(\alpha)}\|^2 &= \mathsf{E}\,\|\{\hat{\mathscr{R}}_{M_iO_i(-i)}\hat{\mathscr{R}}_{O_iO_i(-i)}^{(\alpha)-1} - \mathscr{R}_{M_iO_i}\mathscr{R}_{O_iO_i}^{(\alpha)-1}\}\mathscr{R}_{O_iO_i}^{1/2}\|_2^2 \\
&\leq \mathsf{E}[\|\{\hat{\mathscr{R}}_{M_iO_i(-i)} - \mathscr{R}_{M_iO_i}\}\hat{\mathscr{R}}_{O_iO_i(-i)}^{(\alpha)-1}\mathscr{R}_{O_iO_i}^{1/2}\|_2 \\
&\quad + \|\mathscr{R}_{M_iO_i}\{\hat{\mathscr{R}}_{O_iO_i(-i)}^{(\alpha)-1} - \mathscr{R}_{O_iO_i}^{(\alpha)-1}\}\mathscr{R}_{O_iO_i}^{1/2}\|_2]^2.
\end{aligned}$$

The proof is complete upon computing

$$\begin{aligned}
\mathsf{E}\,\|\{\hat{\mathscr{R}}_{M_iO_i(-i)} - \mathscr{R}_{M_iO_i}\}\hat{\mathscr{R}}_{O_iO_i(-i)}^{(\alpha)-1}\mathscr{R}_{O_iO_i}^{1/2}\|_2^2 &\leq \mathsf{E}\{\|\hat{\mathscr{R}}_{M_iO_i(-i)} - \mathscr{R}_{M_iO_i}\|_2^2 \\
&\quad \times \|\hat{\mathscr{R}}_{O_iO_i(-i)}^{(\alpha)-1}\|_\infty^2 \|\mathscr{R}_{O_iO_i}^{1/2}\|_\infty^2\} \\
&\leq \mathsf{E}\,\|\hat{\mathscr{R}}_{M_iO_i(-i)} - \mathscr{R}_{M_iO_i}\|_2^2 \alpha^{-2}\lambda_{O_iO_i1} \\
&= \alpha^{-2}O(n^{-1}),
\end{aligned}$$

$$\begin{aligned}\mathsf{E}\,\|\mathscr{R}_{M_iO_i}\{\hat{\mathscr{R}}^{(\alpha)-1}_{O_iO_i(-i)}-\mathscr{R}^{(\alpha)-1}_{O_iO_i}\}\mathscr{R}^{1/2}_{O_iO_i}\|_2^2 &\le \mathsf{E}\{\|\mathscr{R}_{M_iO_i}\|_\infty^2\|\hat{\mathscr{R}}^{(\alpha)-1}_{O_iO_i(-i)}\|_\infty^2\\ &\quad\times\|\hat{\mathscr{R}}_{O_iO_i(-i)}-\mathscr{R}_{O_iO_i}\|_2^2\|\mathscr{R}^{(\alpha)-1}_{O_iO_i}\mathscr{R}^{1/2}_{O_iO_i}\|_\infty^2\}\\ &\le \|\mathscr{R}_{M_iO_i}\|_\infty^2\alpha^{-2}\,\mathsf{E}\,\|\hat{\mathscr{R}}_{O_iO_i(-i)}-\mathscr{R}_{O_iO_i}\|_2^2\alpha^{-1}\\ &= \alpha^{-3}O(n^{-1}).\end{aligned}$$

References

Antoniadis, A. and Sapatinas, T. (2003). Wavelet methods for continuous-time prediction using Hilbert-valued autoregressive processes. *J. Multivariate Anal.*, 87(1):133–158.

Aston, J. A. D. and Kirch, C. (2012). Detecting and estimating changes in dependent functional data. *J. Multivariate Anal.*, 109:204–220.

Benko, M., Härdle, W., and Kneip, A. (2009). Common functional principal components. *Ann. Statist.*, 37(1):1–34.

Bosq, D. (2000). *Linear Processes in Function Spaces*. Springer, New York.

Bugni, F. A. (2012). Specification test for missing functional data. *Econometric Theory*, 28(5):959–1002.

Cai, T. T. and Hall, P. (2006). Prediction in functional linear regression. *Ann. Statist.*, 34(5):2159–2179.

Cardot, H., Ferraty, F., and Sarda, P. (1999). Functional linear model. *Statist. Probab. Lett.*, 45(1):11–22.

Cardot, H., Mas, A., and Sarda, P. (2007). CLT in functional linear regression models. *Probab. Theory Related Fields*, 138(3-4):325–361.

Dauxois, J., Pousse, A., and Romain, Y. (1982). Asymptotic theory for the principal component analysis of a vector random function: some applications to statistical inference. *J. Multivariate Anal.*, 12(1):136–154.

Delaigle, A. and Hall, P. (2013). Classification using censored functional data. *J. Amer. Statist. Assoc.*, 108(504):1269–1283.

Didericksen, D., Kokoszka, P., and Zhang, X. (2012). Empirical properties of forecasts with the functional autoregressive model. *Comput. Statist.*, 27(2):285–298.

Ferraty, F. and Romain, Y., editors (2011). *The Oxford Handbook of Functional Data Analysis*. Oxford University Press, Oxford.

Ferraty, F. and Vieu, P. (2006). *Nonparametric Functional Data Analysis*. Springer, New York.

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

Groetsch, C. W. (1993). *Inverse Problems in the Mathematical Sciences*. Vieweg, Braunschweig.

Hall, P. and Horowitz, J. L. (2007). Methodology and convergence rates for functional linear regression. *Ann. Statist.*, 35(1):70–91.

Hall, P. and Hosseini-Nasab, M. (2006). On properties of functional principal components analysis. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 68(1):109–126.

Hall, P., Müller, H.-G., and Wang, J.-L. (2006). Properties of principal component methods for functional and longitudinal data analysis. *Ann. Statist.*, 34(3):1493–1517.

He, G., Müller, H.-G., and Wang, J.-L. (2003). Functional canonical analysis for square integrable stochastic processes. *J. Multivariate Anal.*, 85(1):54–77.

He, G., Müller, H.-G., Wang, J.-L., and Yang, W. (2010). Functional linear regression via canonical analysis. *Bernoulli*, 16(3):705–729.

Horváth, L., Hušková, M., and Kokoszka, P. (2010). Testing the stability of the functional autoregressive process. *J. Multivariate Anal.*, 101(2):352–367.

Horváth, L. and Kokoszka, P. (2012). *Inference for Functional Data with Applications*. Springer, New York.

Horváth, L., Kokoszka, P., and Reeder, R. (2013). Estimation of the mean of functional time series and a two-sample problem. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 75(1):103–122.

James, G. M. and Hastie, T. J. (2001). Functional linear discriminant analysis for irregularly sampled curves. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 63(3):533–550.

James, G. M., Hastie, T. J., and Sugar, C. A. (2000). Principal component models for sparse functional data. *Biometrika*, 87(3):587–602.

Jarušková, D. (2013). Testing for a change in covariance operator. *J. Statist. Plann. Inference*, 143(9):1500–1511.

Jolliffe, I. T. (2002). *Principal Component Analysis*. Springer, New York.

Kallenberg, O. (2002). *Foundations of Modern Probability*. Springer, New York.

Kargin, V. and Onatski, A. (2008). Curve forecasting by functional autoregression. *J. Multivariate Anal.*, 99(10):2508–2526.

Kraus, D. and Panaretos, V. M. (2012). Dispersion operators and resistant second-order functional data analysis. *Biometrika*, 99(4):813–832.

Krzanowski, W. J. (2000). *Principles of Multivariate Analysis*. Oxford University Press, Oxford.

Liebl, D. (2013). Modeling and forecasting electricity spot prices: a functional data perspective. *Ann. Appl. Stat.*, 7(3):1562–1592.

Liu, B. and Müller, H.-G. (2009). Estimating derivatives for samples of sparsely observed functions, with application to online auction dynamics. *J. Amer. Statist. Assoc.*, 104(486):704–717.

Mas, A. (2007). Testing for the mean of random curves: a penalization approach. *Stat. Inference Stoch. Process.*, 10(2):147–163.

Müller, H.-G. and Stadtmüller, U. (2005). Generalized functional linear models. *Ann. Statist.*, 33(2):774–805.

Panaretos, V. M., Kraus, D., and Maddocks, J. H. (2010). Second-order comparison of Gaussian random functions and the geometry of DNA minicircles. *J. Amer. Statist. Assoc.*, 105(490):670–682.

Pruijm, M., Ponte, B., Ackermann, D., Vuistiner, P., Paccaud, F., Guessous, I., Ehret, G., Eisenberger, U., Mohaupt, M., Burnier, M., Martin, P.-Y., and Bochud, M. (2013). Heritability, determinants and reference values of renal length: a family-based population study. *European Radiology*, 23(10):2899–2905.

Ramsay, J. O., Hooker, G., and Graves, S. (2009). *Functional Data Analysis with R and MATLAB*. Springer, New York.

Ramsay, J. O. and Silverman, B. W. (2002). *Applied Functional Data Analysis*. Springer, New York.

Ramsay, J. O. and Silverman, B. W. (2005). *Functional Data Analysis*. Springer, New York.

Sangalli, L. M., Secchi, P., Vantini, S., and Veneziani, A. (2009). A case study in exploratory functional data analysis: geometrical features of the internal carotid artery. *J. Amer. Statist. Assoc.*, 104(485):37–48.

Yao, F., Müller, H.-G., and Wang, J.-L. (2005a). Functional data analysis for sparse longitudinal data. *J. Amer. Statist. Assoc.*, 100(470):577–590.

Yao, F., Müller, H.-G., and Wang, J.-L. (2005b). Functional linear regression analysis for longitudinal data. *Ann. Statist.*, 33(6):2873–2903.