SFB 823

Dynamic functional principal components

Siegfried Hörmann, Lukasz Kidzinski,
Marc Hallin

Nr. 48/2012

Discussion Paper

Dynamic Functional Principal Components

Siegfried Hörmann[1*] Łukasz Kidziński[2]
Marc Hallin[3,4]

[1,2] Department of Mathematics, Université libre de Bruxelles CP210, Bd. du Triomphe, B-1050 Brussels, Belgium.
[3] ECARES, Université libre de Bruxelles CP 114/04 50, avenue F.D. Roosevelt B-1050 Brussels, Belgium.
[4] ORFE, Princeton University, Sherrerd Hall, Princeton, NJ 08540, USA.

Abstract. In this paper, we address the problem of dimension reduction for sequentially observed functional data $(X_k \colon k \in \mathbb{Z})$. Such functional time series arise frequently, e.g., when a continuous time process is segmented into some smaller natural units, such as days. Then each X_k represents one intraday curve. We argue that functional principal component analysis (FPCA), though a key technique in the field and a benchmark for any competitor, does not provide an adequate dimension reduction in a time series setting. FPCA is a static procedure which ignores valuable information in the serial dependence of the functional data. Therefore, inspired by Brillinger's theory of dynamic principal components, we propose a dynamic version of FPCA which is based on a frequency domain approach. By means of a simulation study and an empirical illustration, we show the considerable improvement our method entails when compared to the usual (static) procedure. While the main part of the article outlines the ideas and the implementation of dynamic FPCA for functional X_k, we provide in the appendices a rigorous theory for general Hilbertian data.

1. Introduction

The tremendous technical improvements in data collection and storage allow to get an increasingly complete picture of many common phenomena. In principle, all processes in real life are continuous in time and, with improved data acquisition techniques, they can be recorded at arbitrarily high frequency. To benefit from increasing information, we need appropriate statistical tools that can help us extracting

*Corresponding author. Email: shormann@ulb.ac.be

the most important characteristics of some possibly high-dimensional specifications. Functional data analysis (FDA) has proven in recent years to be an appropriate tool in many such cases and has consequently evolved into a very important field of research in the statistical community.

Most classically, functional data are considered as realizations of (smooth) random curves. Then every observation X is a curve $(X(u)\colon u \in \mathcal{U})$. One generally assumes, for simplicity, that $\mathcal{U} = [0, 1]$, but $\mathcal{U}$ could be a more complex domain like a cube or the surface of a sphere. Since observations are functions, we are dealing with high-dimensional, in fact intrinsically infinite-dimensional objects. So, not surprisingly, there is a clear demand for efficient data reduction techniques. As such, *functional principal component analysis* (FPCA) has taken a leading role in FDA. Arguably, it can be seen as *the* key technique in the field. In analogy to classical multivariate PCA (see Jolliffe [21]), functional PCA heavily relies on an eigendecomposition of the underlying covariance function. The mathematical foundations for this have been laid several decades ago in the pioneering papers by Karhunen [22] and Loève [25], but it took a while until the method was popularized in the statistical community. Some earlier contributions are Besse and Ramsay [4], Ramsay and Dalzell [28] and, later, the influential books by Ramsay and Silverman [29], [30] and Ferraty and Vieu [10]. Statisticians have been working on problems related to estimation and inference (Kneip and Utikal [23], Benko et al. [3]), asymptotics (Dauxois et al. [9] and Hall and Hosseini-Nasab [14]), smoothing techniques (Silverman [32]), sparse data (James et al. [20], Hall et al. [15]), and robustness issues (Locantore et al. [24], Gervini [11]), to name just a few. Important applications include FPC-based estimation of functional linear models (Cardot et al. [8], Reiss and Ogden [31]) or forecasting (Hyndman and Ullah [19], Aue et al. [1]). The usefulness of functional PCA has also been recognized in other scientific disciplines, like chemical engineering (Gokulakrishnan et al. [13]) or functional magnetic resonance imaging (Aston and Kirch [2], Viviani et al. [34]). Many more references can be found in the above cited papers and in Sections 8–10 of Ramsay and Silverman [30], where we refer to for background reading. A further reason for the success of FPCA seems to be the fact that, in contrast to their multivariate counterpart, FPCs do not suffer from the lack of scale invariance. Roughly speaking, while in the vector case different components can have completely different measuring units, all points $X(u)$, $u \in [0, 1]$, of some curve are expressed in the same units, and rescaling at different u values is usually not meaningful.

Most existing concepts and methods in FDA, even though they may tolerate serial dependence, have been developed for independent observations. This is a serious weakness, as in numerous applications the functional data under study are obviously dependent, either in time or in space. Examples include daily curves of financial transactions, daily patterns of geophysical and environmental data, annual temperatures measured on the surface of the earth, etc. In such cases, we should view the data as the realization of a *functional time series* $(X_t(u)\colon t \in \mathbb{Z})$, where the time parameter t is discrete and the parameter u is continuous. For example, in case of daily observations, the curve $X_t(u)$ may be viewed as the observation on day t with intraday time parameter u. A key reference on functional time series techniques is Bosq [7], who studied functional versions of AR processes. We also refer to Hörmann and Kokoszka [18] for a survey.

Ignoring time dependence in this time series context may result in misleading, or even completely wrong, findings, and highly inefficient procedures. Similar conclusions motivated Hörmann and Kokoszka [17] to investigate the robustness properties of some classical FDA methods in the presence of serial dependence. In particular, they show that usual FPCs still can be consistently estimated within a quite general dependence framework. Yet, the basic problem remains that FPCA operates in a *static* way: when applied to serially dependent curves, it fails to take into account the potentially very valuable information carried by the past values of the functional observations under study. In particular, a static FPC with small eigenvalue, hence negligible instantaneous impact on X_t, may have a major impact on X_{t+1}, and high predictive value. Neglecting it, as FPCA does, may have serious consequences.

Besides their failure to produce adequate dimension reduction, static FPCs, while cross-sectionally uncorrelated at fixed time t, typically still exhibit lagged cross-correlations. Therefore, unlike in the i.i.d. case, the resulting FPC scores cannot be analyzed componentwise, but need to be considered as vector time series, which are less easy to handle and interpret.

These shortcomings motivated our development of *dynamic functional principal components*. The idea is to transform the functional time series into a vector time series (of low dimension 3 or 4, say), where the individual component processes are mutually uncorrelated, and account for most of the dynamics and variability of the original process. The analysis of the functional time series can then be performed on those *dynamic principal components*. Since the transformed variables are non-correlated, we can even perform any second-order based analysis componentwise. In analogy to the static FPCA, the curves can be optimally reconstructed/approximated from the low dimensional time series via a dynamic version of the celebrated *Karhunen-Loève expansion*.

Dynamic PCs first have been suggested by Brillinger [5] for vector time series. The purpose of this article is to extend the Brillinger approach to a functional, or more general *Hilbert space* setting. The methodology heavily relies on a frequency domain analysis for functional data, which has been only recently brought forth by Panaretos and Tavakoli [26].

An impression of how well the proposed method works can be obtained from Figure 1. Its left panel shows ten consecutive intraday curves of some pollutant level. (A detailed description of the underlying data is given in Section 5.) The two panels to the right show the reconstructions of these curves after performing of a dimension reduction to dimension one. We used static FPCA in the central panel and dynamic FPCA in the right panel. The difference is striking. While the static method solely reproduces an average level and exhibits a spurious intraday symmetry, the dynamic counterpart to a large extent catches the evolution of the curves. In particular, it retrieves remarkably well the intraday trend of the pollution levels.

The rest of the paper is organized as follows. In Section 2, we describe our approach and state a number of relevant propositions. In Section 3, we discuss its

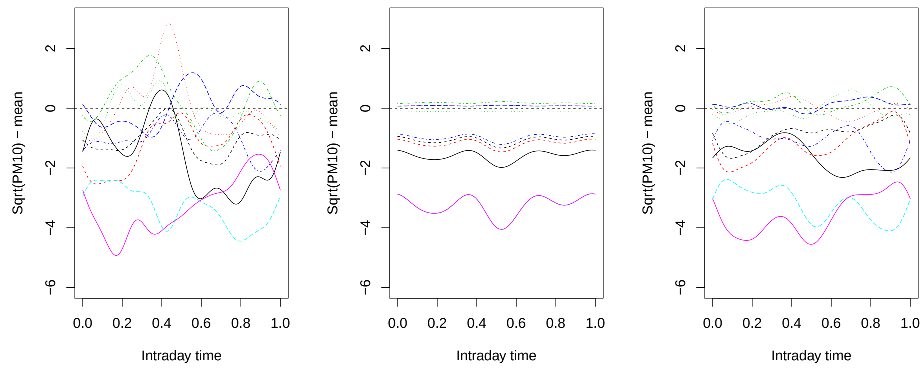

Figure 1: Ten subsequent observations (left panel), the corresponding static Karhunen-Loève expansion with one component (middle panel) and the dynamic Karhunen-Loève expansion with one component (right panel).

practical implementation and the related numerical costs. After a simulation study in Section 4, we illustrate the methodology by a real data example on pollution curves. Appendix A contains a rigorous mathematical framework and the proofs in a general Hilbertian setting. Finally, in Appendix B, we justify the proposed estimation steps by providing some asymptotics.

2. Methodology for L^2 curves

In this section, we introduce some necessary notation and tools. Most of the discussion on technical details is postponed to Appendices A and B. While focusing here on $L^2([0,1])$-valued processes, i.e. square integrable functions defined on the unit interval, we will work, in the technical appendices, with processes taking values in some arbitrary separable Hilbert space. Such general setup facilitates notation and makes theory clearer, but we postpone it until Appendix A, to make the paper more easily accessible for readers less familiar with functional analysis.

2.1. Notation and setup

Throughout this section, we consider a functional time series $(X_t\colon t \in \mathbb{Z})$, where X_t takes values in the space $H := L^2([0,1])$ of complex-valued square-integrable functions on $[0,1]$. This means that $X_t = (X_t(u)\colon u \in [0,1])$ and

$$\int_0^1 |X_t(u)|^2 du < \infty,$$

where $|z| := \sqrt{z\bar{z}}$, with $\bar{z}$ the complex conjugate of z, denotes the modulus of $z \in \mathbb{C}$. In most applications, observations are real, but, since we will use spectral methods, a complex vector space definition will serve useful.

The space H then is a Hilbert space, equipped with the inner product $\langle x, y\rangle := \int_0^1 x(t)\bar{y}(t)dt$, so that $\|x\| = \sqrt{\langle x, x\rangle}$ defines a norm. The notation $X \in L_H^p$ is used to indicate that, for some $p > 0$, $E[\|X\|^p] < \infty$. Any $X \in L_H^1$ then possesses a mean curve $\mu = (E[X(t)] \colon t \in [0,1])$, and any $X \in L_H^2$ a covariance operator C, defined by $C(x) := E[\langle X - \mu, x\rangle \overline{(X - \mu)}]$. The operator C is a kernel operator given by

$$C(x)(u) = \int_0^1 c(u, v)x(v)dv, \text{ with } c(u, v) := \operatorname{cov}(X(u), X(v)), \; u, v \in [0, 1].$$

The process $(X_t \colon t \in \mathbb{Z})$ is called *weakly stationary* if for all t we have (i) $X_t \in L_H^2$, (ii) $EX_t = EX_0$ and (iii) for all $h \in \mathbb{Z}$ and $u, v \in [0, 1]$

$$\operatorname{cov}(X_{t+h}(u), X_t(v)) = \operatorname{cov}(X_h(u), X_0(v)) =: c_h(u, v).$$

Denote by $C_h, h \in \mathbb{Z}$, the operator corresponding to the autocovariance kernels c_h. Clearly, $C_0 = C$. For our problem, the mean is not important, so we will throughout suppose that random elements are centered. *For the rest of the paper, it will be tacitly imposed that $(X_t \colon t \in \mathbb{Z})$ is a weakly stationary, zero mean process defined on some probability space $(\Omega, \mathcal{A}, P)$.*

As in the multivariate case, the covariance operator C of a random element $X \in L_H^2$ admits an eigendecomposition (see, e.g., p. 178, Theorem 5.1 in [12])

$$C(x) = \sum_{\ell=1}^{\infty} \lambda_\ell \langle x, v_\ell\rangle v_\ell,$$

where $(\lambda_\ell \colon \ell \geq 1)$ are C's eigenvalues (in descending order) and $(v_\ell \colon \ell \geq 1)$ the corresponding normalized eigenfunctions, so that $C(v_\ell) = \lambda_\ell v_\ell$ and $\|v_\ell\| = 1$. If C has full rank, then the sequence $(v_\ell \colon \ell \geq 1)$ forms an orthonormal basis (ONB) of $L^2([0, 1])$. Hence X admits the representation

$$X = \sum_{\ell=1}^{\infty} \langle X, v_\ell\rangle v_\ell, \tag{1}$$

which is called the *Karhunen-Loève (KL)* expansion of X. The eigenfunctions v_ℓ are called *the (static) functional principal components (FPCs)* and the coefficients $\langle X, v_\ell\rangle$ are called *the (static) FPC scores* or *loadings*. It is well known that the basis $(v_\ell \colon \ell \geq 1)$ is optimal in representing X in the following sense: if $(w_\ell \colon \ell \geq 1)$ is any other ONB of H, then

$$E\|X - \sum_{\ell=1}^{p} \langle X, v_\ell\rangle v_\ell\|^2 \leq E\|X - \sum_{\ell=1}^{p} \langle X, w_\ell\rangle w_\ell\|^2, \quad \forall p \geq 1. \tag{2}$$

Property (2) shows that a finite number of FPCs can be used to transform the function X to a vector of given dimension p with a minimum loss of "instantaneous" information. It should be noted, though, that this transformation is *static* in its nature, meaning that it is performed observation by observation, and does not take into account the possible serial dependence of the X_t's, which is likely to exist in a time series context. Globally speaking, we should be looking for a transformation which involves *all* observations, and is based on the whole family $(C_h\colon h \in \mathbb{Z})$ rather than on C_0 only. To achieve this goal, we introduce below the *spectral density operator*, which contains the full information on the family of operators $(C_h\colon h \in \mathbb{Z})$.

2.2. The Spectral Density Operator

Existence of the operator to be defined below requires a summability condition on the autocovariance operators C_h. Specifically, we assume that

$$\sum_{h\in\mathbb{Z}} \left(\int_0^1 \int_0^1 |c_h(u,v)|^2 du dv \right)^{1/2} < \infty, \tag{3}$$

a condition that is more conveniently expressed as

$$\sum_{h\in\mathbb{Z}} \|C_h\|_{\mathcal{S}} < \infty, \tag{4}$$

where $\|\cdot\|_{\mathcal{S}}$ denotes the Hilbert-Schmidt norm (see Section A.1). A simple sufficient condition for (4) will be provided in Proposition 7. Now, set

$$f_\theta^X(u,v) := \frac{1}{2\pi} \sum_{h\in\mathbb{Z}} c_h(u,v) e^{-\mathrm{i}h\theta}, \quad \theta \in [-\pi, \pi],$$

where i denotes the imaginary unit. By (3), this series converges in mean square for all θ.

Definition 1. *Let (X_t) be a stationary process. The operator $\mathcal{F}_\theta^X$ whose kernel is $f_\theta^X(\cdot,\cdot)$ is called the* spectral density operator of (X_t) at frequency θ*.*

This concept of a spectral density operator has been very recently introduced by Panaretos and Tavakoli [26], where we refer to for many interesting details on estimation and asymptotics. In our context, this operator is used to create particular *functional filters* (see Sections 2.3 and A.3) which are the building blocks for the construction of *dynamic* FPCs. A functional filter is defined via a sequence $\mathbf{\Phi} = (\Phi_\ell\colon k \in \mathbb{Z})$ of linear operators between two spaces H and H'. The filtered variables Y_t have the form $Y_t = \sum_{k\in\mathbb{Z}} \Phi_\ell(X_{t-\ell})$. For the important case when $H' = \mathbb{R}^p$, the following proposition relates the spectral density operator of (X_t) to the spectral density matrix of such a filtered sequence (Y_t). This simple result plays a crucial role in our construction. Let $\|\Psi\| := \sup_{\|x\|\leq 1} \|\Psi(x)\|$ denote the operator norm of some operator Ψ.

Proposition 1. *Assume that $\Phi_\ell(x) = (\langle x, \phi_{1\ell}\rangle, \langle x, \phi_{2\ell}\rangle, \ldots, \langle x, \phi_{p\ell}\rangle)'$, with $\phi_{m\ell} \in H$ and $\sum_{\ell\in\mathbb{Z}} \|\Phi_\ell\| < \infty$. Then $\sum_{\ell\in\mathbb{Z}} \Phi_\ell(X_{t-\ell})$ converges in mean square to a limit Y_t.*

The p-dimensional vector process is stationary and has a spectral density matrix $\mathcal{F}_\theta^Y$ given by

$$\mathcal{F}_\theta^Y = \begin{pmatrix} \langle \mathcal{F}_\theta^X(\phi_1^\star(\theta)), \phi_1^\star(\theta)\rangle & \cdots & \langle \mathcal{F}_\theta^X(\phi_p^\star(\theta)), \phi_1^\star(\theta)\rangle \\ \vdots & \ddots & \vdots \\ \langle \mathcal{F}_\theta^X(\phi_1^\star(\theta)), \phi_p^\star(\theta)\rangle & \cdots & \langle \mathcal{F}_\theta^X(\phi_p^\star(\theta)), \phi_p^\star(\theta)\rangle \end{pmatrix},$$

where $\phi_m^\star(\theta) := \sum_{\ell\in\mathbb{Z}} \phi_{m\ell} e^{\mathrm{i}\ell\theta}$.

To explain the important consequence of this result, first observe that, for every frequency θ, the operator $\mathcal{F}_\theta^X$ is a non-negative, self-adjoint Hilbert-Schmidt operator. Hence, assuming for the moment that the kernel $f_\theta^X(u,v)$ is continuous in u and v, we obtain, by Mercer's theorem (see, e.g., p. 197, Theorem 3.1 in [12]),

$$f_\theta^X(u,v) = \sum_{m\geq 1} \lambda_m(\theta)\varphi_m(u|\theta)\overline{\varphi_m(v|\theta)}. \tag{5}$$

Here, $\varphi_m(u|\theta)$ (in short, $\varphi_m(\theta)$) and $\lambda_m(\theta)$ are the eigenfunctions and eigenvalues, respectively, of $\mathcal{F}_\theta^X$. The series (5) converges absolutely and uniformly on $[0,1]^2$. We impose the order $\lambda_1(\theta) \geq \lambda_2(\theta) \geq \ldots \geq 0$ for all $\theta \in [-\pi,\pi]$, and require that the eigenfunctions be standardized, so that $\|\varphi_m(\theta)\| = 1$ for all $m \geq 1$ and $\theta \in [-\pi,\pi]$. Then the sequences $(\varphi_m(\theta) \colon m \geq 1)$ form orthonormal bases of the closure $\overline{\mathrm{Im}(\mathcal{F}_\theta^X)}$ of the image of $\mathcal{F}_\theta^X$. If $\mathcal{F}_\theta^X$ is not full-rank, we can always extend $(\varphi_m(\theta) \colon m \geq 1)$ into a basis of H, and thus, without loss of generality, we assume that the closed span $\overline{\mathrm{sp}}(\varphi_m(\theta) \colon m \geq 1)$ is H.

Assume now that we could choose the functional filters $(\phi_{m\ell} \colon \ell \in \mathbb{Z})$ such that $\phi_m^\star(\theta) = \varphi_m(\theta)$. We then have $\mathcal{F}_\theta^Y = \mathrm{diag}(\lambda_1(\theta), \ldots, \lambda_m(\theta))$, implying that the coordinate processes of (Y_t) are uncorrelated at any lag: $\mathrm{cov}(Y_{mt}, Y_{m's}) = 0$ for all s,t if $m \neq m'$. As discussed in the Introduction, this is a highly desirable property which the static FPCs do not possess.

2.3. Dynamic FPCs

Motivated by the discussion above, we wish to define $\phi_{m\ell}$ in such a way that

$$\phi_m^\star(\theta) := \sum_{\ell\in\mathbb{Z}} \phi_{m\ell} e^{\mathrm{i}\ell\theta} = \varphi_m(\theta),$$

which is the case if the $\phi_{m\ell}$'s are the coefficients of the Fourier expansion of $\varphi_m(\theta)$ as a function in θ. Since $\phi_{m\ell} = \phi_{m\ell}(u)$ is a curve, the concept of Fourier expansion requires some explanation here. By Fubini's theorem,

$$2\pi = \int_{-\pi}^{\pi}\int_0^1 \varphi_m^2(u|\theta)dud\theta = \int_0^1\int_{-\pi}^{\pi} \varphi_m^2(u|\theta)d\theta du,$$

showing that, for all m and almost all $u \in [0,1]$, $\varphi_m(u|\theta)$ is square integrable with respect to $\theta \in [-\pi,\pi]$. Thus, for almost all u, the Fourier expansion of $\varphi_m(u|\theta)$ as a function in θ exists and takes the form

$$\varphi_m(u|\theta) = \sum_{\ell\in\mathbb{Z}} \frac{1}{2\pi}\int_{-\pi}^{\pi} \varphi_m(u|s)e^{-\mathrm{i}\ell s}ds\, e^{\mathrm{i}\ell\theta} =: \sum_{\ell\in\mathbb{Z}} \phi_{m\ell}(u)e^{\mathrm{i}\ell\theta}. \tag{6}$$

This leads to the following definition.

Definition 2 (Dynamic functional principal components). *Assume that* $(X_t\colon t \in \mathbb{Z})$ *is a mean zero stationary process with values in* L^2_H *satisfying assumption* (4). *Let* $\phi_{m\ell}$ *be defined as in* (6). *Then the* m-th dynamic functional principal component (DFPC) score of X_t *is*

$$Y_{mt} := \sum_{\ell \in \mathbb{Z}} \langle X_{t-\ell}, \phi_{m\ell} \rangle, \quad t \in \mathbb{Z},\ m \geq 1. \tag{7}$$

We call $\Phi_m := (\phi_{m\ell}\colon \ell \in \mathbb{Z})$ *the* m-th DFPC filter coefficients.

The rest of this section is devoted to some important properties of dynamic FPCs.

Proposition 2 (Elementary properties). *Assume that* $(X_t\colon t \in \mathbb{Z})$ *is a real-valued stationary process satisfying* (4) *and let* Y_{mt} *be its dynamic FPC scores. Then,*

(a) the eigenfunctions $\varphi_m(\theta)$ *are Hermitian, and hence* Y_{mt} *is real;*

(b) if $C_h = 0$ *for* $h \neq 0$*, the dynamic FPC scores coincide with the static ones.*

Our construction of $\phi_{m\ell}$ was motivated through Proposition 1. In order to apply it to the thus defined functional filters, we shall now impose for some of the subsequent results that

$$\sum_{k \in \mathbb{Z}} \|\phi_{m\ell}\| < \infty. \tag{8}$$

Proposition 3 (Second-order properties). *Assume* $(X_t\colon t \in \mathbb{Z})$ *is a stationary process satisfying* (4) *and let* Y_{mt} *be its dynamic FPC scores. Then,*

(a) the series defining Y_{mt} *is mean-square convergent, with*

$$EY_{mt} = 0 \quad \textit{and} \quad EY_{mt}^2 = \sum_{\ell \in \mathbb{Z}} \sum_{k \in \mathbb{Z}} \langle C_{\ell-k}(\phi_{m\ell}), \phi_{mk} \rangle.$$

Assume that, in addition to the previous assumptions, (8) *holds. Then,*

(b) for $m \neq m'$*, the dynamic FPC scores* Y_{mt} *and* $Y_{m's}$ *are uncorrelated for all* s, t*;*

(c) the long-run variance of the m*-th dynamic FPC score sequence is*

$$\lim_{n \to \infty} \frac{1}{n} \mathrm{Var}(Y_{m1} + \cdots + Y_{mn}) = 2\pi \lambda_m(0).$$

It is important to note that Part (a) of Proposition 3 holds without assumption (8). Thus, the definition of (7) is always meaningful. However, for the derivation of more general properties of dynamic FPC scores, condition (8) seems to be a minimal requirement.

The next proposition tells us how we can recover the original process $(X_t(u)\colon t \in \mathbb{Z},\ u \in [0,1])$ from $(Y_{mt}\colon t \in \mathbb{Z},\ m \geq 1)$. It is the dynamic analogue of the static Karhunen-Loève expansion (1) associated with static principal components.

Proposition 4 (Inversion formula)**.** *Let Y_{mt} be the DFPC scores related to the process $(X_t(u)\colon t \in \mathbb{Z},\, u \in [0,1])$. Assume that* (8) *holds. Then,*

$$X_t(u) = \sum_{m \geq 1} X_{mt}(u) \quad \textit{with} \quad X_{mt}(u) := \sum_{\ell \in \mathbb{Z}} Y_{m,t+\ell}\phi_{m\ell}(u) \tag{9}$$

(where convergence is in mean square). We call (9) *the* dynamic Karhunen-Loève expansion of X_t.

The random variables defined by $\sum_{m=1}^{p} X_{mt}(u)$, $p \geq 1$, can be seen as p-dimensional reconstructions of $X_t(u)$, which only involve the p time series $(Y_{mt}\colon t \in \mathbb{Z})$, $1 \leq m \leq p$. Competitors to this reconstruction are obtained by replacing $\phi_{m\ell}$ in (7) and (9) with other elements $\psi_{m\ell}$ and $\upsilon_{m\ell}$. The next theorem shows that within this class of p-dimensional processes, $\sum_{m=1}^{p} X_{mt}(u)$ approximates $X_t(u)$ in an optimal way.

Proposition 5 (Optimality)**.** *Let Y_{mt} be the DFPC scores related to the process $(X_t\colon t \in \mathbb{Z})$ and let X_{mt} be defined as in Proposition 4 and assume that* (8) *holds. Let $\tilde{X}_{mt} = \sum_{\ell \in Z} \tilde{Y}_{m,t+\ell}\upsilon_{m\ell}$, with $\tilde{Y}_{mt} = \sum_{\ell \in Z} \langle X_{t-\ell}, \psi_{m\ell}\rangle$, where $(\psi_{mk}\colon k \in \mathbb{Z})$ and $(\upsilon_{mk}\colon k \in \mathbb{Z})$ are elements of H, such that we have $\sum_{k \in \mathbb{Z}} \|\psi_{mk}\| < \infty$ and $\sum_{k \in \mathbb{Z}} \|\upsilon_{mk}\| < \infty$. Then,*

$$E\|X_t - \sum_{m=1}^{p} X_{mt}\|^2 = \sum_{m>p} \int_{-\pi}^{\pi} \lambda_m(\theta)d\theta \leq E\|X_t - \sum_{m=1}^{p} \tilde{X}_{mt}\|^2 \quad \forall p \geq 1. \tag{10}$$

Inequality (10) can be interpreted as the dynamic version of (2). Proposition 5 also suggests the proportion of variance explained by the first p dynamic FPCs as a natural measure of how well a functional time series can be represented in dimension p. This proportion is given by

$$\sum_{m \leq p} \int_{-\pi}^{\pi} \lambda_m(\theta)d\theta \Big/ E\|X_1\|^2. \tag{11}$$

3. Practical Implementation

In order to handle observed functional series in practice, some preprocessing of the data material is required. The approach which is commonly taken consists in representing a curve $x(u)$ which is observed on grid points $0 \leq u_1 < u_2 < \cdots < u_r \leq 1$ as a functional observation with a finite number of basis functions $(v_k\colon 1 \leq k \leq d)$, i.e. as $x(u) = \sum_{k=1}^{d} x_k v_k(u)$. Commonly, Fourier bases, b-splines or wavelets are used. A good choice of the basis and the number of basis functions will heavily rely on the underlying data. The coefficients x_k can be obtained, for example, via least-squares fitting or some penalized form thereof. We will not go into details here, but refer, e.g., to Ramsey and Silverman [30, Chapters 3–5]. Once such a representation is established, the analysis is reduced to that of the space $H_d = \overline{\mathrm{sp}}(v_k\colon 1 \leq k \leq d)$ spanned by the d basis functions.

In the sequel, we write $(a_{ij}\colon 1 \leq i, j \leq d)$ for a $d \times d$ matrix with entry a_{ij} in row i and column j.

3.1. Representation in finite dimension

Let $x \in H_d$, i.e. of the form $x = \mathbf{v}'\mathbf{x}$ where $\mathbf{v} = (v_1, \dots, v_d)'$ and $\mathbf{x} = (x_1, \dots, x_d)'$. We assume that the basis functions are linearly independent, but they need not be orthogonal. Any statement about x then can be expressed as an equivalent statement about $\mathbf{x}$. In particular, if $A : H_d \to H_d$ is a linear operator, then, for $x \in H_d$,

$$A(x) = \sum_{k=1}^{d} x_k A(v_k) = \sum_{k=1}^{d} \sum_{k'=1}^{d} x_k \langle A(v_k), v_k \rangle v_{k'} = \mathbf{v}'\mathfrak{A}\mathbf{x},$$

where $\mathfrak{A}' = (\langle A(v_i), v_j \rangle \colon 1 \le i, j \le d)$. We call $\mathfrak{A}$ the *corresponding matrix* of A and $\mathbf{x}$ the *corresponding vector* of x.

The following simple results are stated without proof.

Lemma 1. *Let A, B be linear operators on H_d and let $\mathfrak{A}$ and $\mathfrak{B}$ be their corresponding matrices. Then,*

(i) for any $\alpha, \beta \in \mathbb{C}$, the corresponding matrix of $\alpha A + \beta B$ is $\alpha\mathfrak{A} + \beta\mathfrak{B}$;

(ii) $A(e) = \lambda e$ iff $\mathfrak{A}\mathbf{e} = \lambda\mathbf{e}$, where $e = \mathbf{v}'\mathbf{e}$;

(iii) letting $A := \sum_{i=1}^{p} \sum_{j=1}^{p} g_{ij} v_i \otimes v_j$, $G := (g_{ij} \colon 1 \le i, j \le d)$, where $g_{ij} \in \mathbb{C}$, and $V := (\langle v_i, v_j \rangle \colon 1 \le i, j \le d)$, the corresponding matrix of A is $\mathfrak{A} = GV'$.

To obtain the corresponding matrix of the spectal density operators $\mathcal{F}_\theta^X$, first observe that, if $X_k = \sum_{i=1}^{d} X_{ki} v_i =: \mathbf{v}'\mathbf{X}_k$, then

$$C_h^X = EX_h \otimes X_0 = \sum_{i=1}^{d} \sum_{j=1}^{d} EX_{hi} X_{0j} v_i \otimes v_j.$$

Let $C_h^{\mathbf{X}} := E\mathbf{X}_h\mathbf{X}_0'$. Then, by Lemma 1 (iii), we get $\mathfrak{C}_h^X = C_h^{\mathbf{X}} V'$ as the corresponding matrix of C_h^X, and by the linearity property (i), the corresponding matrix of $\mathcal{F}_\theta^X$ is

$$\mathfrak{F}_\theta^X = \frac{1}{2\pi} \left(\sum_{h \in \mathbb{Z}} C_h^{\mathbf{X}} e^{-\mathrm{i}h\theta} \right) V'. \tag{12}$$

Assume that $\lambda_m(\theta)$ is the m-th largest eigenvalue of $\mathfrak{F}_\theta^X$, with eigenvector $\boldsymbol{\varphi}_m(\theta)$. Then $\lambda_m(\theta)$ is also an eigenvalue of $\mathcal{F}_\theta^X$ and $\mathbf{v}'\boldsymbol{\varphi}_m(\theta)$ is the corresponding eigenfunction, from which we can compute, via its Fourier expansion, the dynamic FPCs. In particular, we have

$$\phi_{mk} = \frac{\mathbf{v}'}{2\pi} \int_{-\pi}^{\pi} \boldsymbol{\varphi}_m(s) e^{-\mathrm{i}ks} ds =: \mathbf{v}'\boldsymbol{\phi}_{mk},$$

and hence

$$Y_{mt} = \sum_{k \in \mathbb{Z}} \int_0^1 \mathbf{X}_{t-k}' \mathbf{v}(u) \mathbf{v}'(u) \boldsymbol{\phi}_{mk} du = \sum_{k \in \mathbb{Z}} \mathbf{X}_{t-k}' V \boldsymbol{\phi}_{mk}. \tag{13}$$

3.2. Estimators

In view of (12), our main task is to replace the spectral density matrix

$$\mathcal{F}_\theta^{\mathbf{X}} = \frac{1}{2\pi}\sum_{h\in\mathbb{Z}} C_h^{\mathbf{X}} e^{-\mathrm{i}h\theta}$$

of the coefficient sequence $(\mathbf{X}_k)$ by some estimate. For this purpose, we can use existing multivariate techniques. Classically, we would put for $|h| < n$

$$\hat{C}_h^{\mathbf{X}} := \frac{1}{n}\sum_{k=h+1}^{n} \mathbf{X}_k\mathbf{X}_{k-h}', \quad h \geq 0, \quad \text{and} \quad \hat{C}_h^{\mathbf{X}} := \hat{C}_{-h}^{\mathbf{X}}, \quad h < 0,$$

(recall that we throughout assume that the data have been centered) and use, for example, some lag window estimator

$$\hat{\mathcal{F}}_\theta^{\mathbf{X}} := \frac{1}{2\pi}\sum_{|h|\leq q} w(h/q)\hat{C}_h^{\mathbf{X}} e^{-\mathrm{i}h\theta}, \tag{14}$$

where w is some appropriate weight function and $q = q_n \to \infty$. We refer to Chapters 10–11 in Brockwell and Davis [6] or to Politis [27]. In the examples of the following sections we shall use the Bartlett kernel $w(x) = 1 - |x|$. We then set $\hat{\mathfrak{F}}_\theta^X := \hat{\mathcal{F}}_\theta^{\mathbf{X}} V'$ and compute the eigenvalues and eigenfunctions $\hat{\lambda}_m(\theta)$ and $\hat{\boldsymbol{\varphi}}_m(\theta)$ thereof, which serve as estimators of $\lambda_m(\theta)$ and $\boldsymbol{\varphi}_m(\theta)$, respectively. We estimate the filter coefficients by $\hat{\phi}_{mk} = \frac{\mathbf{v}'}{2\pi}\int_{-\pi}^{\pi}\hat{\boldsymbol{\varphi}}_m(s)e^{\mathrm{i}ks}ds$. Usually, no analytic form of $\hat{\boldsymbol{\varphi}}_m(s)$ will be available, and one has to perform numerical integration. One may use, for example,

$$\hat{\phi}_{mk} = \frac{\mathbf{v}'}{2\pi(2N_\theta+1)}\sum_{j=-N_\theta}^{N_\theta}\hat{\boldsymbol{\varphi}}_m(\pi j/N_\theta)e^{\mathrm{i}ks} =: \mathbf{v}'\hat{\boldsymbol{\phi}}_{mk}, \quad (N_\theta \gg 1).$$

Since in our sample we only observe $X_1, \ldots, X_n$, we cannot just substitute $\hat{\boldsymbol{\phi}}_{mk}$ into (13). For example, one may define

$$\hat{Y}_{mt} = \sum_{k=-L}^{L}\mathbf{X}_{t-k}' V\hat{\boldsymbol{\phi}}_{mk}, \quad t \in \{L+1, \ldots, n-L\}. \tag{15}$$

In this case we loose the first L and the last L observations of our sample. Such boundary problems for moving averages are well known in time series analysis (e.g., for exponential smoothing) and can be partly remedied with properly weighted sums. A simple solution for obtaining $\hat{Y}_{mt}$ when $1 \leq t \leq L$ or $n-L+1 \leq t \leq n$ is to set $X_{-L+1} = \cdots = X_0 = 0$ and $X_{n+1} = \cdots = X_{n+L} = 0$.

With $\hat{\phi}_{mk}$ defined above, along the same line of argumentation as before, we obtain a p-term dynamic Karhunen-Loève expansion

$$\hat{X}_t = \sum_{m=1}^{p}\sum_{k=-L}^{L}\hat{Y}_{m,t+k}\hat{\phi}_{mk}, \quad t \in \{2L+1, \ldots, n-2L\}. \tag{16}$$

Parallel to (11), the proportion of variance explained by the first p dynamic FPCs can be estimated through

$$\mathrm{PV}_{\mathrm{dyn}}(p) := \frac{\pi}{N_\theta} \sum_{m \leq p} \sum_{j=-N_\theta}^{N_\theta} \hat{\lambda}_m(\pi j / N_\theta) \Big/ \frac{1}{n} \sum_{k=1}^{n} \|X_k\|^2.$$

Alternatively, we may use $1 - \mathrm{PV}_{\mathrm{dyn}}(p)$ or (see Proposition 5) the normalized mean squared errors

$$\mathrm{NMSE}(p, L) := \sum_{k=2L+1}^{n-2L} \|X_k - \hat{X}_k\|^2 \Big/ \sum_{k=2L+1}^{n-2L} \|X_k\|^2 \tag{17}$$

as measures for the loss of information when considering a p term dynamic KL expansion. Notice that $1 - \mathrm{PV}_{\mathrm{dyn}}(p)$ and $\mathrm{NMSE}(p, L)$ will, in general, not coincide. The latter depends on L and, from this point of view, it may look less practical than $1 - \mathrm{PV}_{\mathrm{dyn}}(p)$. On the other hand, the determination of $\hat{Y}_k$ and $\hat{X}_k$ also depends on the choice of L, and so $\mathrm{NMSE}(p, L)$ is a more 'honest' estimate which we thus recommend.

3.3. Complexity

The practical implementation of dynamic functional principal components comes along with a number of calculations. In this section, we shall summarize the numerical costs and compare them with those needed for the computation of static FPCs. Of course, efficiency and quality of algorithms play an important role in this context, and the numerical complexity we provide is related to those algorithms we used and implemented in our simulation study (Section 4) and the real data application (Section 5).

In Table 1, we list the parameters on which the computation time depends.

n	sample size
d	number of basis functions used to represent curves
N_θ	number of integration points $\theta \in [-\pi, \pi]$
q	lag window size in (14)
L	truncation level for filters in (15)
p	number of dynamic FPCs to be computed

Table 1: Parameters involved for computing DFPCs.

Table 2 displays the building blocks required for the computation of dynamic FPCs, along with the numerical complexity (number of summations, multiplications and storing) involved. All these quantities have to be computed over a set of different parameter values, which introduces an additional multiplicative factor, shown in the last column of Table 2. Objects obtained in step $i-1$ are stored and can be used for step i, for $2 \leq i \leq 5$. The computation time $\mathrm{CT}_{\mathrm{dyn}}$ for the dynamic procedure is thus $O(d \times [ndq + dqN_\theta + pd^2N_\theta + LpN_\theta + Lnpd])$. In comparison, the computation time $\mathrm{CT}_{\mathrm{stat}}$ for static FPCA is $O(d \times [nd + pd^2 + np])$. In practice, q and L will be adapted to the sample size n. From our computational experience, we would recommend to put $q = O(\sqrt{n})$ and $L = O(\sqrt{n})$, while p is usually small and fixed. Then we have $\mathrm{CT}_{\mathrm{dyn}} = O(d^2 \times \max\{n^{3/2}, \sqrt{n}N_\theta, dN_\theta\})$. One may conclude that $\mathrm{CT}_{\mathrm{dyn}} \ll \max\{\sqrt{n}, N_\theta\}\mathrm{CT}_{\mathrm{stat}}$.

step	object	complexity	multiplicity
1	$\hat{C}_h^{\mathbf{X}}$	$O(nd^2)$	q
2	$\hat{\mathcal{F}}_\theta^{\mathbf{X}}$	$O(d^2q)$	N_θ
3	$\hat{\boldsymbol{\varphi}}_m(\theta)$	$O(d^3)$	pN_θ
4	$\hat{\phi}_{mk}$	$O(N_\theta d)$	pL
5	Y_{mk}	$O(d^2L)$	pn

Table 2: Computation complexity for obtaining the different objects required in our procedure. These quantities have to be computed over a set of different parameter values, the impact of which is reflected in the multiplicative factors shown in the third column.

4. Simulation study

In this simulation study, we compare the performance of dynamic PCA with that of static PCA as follows. For a given time series (X_t), we perform a static and a dynamic FPC analysis. From the resulting scores, we recover two functional series $(\hat{X}_t^{\mathrm{stat}})$ and $(\hat{X}_t^{\mathrm{dyn}})$ using the static and dynamic Karhunen-Loève expansions, respectively. Performance is then measured in terms of the respective normalized mean square errors $E\|X_t - \hat{X}_t^{\mathrm{stat}}\|^2/E\|X_t\|^2$ and $E\|X_t - \hat{X}_t^{\mathrm{dyn}}\|^2/E\|X_t\|^2$. The latter quantity can be estimated by $\mathrm{NMSE}(p, L)$ or by $1 - \mathrm{PV}_{\mathrm{dyn}}(p)$. For the static FPCA, we use the estimate $1 - \mathrm{PV}_{\mathrm{stat}}(p)$, where $\mathrm{PV}_{\mathrm{stat}}(p)$ is the proportion of variance explained by the first p static FPCs.

For computations we employed the statistical software `R` along with the `fda` package. The data was represented as discussed in Section 3.1 using Fourier basis functions $(v_i\colon 1 \leq i \leq d)$, where $d = 5, 11, 21$. We then set $H = \overline{\mathrm{sp}}(v_i\colon 1 \leq i \leq d)$. In each run we sample a matrix Ψ with i.i.d. standard normal entries and normalize it to $\|\Psi\| = \kappa$, where $\kappa = 0.1, 0.3, 0.6, 0.9$. This matrix is then used as the corresponding matrix of an operator (with slight abuse of notation, we denote the operator also by Ψ) on H. With the operator we generate 400 observations from an autoregressive Hilbertian process of order 1, defined by $X_t = \Psi(X_{t-1}) + \varepsilon_t$. The noise (ε_t) is i.i.d. Gaussian, obtained as a linear combination of the functions $(v_i\colon 1 \leq i \leq d)$ with i.i.d. standard normal coefficients.

Efficiency of our method also relies on the estimation of the spectral density operator. We follow the methodology introduced in Section 3 and use a Barlett kernel and $q = 20$ in (14). The numerical integration for obtaining $\hat{\phi}_{mk}$ is performed on the basis of 400 equidistant integration points. We test truncation levels $L = 5, 10, 15, 20, 25$ for the filters in (15).

For each choice of d and L, the experiment was repeated 200 times. Results are presented as boxplots in Figure 2. The dashed lines correspond to the average of $1 - \mathrm{PV}_{\mathrm{dyn}}(p)$, $p = 1, 2, 3$, while the solid lines correspond to $1 - \mathrm{PV}_{\mathrm{stat}}(p)$.

We see that in this sense dynamic FPCA quite significantly outperforms static FPCA. As one can expect, performance increases with the dependence coefficient κ.

5. Illustrative data example

In this section, we compute and interpret the first dynamic FPC score sequence of daily air pollution curves and draw a comparison with its static counterpart. The observations are half-hourly measurements of the concentration (measured in $\mu g m^{-3}$) of particulate matter with an aerodynamic diameter of less than $10\mu m$, abbreviated as `PM10`, in ambient air taken in Graz, Austria from October 1, 2010 until March 31, 2011. Following Stadlober et al. [33], a square-root transformation was applied to the data in order to stabilize the variance and avoid heavy-tailed observations. The data have been already explored in Aue et al. [1] in the context of curve prediction. Following their approach, we remove some outliers and a seasonal (weekly) pattern coming from different traffic intensities on business days and weekends. Then we use the software `R` to transform the data to functional data. The only difference in our approach is that we use 15 Fourier basis functions instead of b-splines. This simplifies our computations, but easily can be changed. Eventually, 175 daily functional observations, say, $X_1, \ldots, X_{175}$, were obtained, roughly representing one winter season for which pollution levels are known to be high. They are displayed in Figure 3.

Next we compute the (estimated) first dynamic functional PC score sequence $(\hat{Y}_{1t}^{\mathrm{dyn}} \colon\ 1 \leq\ t \leq\ 175)$. To this end, we first center the data by their empirical mean $\hat{\mu}(u)$ and then follow the procedure introduced in Section 3; in particular, we set $q = 15$ in (14) and use the Barlett kernel to obtain an estimator for the spectral density operator. From this, we obtain the estimated filter elements $\hat{\phi}_{1t}$. Since $\|\hat{\phi}_{1t}\|$ seems to converge to zero fast, we simply set $L = 10$ in (15). To obtain DFPC scores Y_{1t}^{dyn} for the case $1 \leq t \leq 10$ and $166 \leq t \leq 175$ we set $x_{-9} = \cdots = x_0 = \hat{\mu}$ and $x_{176} = \cdots = x_{185} = \hat{\mu}$. The corresponding time series $(\hat{Y}_{1t}^{\mathrm{dyn}} \colon 1 \leq t \leq 175)$ is shown in Figure 4. We shall focus here on one component only, since the first dynamic FPC already explains 80.2% of the variance. This should be compared to 73.8% explained by the first static FPC.

Figure 4 shows that the static score sequence $(\hat{Y}_{1t}^{\mathrm{stat}} \colon 1 \leq t \leq 175)$ is almost identical to the dynamic one. This is remarkable, as they have been computed from quite different methods. To get some interpretation, let us analyze the first static sample FPC $\hat{v}_1(u)$, say, and the DFPC filters $\hat{\phi}_{1t}(u)$. They are displayed in Figure 5. We see that $\hat{v}_1(u) \approx 1$ for all $u \in [0, 1]$, and hence the FPC score $Y_{1t}^{\mathrm{stat}} = \int_0^1 (X_t(u) - \hat{\mu}(u))\hat{v}_1(u)du$ roughly is the average deviation of $X_t(u)$ from

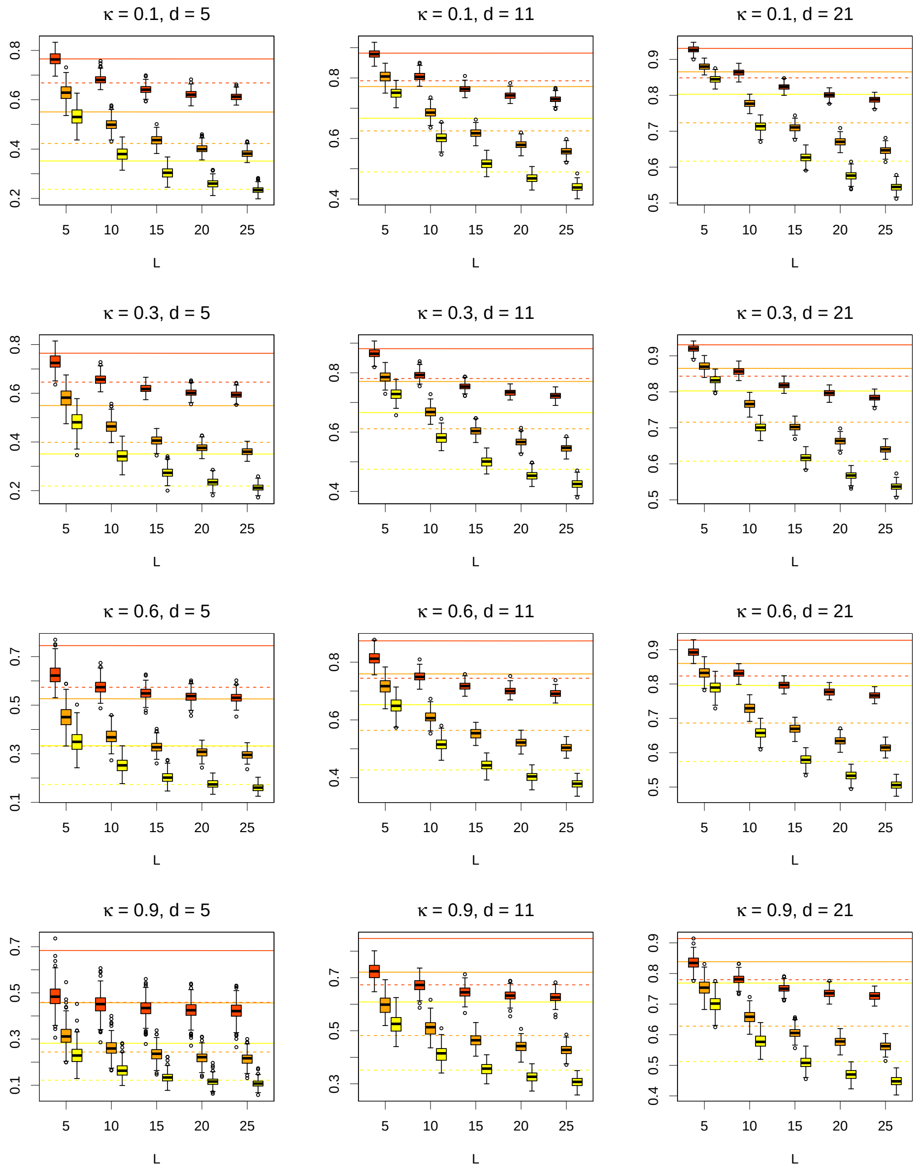

Figure 2: Boxplots: $\mathrm{NMSE}(p, L)$ with $L = 5, 10, 15, 20, 25$, where for each L red, orange and yellow boxplot corresponds to $p = 1, 2, 3$, respectively. Solid lines (red, orange and yellow): $1 - \mathrm{PV}_{\mathrm{stat}}(p)$ with $p = 1, 2, 3$, respectively. Dashed lines: $1 - \mathrm{PV}_{\mathrm{dyn}}(p)$ with $p = 1, 2, 3$, respectively.

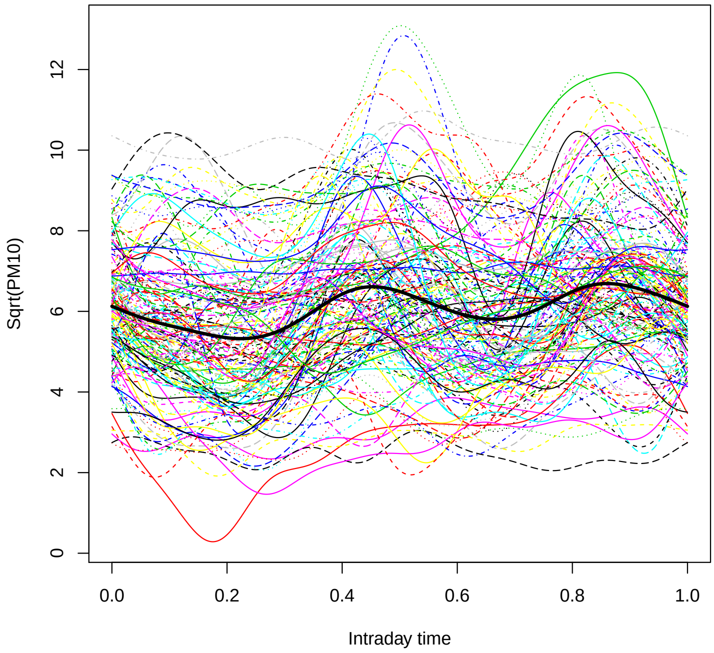

Figure 3: We display $x_t(u)$, $1 \leq t \leq 175$, where $x_t(u)$ are the square-root transformed and d etrended daily functional observations of PM10, represented with 15 Fourier basis functions. The solid black line represents the sample mean curve $\hat{\mu}(u)$.

the mean. The effect of a large (small) first score corresponds to a large (small) daily average of $\sqrt{\texttt{PM10}}$. In view of the similarity of Y_{1t}^{dyn} and Y_{1t}^{stat}, it is possible to attribute the same meaning to the dynamic FPC scores. However, regarding the dynamic KL expansion, dynamic FPC scores should be interpreted sequentially and not in a static way. To this end, let us take advantage of the fact that all functions $\hat{\phi}_{1t}$, $|t| > 1$, are close to zero (see Figure 5) and thus, in the approximation by a single-term dynamic KL expansion, we roughly have

$$X_t(u) \approx \hat{\mu}(u) + \hat{Y}_{1,t-1}^{\mathrm{dyn}} \hat{\phi}_{1,-1}(u) + \hat{Y}_{1t}^{\mathrm{dyn}} \hat{\phi}_{1,0}(u) + \hat{Y}_{1,t+1}^{\mathrm{dyn}} \hat{\phi}_{1,1}(u).$$

This suggests to study the effect of triplets $(\hat{Y}_{1,t-1}^{\mathrm{dyn}}, \hat{Y}_{1t}^{\mathrm{dyn}}, \hat{Y}_{1,t+1}^{\mathrm{dyn}})$ of consecutive scores on the pollution level of day t, which can be done by adding the functions

$$\mathrm{eff}(\delta_1, \delta_2, \delta_3) := \delta_1 \hat{\phi}_{1,-1}(u) + \delta_2 \hat{\phi}_{1,0}(u) + \delta_3 \hat{\phi}_{1,1}(u), \quad \delta_i = \mathrm{const} \times \pm 1,$$

to the overall mean curve $\hat{\mu}(u)$. We do this in Figure 6 with $\delta_i = \pm 1$. For instance, the upper left panel shows $\hat{\mu}(u) + \mathrm{eff}(-1, -1, -1)$; this corresponds to the effect of three subsequent small DFPC scores. Not surprisingly, they result in a negative shift of the mean curve. The second panel from the left in top row shows $\hat{\mu}(u) + \mathrm{eff}(-1, -1, +1)$. The picture is similar as before, but now the level increases as u

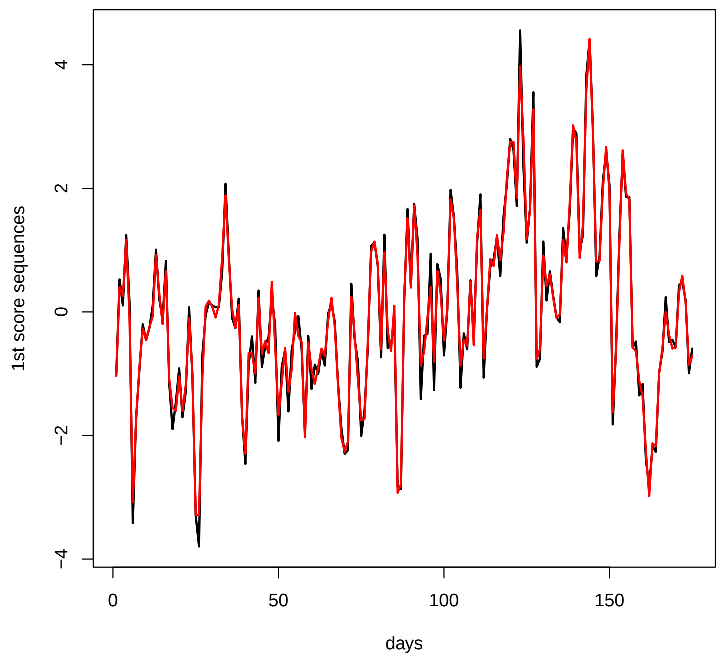

Figure 4: The sequence of the first static FPC scores (red) and the dynamic ones (black).

approaches 1. This has a simply explanation. A large value of $\hat{Y}^{\mathrm{dyn}}_{1,t+1}$, implies a large average concentration of $\sqrt{\mathtt{PM10}}$ on day $t+1$, and since the pollution curves are highly correlated at the transition from day t to day $t+1$, this should indeed be reflected by an increase of $\sqrt{\mathtt{PM10}}$ towards the end of day t. By the same line of argumentation, it becomes clear why the pollution level is low for the rest of the day. Similar interpretations can be given for the other panels of Figure 6.

It is interesting to observe that, in this example, the first dynamic FPC seems to take the role of the first two static FPCs. The second static FPC (see the left panel in Figure 5) can be interpreted as an intraday trend effect; if the second static score of day t is large (small), then $X_t(u)$ is increasing (decreasing) over $u \in [0,1]$. However, since we are working with sequentially dependent data, we can get information about such a trend from future and past observations, too. This is exemplified in Figure 1 of Section 1. It shows the ten consecutive curves $x_{71}(u)-\hat{\mu}(u),\ldots,x_{80}(u)-\hat{\mu}(u)$ (left panel) and compares them to the single-term static (middle panel) and the single-term dynamic KL expansion (right panel). We see that the dynamic KL version not only recovers the level, but also the intraday trend.

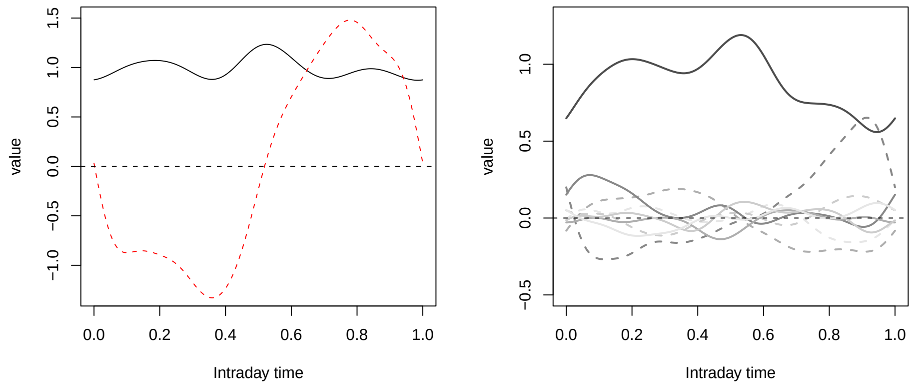

Figure 5: First static FPC $\hat{v}_1(u)$ (solid line left), and second static FPC $\hat{v}_2(u)$ (dashed line left), and the filters corresponding to first dynamic FPC (right). The filters $\hat{\phi}_{1t}(u)$, $t \geq 1$, are dashed and the filters $\hat{\phi}_{1t}(u)$, $t \leq 0$, are solid. The larger $|t|$, the lighter the curve.

6. Conclusion

Functional principal component analysis is taking a leading role in the functional data literature. As an extremely effective tool for dimension reduction, it is useful for empirical data analysis as well as for many FDA-related methods, like functional linear models. A frequent situation in practice is that functional data are observed sequentially and are serially dependent. For example, this occurs when observations stem from a continuous time process which is segmented into smaller units, e.g., days. In such cases, classical static FPCs still can be consistently estimated, but, in contrast to the i.i.d. setup, they will not lead to an adequate dimension reduction technique.

In this paper, we have proposed a *dynamic version of functional PCA* which takes into account a potential serial dependence of the functional observations. In the special case of uncorrelated data, the dynamic methodology reduces to the usual static FPCA. We have complemented the methodology with (i) practical guidelines for implementation, (ii) simulations, (iii) a toy example with `PM10` pollution data and (iv) a rigorous mathematical theory, including some asymptotics. Our empirical work shows that dynamic FPCs have a clear edge over static FPCs in terms of their ability to represent dependent functional data in small dimension. While we have presented the method for functional (L^2-valued) data, our proofs are general and cover the theory for separable Hilbert spaces.

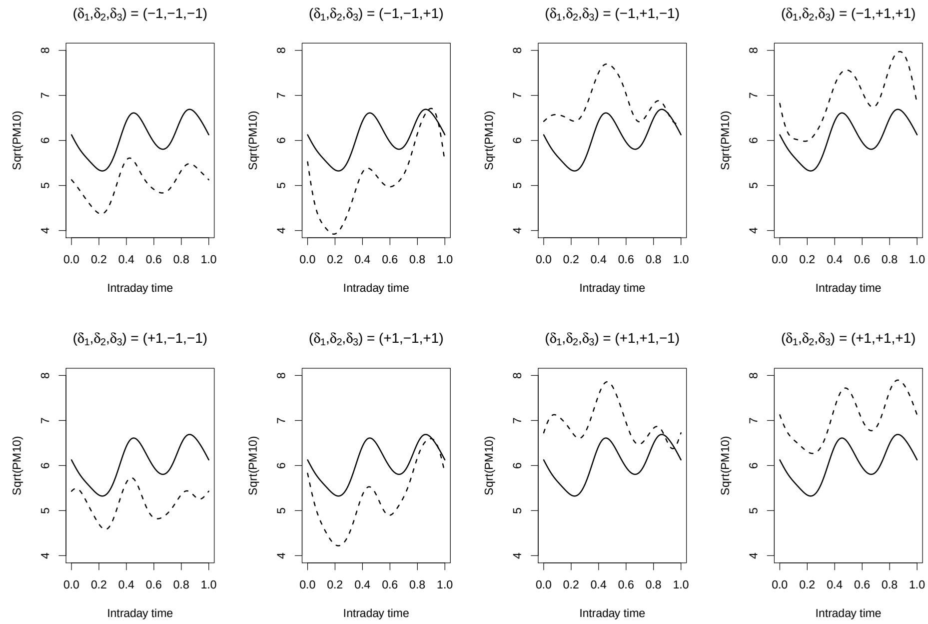

Figure 6: Mean curve $\hat{\mu}(u)$ (solid line) and $\hat{\mu}(u) + \text{eff}(\delta_1, \delta_2, \delta_3)$ with $\delta_i = \pm 1$.

A. General Methodology and Proofs

A.1. Hilbertian Framework

In this subsection, we give a mathematically rigorous description of the methodology introduced in Section 2.1. We adopt a more general framework which can be specialized to the functional setup of Section 2.1. Throughout H denotes some (complex) Hilbert space. We work in complex spaces, since our theory is based on a frequency domain analysis. Nevertheless, all our functional time series observations X_t are assumed to be real-valued functions. A crucial structural assumption that we impose is that H is separable, i.e. possesses a countable orthonormal basis (ONB).

Linear operators. We consider the class $\mathcal{L}(H, H')$ of bounded linear operators between two Hilbert spaces H and H'. With a slight abuse of notation, we use $\|\cdot\|$ and $\langle\cdot,\cdot\rangle$, for norm and inner product on both, H and H'. For $\Psi \in \mathcal{L}(H, H')$, the operator norm is defined as $\|\Psi\|_{\mathcal{L}} := \sup_{\|x\|\leq 1} \|\Psi(x)\|$. The simplest operators can be defined via a tensor product $v \otimes w$; then $v \otimes w(z) := v\langle z, w\rangle$. Every operator $\Psi \in \mathcal{L}(H, H')$ possesses an adjoint $\Psi^* \in \mathcal{L}(H', H)$ which satisfies $\langle \Psi(x), y\rangle = \langle x, \Psi^*(y)\rangle$ for all $x \in H$ and $y \in H'$. It holds that $\|\Psi^*\|_{\mathcal{L}} = \|\Psi\|_{\mathcal{L}}$. If $H = H'$, then Ψ is called *self-adjoint* if $\Psi = \Psi^*$. It is called non-negative definite if $\langle \Psi x, x\rangle \geq 0$ for all $x \in H$.

A linear operator Ψ is said to be *Hilbert-Schmidt* if we have $\|\Psi\|_{\mathcal{S}}^2 := \sum_{k\geq 1} \|\Psi(v_k)\|^2 < \infty$ for some ONB $(v_k\colon k \geq 1)$ of H. Then $\|\Psi\|_{\mathcal{S}}$ defines a norm, the so-called Hilbert-Schmidt norm of Ψ. It bounds the operator norm: $\|\Psi\|_{\mathcal{L}} \leq \|\Psi\|_{\mathcal{S}}$, and can be shown to be independent of the choice of the ONB. Every Hilbert-Schmidt operator is compact. The class of Hilbert-Schmidt operators between H and H' defines again a separable Hilbert space $\mathcal{H}$ with inner product $\langle \Psi, \Psi' \rangle_{\mathcal{S}} := \sum_{k \geq 1} \langle \Psi(v_k), \Psi'(v_k) \rangle$.

If $\Psi \in \mathcal{L}(H, H')$ and $\Upsilon \in \mathcal{L}(H'', H)$ then $\Psi\Upsilon$ is the operator which maps $x \in H''$ to $\Psi(\Upsilon(x)) \in H'$. Assume that Ψ is a compact operator in $\mathcal{L}(H, H')$ and let (s_j^2) be the eigenvalues of $(\Psi^*)\Psi$. Then Ψ is said to be *trace class* if $\|\Psi\|_{\mathcal{T}} := \sum_{j \geq 1} s_j < \infty$. In this case $\|\Psi\|_{\mathcal{T}}$ defines a norm, the so-called *Schatten 1-norm*. We have $\|\Psi\|_{\mathcal{S}} \leq \|\Psi\|_{\mathcal{T}}$, and hence any trace-class operator is Hilbert-Schmidt. For self-adjoint non-negative operators, it holds that $\|\Psi\|_{\mathcal{T}} = \mathrm{tr}(\Psi) := \sum_{k \geq 1} \langle \Psi(v_k), v_k \rangle$. If $\tilde{\Psi}\tilde{\Psi} = \Psi$, then we have $\mathrm{tr}(\Psi) = \|\tilde{\Psi}\|_{\mathcal{S}}^2$.

For further background on the theory of linear operators we refer to [12].

Random sequences in Hilbert spaces. All random elements that appear in the sequel are assumed to be defined on a common probability space $(\Omega, \mathcal{A}, P)$. We write $X \in L_H^p(\Omega, \mathcal{A}, P)$ (in short, $X \in L_H^p$) if $E\|X\|^p < \infty$. Every element $X \in L_H^1$ possesses an expectation, which is the unique $\mu \in H$ satisfying $E\langle X, y \rangle = \langle \mu, y \rangle$ for all $y \in H$. Provided X and Y are in L_H^2, we can define the cross-covariance operator as $C_{XY} := E(X - \mu_X) \otimes (Y - \mu_Y)$, where μ_X and μ_Y are the expectations of X and Y, respectively. We have that $\|C_{XY}\|_{\mathcal{T}} \leq E\|(X - \mu_X) \otimes (Y - \mu_Y)\|_{\mathcal{T}} = E\|X - \mu_X\|\|Y - \mu_Y\|$, and so these operators are trace-class. An important specific role is played by the covariance operator C_{XX}. This operator is non-negative definite, self-adjoint with $\mathrm{tr}(C_{XX}) = E\|X - \mu_X\|^2$. We call an H-valued process (X_t) *(weakly) stationary*, if $(X_t) \in L_H^2$ and if EX_t and $C_{X_{t+h}X_t}$ do not depend on t. In this case, we write C_h^X, or shortly C_h, for $C_{X_{t+h}X_t}$ if it is clear to which process it belongs. Two weakly stationary processes (X_t) and (Y_t) are called *costationary* if $C_{X_{t+h}Y_t}$ does not depend on t. Then we write C_h^{XY} for the covariance operator $C_{X_{t+h}Y_t}$.

Many useful results on random processes in Hilbert spaces or more general Banach spaces are collected in Chapters 1 and 2 of [7].

Fourier series in Hilbert spaces. For $p \geq 1$, consider the space $L_H^p([-\pi, \pi])$, that is the space of measurable mappings $x : [-\pi, \pi] \to H$ which satisfy $\int_{-\pi}^{\pi} \|x(\theta)\|^p d\theta < \infty$. For $p = 2$ this space is again a Hilbert space, with inner product

$$(x, y) := \frac{1}{2\pi} \int_{-\pi}^{\pi} \langle x(\theta), y(\theta) \rangle d\theta$$

and norm $\|x\|_2 = \sqrt{(x, x)}$. One can show (see e.g. [7, Lemma 1.4]) that, for any $x \in L_H^1([-\pi, \pi])$, there exists a unique element $I(x) \in H$ which satisfies

$$\int_{-\pi}^{\pi} \langle x(\theta), v \rangle d\theta = \langle I(x), v \rangle \quad \forall v \in H. \tag{18}$$

Then we define $\int_{-\pi}^{\pi} x(\theta) d\theta := I(x)$.

For $x \in L_H^2([-\pi, \pi])$ we can now set the k-th Fourier coefficient equal to

$$f_k := \frac{1}{2\pi} \int_{-\pi}^{\pi} x(\theta) e^{-\mathrm{i}k\theta} d\theta, \quad k \in \mathbb{Z}. \tag{19}$$

Below we write e_k for the function $\theta \mapsto e^{\mathrm{i}k\theta}$, $\theta \in [-\pi, \pi]$.

Proposition 6. *Suppose $x \in L^2_H([-\pi,\pi])$ and define f_k by equation* (19)*. Then, the sequence $S_n := \sum_{k=-n}^n f_k e_k$ has a mean square limit in $L^2_H([-\pi,\pi])$. If we denote the limit by S, then $x(\theta) = S(\theta)$ for almost all θ.*

Proof. Let $0 < m < n$ and notice that

$$\begin{aligned}\|S_n - S_m\|_2^2 &= \left(\sum_{m\le|k|\le n} f_k e_k, \sum_{m\le|\ell|\le n} f_\ell e_\ell \right) \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi} \sum_{m\le|k|\le n} \sum_{m\le|\ell|\le n} \langle f_k, f_\ell\rangle e^{\mathbb{i}(k-\ell)\theta} d\theta = \sum_{m\le|k|\le n} \|f_k\|^2.\end{aligned}$$

To prove the first statement, we need to show that (S_n) defines a Cauchy sequence in $L^2_H([-\pi,\pi])$, which follows if we show that $\sum_{k\in\mathbb{Z}} \|f_k\|^2 < \infty$. We use the fact that, for any $v \in H$, the function $\langle x(\theta), v\rangle$ belongs to $L^2([-\pi,\pi])$. Then, by Parseval's identity and (18), we have, for any $v \in H$,

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} |\langle x(\theta), v\rangle|^2 d\theta = \sum_{k\in\mathbb{Z}} \left(\frac{1}{2\pi}\int_{-\pi}^{\pi} \langle x(s), v\rangle e^{-\mathbb{i}ks} ds \right)^2 = \sum_{k\in\mathbb{Z}} |\langle f_k, v\rangle|^2.$$

Let $(v_k\colon k \ge 1)$ be an ONB of H. Then, by the last result and Parseval's identity again, it follows that

$$\begin{aligned}\|x\|_2^2 &= \frac{1}{2\pi}\int_{-\pi}^{\pi} \sum_{\ell\ge1} |\langle x(\theta), v_\ell\rangle|^2 d\theta = \frac{1}{2\pi}\sum_{\ell\ge1}\int_{-\pi}^{\pi} |\langle x(\theta), v_\ell\rangle|^2 d\theta \\ &= \sum_{\ell\ge1}\sum_{k\in\mathbb{Z}} |\langle f_k, v_\ell\rangle|^2 = \sum_{k\in\mathbb{Z}} \|f_k\|^2.\end{aligned}$$

As for the second statement, we conclude from classical Fourier analysis results that, for each $v \in H$,

$$\lim_{n\to\infty} \frac{1}{2\pi}\int_{-\pi}^{\pi} \left(\langle x(\theta), v\rangle - \sum_{k=-n}^{n} \left(\frac{1}{2\pi}\int_{-\pi}^{\pi} \langle x(s), v\rangle e^{-\mathbb{i}ks} ds \right) e^{\mathbb{i}k\theta} \right)^2 d\theta = 0.$$

Now, by definition of S_n, this is equivalent to

$$\lim_{n\to\infty} \frac{1}{2\pi}\int_{-\pi}^{\pi} \langle x(\theta) - S_n(\theta), v\rangle^2 \, d\theta = 0, \quad \forall v \in H.$$

Combined with the first statement of the proposition and

$$\begin{aligned}\int_{-\pi}^{\pi} \langle x(\theta) - S(\theta), v\rangle^2 \, d\theta \le{}& 2\int_{-\pi}^{\pi} \langle x(\theta) - S_n(\theta), v\rangle^2 \, d\theta \\ &+ 2\|v\|^2 \int_{-\pi}^{\pi} \|S_n(\theta) - S(\theta)\|^2 d\theta,\end{aligned}$$

this implies that

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} \langle x(\theta) - S(\theta), v\rangle^2 \, d\theta = 0, \quad \forall v \in H. \tag{20}$$

Let (v_i) be an ONB of H and $A_i := \{\theta \in [-\pi, \pi] \colon \langle x(\theta) - S(\theta), v_i \rangle \neq 0\}$. By (20), we have that $\lambda(A_i) = 0$ (λ denotes the Lebesgue measure), and hence $\lambda(A) = 0$ for $A = \cup_{i \geq 1} A_i$. Consequently, since (v_i) define an ONB, for any $\theta \in [-\pi, \pi] \backslash A$, we have $\langle x(\theta) - S(\theta), v \rangle = 0$ for all $v \in H$, which in turn implies that $x(\theta) - S(\theta) = 0$. □

A.2. On the Spectral Density Operator

Assume that the H-valued process $(X_t \colon t \in \mathbb{Z})$ is stationary with lag h autocovariance operator C_h and spectral density operator

$$\mathcal{F}_\theta^X := \frac{1}{2\pi} \sum_{h \in \mathbb{Z}} C_h e^{-\mathrm{i} h \theta}.$$

In order to guarantee convergence of this series, we *tacitly impose assumption* (4) *throughout this section.* It can be easily seen that the operator $\mathcal{F}_\theta^X$ is self-adjoint, non-negative definite and Hilbert-Schmidt. Below, we introduce a weak dependence assumption established in [17] from which we can derive a sufficient condition for (4).

Definition 3 (L^p–m–approximability)**.** *A random H–valued sequence $(X_n \colon n \in \mathbb{Z})$ is called L^p–m–approximable if it can be represented as*

$$X_n = f(\delta_n, \delta_{n-1}, \delta_{n-2}, ...)$$

where the δ_i's are i.i.d. elements taking values in some measurable space S and f is a measurable function $f : S^\infty \to H$. Moreover, if $\delta_1', \delta_2', ...$ are independent copies of $\delta_1, \delta_2, ...$ defined on the same measurable space S, then, for

$$X_n^{(m)} := f(\delta_n, \delta_{n-1}, \delta_{n-2}, ..., \delta_{n-m+1}, \delta_{n-m}', \delta_{n-m-1}', ...),$$

we have

$$\sum_{m=1}^{\infty} (E\|X_m - X_m^{(m)}\|^p)^{1/p} < \infty. \tag{21}$$

Hörmann and Kokoszka [17] show that this notion is widely applicable to linear and non-linear functional time series. One of its main advantages is that it is a purely moment-based dependence measure that can be easily verified in many special cases.

Proposition 7. *Assume that (X_t) is L^2–m–approximable. Then* (4) *holds and the operators $\mathcal{F}_\theta^X$, $\theta \in [-\pi, \pi]$, are trace-class.*

Instead of Assumption (4), Panaretos and Tavakoli [26] impose for the definition of a spectral density operator summability of C_h in Schatten 1-norm, i.e. $\sum_{h \in \mathbb{Z}} \|C_h\|_\mathcal{T} < \infty$. Under such slightly more stringent assumption, it immediately follows that the resulting spectral operator is trace-class. The verification of convergence may, however, be a bit delicate. At least, we could not find a simple criterion as in Proposition 7.

Proof of Proposition 7. Without loss of generality, we assume that $EX_0 = 0$. By independence of X_0 and $X_h^{(h)}$, $h \geq 1$, we have

$$\|C_h\|_{\mathcal{S}} = \|EX_0 \otimes (X_h - X_h^{(h)})\|_{\mathcal{S}} \leq (E\|X_0\|^2)^{1/2}(E\|X_h - X_h^{(h)}\|^2)^{1/2}.$$

The first statement of the proposition follows.

Fix θ. Since $\mathcal{F}_\theta^X$ is non-negative and self-adjoint, it is trace class if and only if

$$\mathrm{tr}(\mathcal{F}_\theta^X) = \sum_{m\geq 1} \langle \mathcal{F}_\theta^X(v_m), v_m\rangle < \infty \tag{22}$$

for some ONB (v_m) of H. The trace can be shown to be independent of the choice of the basis. Define $V_{n,\theta} = (2\pi n)^{-1/2} \sum_{k=1}^n X_k e^{\mathrm{i}k\theta}$ and note that, by stationarity,

$$\mathcal{F}_{n,\theta}^X := EV_{n,\theta} \otimes V_{n,\theta} = \frac{1}{2\pi} \sum_{|h|<n} \left(1 - \frac{|h|}{n}\right) EX_0 \otimes X_{-h} e^{-\mathrm{i}h\theta}.$$

It is easily verified that the operators $\mathcal{F}_{n,\theta}^X$ again are non-negative and self-adjoint. Also note that, by the triangular inequality,

$$\|\mathcal{F}_{n,\theta}^X - \mathcal{F}_\theta^X\|_{\mathcal{S}} \leq \sum_{|h|<n} \frac{|h|}{n} \|C_h\|_{\mathcal{S}} + \sum_{|h|\geq n} \|C_h\|_{\mathcal{S}}.$$

By application of (4) and Kronecker's lemma, it easily follows that the latter two terms converge to zero. This implies that $\mathcal{F}_{n,\theta}^X(v)$ converges in norm to $\mathcal{F}_\theta^X(v)$, for any $v \in H$.

Choose $v_m = \varphi_m(\theta)$. Then, by continuity of the inner product and the monotone convergence theorem, we have

$$\begin{aligned}\sum_{m\geq 1} \langle \mathcal{F}_\theta^X(\varphi_m(\theta)), \varphi_m(\theta)\rangle &= \sum_{m\geq 1} \lim_{n\to\infty} \langle \mathcal{F}_{n,\theta}^X(\varphi_m(\theta)), \varphi_m(\theta)\rangle \\ &= \lim_{n\to\infty} \sum_{m\geq 1} \langle \mathcal{F}_{n,\theta}^X(\varphi_m(\theta)), \varphi_m(\theta)\rangle.\end{aligned}$$

Using the fact that the $\mathcal{F}_{n,\theta}^X$'s are self-adjoint and non-negative, we get

$$\begin{aligned}\sum_{m\geq 1} \langle \mathcal{F}_{n,\theta}^X(\varphi_m(\theta)), \varphi_m(\theta)\rangle &= \mathrm{tr}(\mathcal{F}_{n,\theta}^X) = E\|V_n\|^2 \\ &= \frac{1}{2\pi} \sum_{|h|<n} \left(1 - \frac{|h|}{n}\right) E\langle X_0, X_h\rangle e^{-\mathrm{i}h\theta}.\end{aligned}$$

Since $|E\langle X_0, X_h\rangle| = |E\langle X_0, X_h - X_h^{(h)}\rangle|$, by the Cauchy-Schwarz inequality,

$$\sum_{h\in\mathbb{Z}} |E\langle X_0, X_h\rangle| \leq \sum_{h\in\mathbb{Z}} (E\|X_0\|^2)^{1/2}(E(X_h - X_h^{(h)})^2)^{1/2} < \infty,$$

and thus the dominated convergence theorem implies that

$$\mathrm{tr}(\mathcal{F}_\theta^X) = \frac{1}{2\pi} \sum_{h\in\mathbb{Z}} E\langle X_0, X_h\rangle e^{-\mathrm{i}h\theta} \leq \sum_{h\in\mathbb{Z}} |E\langle X_0, X_h\rangle| < \infty.$$

□

The eigendecomposition of $\mathcal{F}_\theta^X$ gives

$$\mathcal{F}_\theta^X = \sum_{m\geq 1} \lambda_m(\theta)\varphi_m(\theta)\otimes\varphi_m(\theta), \quad \theta\in[-\pi,\pi],$$

where $\lambda_1(\theta)\geq\lambda_2(\theta)\geq\cdots$ are the eigenvalues and $\varphi_m(\theta)$ the corresponding eigenfunctions of $\mathcal{F}_\theta^X$. We require $\|\varphi_m(\theta)\|=1$ and hence, if $\lambda_m(\theta)$ has multiplicity 1, then $\varphi_m(\theta)$ is unique up to some rotation $e^{\mathrm{i}\omega}$, $\omega\in[-\pi,\pi]$. Let $\overline{x}$ be the conjugate element of x, i.e. $\langle x,z\rangle=\langle z,\overline{x}\rangle$ for all $z\in H$. Then x is real-valued iff $x=\overline{x}$.

Proposition 8. *Let $\mathcal{F}_\theta^X$ be the spectral density operator of the stationary sequence (X_t) for which the summability condition (4) holds. Let $\lambda_1(\theta)\geq\lambda_2(\theta)\geq\cdots$ denote its eigenvalues and $\varphi_m(\theta)$ be the corresponding eigenfunctions. Then, (a) the functions $\theta\mapsto\lambda_m(\theta)$ are continuous; (b) if we strengthen condition (4) to $\sum_{h\in\mathbb{Z}}|h|\|C_h\|_{\mathcal{S}}<\infty$, the $\lambda_m(\theta)$'s are Lipschitz-continuous functions of θ; (c) assuming that (X_t) is real-valued, for each $\theta\in[-\pi,\pi]$, $\lambda_m(\theta)=\lambda_m(-\theta)$ and $\varphi_m(\theta)=\overline{\varphi_m(-\theta)}$.*

Proof. We have (see e.g. [12], p. 186) that the dynamic eigenvalues satisfy $|\lambda_m(\theta)-\lambda_m(\theta')|\leq\|\mathcal{F}_\theta^X-\mathcal{F}_{\theta'}^X\|_{\mathcal{S}}$. Now,

$$\|\mathcal{F}_\theta^X-\mathcal{F}_{\theta'}^X\|_{\mathcal{S}}\leq\sum_{h\in\mathbb{Z}}\|C_h\||e^{-\mathrm{i}h\theta}-e^{-\mathrm{i}h\theta'}|.$$

The summability condition (4) implies continuity, hence part (a) of the proposition. Using $|e^{-\mathrm{i}h\theta}-e^{-\mathrm{i}h\theta'}|\leq|h||\theta-\theta'|$ yields part (b).

To prove (c), we observe that, for any $\theta\in[-\pi,\pi]$,

$$\lambda_m(\theta)\varphi_m(\theta)=\mathcal{F}_\theta^X(\varphi_m(\theta))=\frac{1}{2\pi}\sum_{h\in\mathbb{Z}}EX_h\langle\varphi_m(\theta),X_0\rangle e^{-\mathrm{i}h\theta}.$$

Since the eigenvalues $\lambda_m(\theta)$ are real, we obtain, by computing the complex conjugate of the above equalities,

$$\lambda_m(\theta)\overline{\varphi_m(\theta)}=\frac{1}{2\pi}\sum_{h\in\mathbb{Z}}EX_h\langle\overline{\varphi_m(\theta)},X_0\rangle e^{\mathrm{i}h\theta}=\mathcal{F}_{-\theta}^X(\overline{\varphi_m(\theta)}).$$

This shows that $\lambda_m(\theta)$ and $\overline{\varphi_m(\theta)}$ are eigenvalue and eigenfunction of $\mathcal{F}_{-\theta}^X$ and they must correspond to a pair $(\lambda_n(-\theta),\varphi_n(-\theta))$; (c) follows. □

Remark 1. *The eigenfunctions $\varphi_m(\theta)$ are unique up to multiplication with a number lying on the complex unit circle. Writing $\varphi_m(\theta)=\overline{\varphi_m(-\theta)}$ more precisely means that $\varphi_m(\theta)=e^{\mathrm{i}\omega}\overline{\varphi_m(-\theta)}$ for some $\omega\in[-\pi,\pi]$.*

Since $\|\varphi_m(\theta)\|^2=1$, we have that $\varphi_m\in L_H^2([-\pi,\pi])$. Hence, we can expand it in a Fourier series in the sense explained in the previous section:

$$\varphi_m=\sum_{\ell\in\mathbb{Z}}\phi_{m\ell}e_\ell, \quad \text{where} \quad \phi_{m\ell}=\frac{1}{2\pi}\int_{-\pi}^{\pi}\varphi_m(s)e^{-\mathrm{i}\ell s}ds.$$

The coefficients $\phi_{m\ell}$ thus defined give rise to the definition of the dynamic FPCs as in (7).

Remark 2. *Since $\varphi_m(\theta)$ is Hermitian, it immediately follows that $\phi_{m\ell}=\overline{\phi_{m\ell}}$, implying that the dynamic FPCs are real if the process (X_t) is real.*

A.3. Functional filters

Computation of dynamic FPCs requires applying time-invariant *functional filters* to the process (X_t). Let $\mathbf{\Psi} = (\Psi_k \colon k \in \mathbb{Z})$ be a sequence of linear operators, each mapping between separable Hilbert spaces H and H'. Further, let B be the backshift or lag operator, given by $B^k X_t := X_{t-k}$, $k \in \mathbb{Z}$. Then the functional filter $\Psi(B) := \sum_{k\in\mathbb{Z}} \Psi_k B^k$, when applied to the sequence (X_t), produces an output series (Y_t) in H' via

$$Y_t = \Psi(B)X_t = \sum_{k\in\mathbb{Z}} \Psi_k(X_{t-k}). \tag{23}$$

We call $\mathbf{\Psi}$ the filter coefficients, and, in the style of scalar or vector time series, we call the mapping $\Psi_\theta : [-\pi, \pi] \to \mathcal{L}(H, H')$ with

$$\Psi_\theta = \Psi(e^{-\mathrm{i}\theta}) = \sum_{k\in\mathbb{Z}} \Psi_k e^{-\mathrm{i}k\theta}$$

the *frequency response function* of the filter $\Psi(B)$.

Proposition 9. *Let $\Psi(B)$ be a functional filter with coefficients satisfying $\sum_{k\in\mathbb{Z}} \|\Psi_k\|_{\mathcal{L}} < \infty$, and let (Y_t) be given as in* (23)*. Then,*

(a) if $(X_t) \in L^2_H$, the series (23) *converges in $L^2_{H'}$;*

(b) the sequence (Y_t) is stationary with autocovariance operator

$$C_h^Y = \sum_{k\in\mathbb{Z}} \sum_{\ell\in\mathbb{Z}} \Psi_k C^X_{\ell-k+h} \Psi^*_\ell.$$

Proof. For (a), we need to show that $S_{n,t} = \sum_{k=-n}^{n} \Psi_k(X_{t-k})$ is a Cauchy sequence. If $m < n$, we get, by application of the Cauchy-Schwarz inequality,

$$\begin{aligned} E\|S_{n,t} - S_{m,t}\|^2 &= \sum_{m<|k|\le n} \sum_{m<|\ell|\le n} E\langle \Psi_k(X_{t-k}), \Psi_\ell(X_{t-\ell})\rangle \\ &\le \sum_{m<|k|\le n} \sum_{m<|\ell|\le n} \|\Psi_k\|_{\mathcal{L}} \|\Psi_\ell\|_{\mathcal{L}} E\|X_{t-k}\|\|X_{t-\ell}\|, \end{aligned}$$

and thus

$$E\|S_{n,t} - S_{m,t}\|^2 \le \Bigg(\sum_{|k|>m} \|\Psi_k\|_{\mathcal{L}}\Bigg)^2 E\|X_0\|^2,$$

which goes to zero as $(m, n) \to \infty$.

To establish (b), first remark that, for two sequences (Z_n) and (Z'_n) in L^2_H with $E\|Z_n - Z\|^2 \to 0$ and $E\|Z'_n - Z'\|^2 \to 0$, we have that

$$\begin{aligned} \|EZ_n \otimes Z'_n - EZ \otimes Z'\|_{\mathcal{S}} &\le E\|Z_n \otimes Z'_n - Z \otimes Z'\|_{\mathcal{S}} \\ &= E\|(Z_n - Z) \otimes Z'_n - Z \otimes (Z' - Z'_n)\|_{\mathcal{S}} \\ &\le E\|Z_n - Z\|^2 E\|Z'_n\|^2 + E\|Z\|^2 E\|Z' - Z'_n\|^2, \end{aligned}$$

and hence $\|EZ_n \otimes Z_n' - EZ \otimes Z'\|_{\mathcal{S}} \to 0$. Observing moreover that $E\Psi(Z) \otimes \Upsilon(Z') = \Psi C_{ZZ'} \Upsilon^*$, the result follows from the fact that

$$C_h^Y = \lim_{n\to\infty} ES_{n,t} \otimes S_{n,t-h} = \lim_{n\to\infty} \sum_{|k|\le n} \sum_{|\ell|\le n} E\Psi_k(X_{t-k}) \otimes \Psi_\ell(X_{t-\ell-h}).$$

□

Proposition 10. *Let $(X_t) \in L_H^2$ and $(X_t') \in L_H^2$ be two costationary processes with $\sum_{h\in\mathbb{Z}} \|C_h^{XX'}\|_{\mathcal{S}} < \infty$ and define their cospectrum as $\mathcal{F}_\theta^{XX'} := (2\pi)^{-1} \sum_{h\in\mathbb{Z}} C_h^{XX'} e^{-\mathrm{i}h\theta}$. Set $Y_t := \Psi(B)X_t$ and $Y_t' := \Upsilon(B)X_t'$, where $\Psi(B)$ and $\Upsilon(B)$ are functional filters with coefficients satisfying the summability conditions $\sum_{k\in\mathbb{Z}} \|\Psi_k\|_{\mathcal{S}} < \infty$ and $\sum_{k\in\mathbb{Z}} \|\Upsilon_k\|_{\mathcal{S}} < \infty$, respectively. Then (Y_t) and (Y_t') are again costationary and $\sum_{h\in\mathbb{Z}} \|C_h^{YY'}\|_{\mathcal{S}} < \infty$. Furthermore, $\mathcal{F}_\theta^{YY'} = \Psi_\theta \mathcal{F}_\theta^{XX'} \Upsilon_\theta^*$, where $\Upsilon_\theta^* := \sum_{k\in\mathbb{Z}} \Upsilon_k^* e^{\mathrm{i}k\theta}$.*

Proof. It is easy to see that (Y_t) and (Y_t') are costationary. Similar as in Proposition 9, we infer that

$$\sum_{h\in\mathbb{Z}} \|C_h^{YY'}\|_{\mathcal{S}} \le \sum_{h\in\mathbb{Z}} \sum_{k\in\mathbb{Z}} \sum_{\ell\in\mathbb{Z}} \|\Psi_k\|_{\mathcal{S}} \|\Upsilon_\ell^*\|_{\mathcal{S}} \|C_{h+\ell-k}^{XX'}\|_{\mathcal{S}}.$$

Then, using $\|\Upsilon^*\|_{\mathcal{S}} = \|\Upsilon\|_{\mathcal{S}}$ and summing first over h yields

$$\sum_{h\in\mathbb{Z}} \|C_h^{YY'}\|_{\mathcal{S}} \le \Big(\sum_{k\in\mathbb{Z}} \|\Psi_k\|_{\mathcal{S}}\Big)\Big(\sum_{k\in\mathbb{Z}} \|\Upsilon_k\|_{\mathcal{S}}\Big) \sum_{h\in\mathbb{Z}} \|C_h^{XX'}\|_{\mathcal{S}} < \infty.$$

Hence the operator $\mathcal{F}_\theta^{YY'}$ is well defined, and we have

$$\begin{aligned}
\mathcal{F}_\theta^{YY'} &= \frac{1}{2\pi} \sum_{h\in\mathbb{Z}} C_h^{YY'} e^{-\mathrm{i}h\theta} = \frac{1}{2\pi} \sum_{h\in\mathbb{Z}} \left(\sum_\ell \sum_k \Psi_k C_{\ell-k+h}^{XX'} \Upsilon_\ell^* \right) e^{-\mathrm{i}h\theta} \\
&= \frac{1}{2\pi} \sum_{h\in\mathbb{Z}} \left(\sum_\ell \sum_k \Psi_k C_{\ell-k+h}^{XX'} e^{-\mathrm{i}(\ell-k+h)\theta} \Upsilon_\ell^* e^{\mathrm{i}(\ell-k)\theta} \right) \\
&= \frac{1}{2\pi} \sum_\ell \sum_k \Psi_k \left(\sum_{h\in\mathbb{Z}} C_{\ell-k+h}^{XX'} e^{-\mathrm{i}(\ell-k+h)\theta} \right) \Upsilon_\ell^* e^{\mathrm{i}(\ell-k)\theta},
\end{aligned}$$

as was to be shown. □

Corollary 1. *Let (Ψ_k) be a functional filter such that $\sum_{k\in\mathbb{Z}} \|\Psi_k\|_{\mathcal{S}} < \infty$ and let (Y_t) be given as in (23). Assume that $(X_t) \in L_H^2$ is stationary with $\sum_{h\in\mathbb{Z}} \|C_h^X\|_{\mathcal{S}} < \infty$. Then $\sum_{h\in\mathbb{Z}} \|C_h^Y\|_{\mathcal{S}} < \infty$ and $\mathcal{F}_\theta^Y = \Psi_\theta \mathcal{F}_\theta^X \Psi_\theta^*$.*

A.4. Proofs for Section 2

We start by observing that Proposition 1 follows directly from Corollary 1. Part (a) of Proposition 2 also has been established in the previous Section (see Remark 2). Part (b) is immediate, and thus we can proceed to the proof of Proposition 3.

Proof of Proposition 3. To prove Part (a) one can proceed along the lines of the proof of Proposition 9, making use of (4) and

$$\sum_{k\in\mathbb{Z}} \|\phi_{mk}\|^2 = \frac{1}{2\pi}\int_{-\pi}^{\pi} \|\varphi_m(\theta)\|^2 d\theta = 1.$$

Then one shows that the partial sums $\sum_{k=-n}^{n}\langle X_{t-k}, \phi_{mk}\rangle$ form a Cauchy sequence in L^2 norm, by using

$$|E\langle X_{t-k}, \phi_{mk}\rangle\langle X_{t-\ell}, \phi_{m\ell}\rangle| = |\langle C_{\ell-k}(\phi_{m\ell}), \phi_{mk}\rangle| \le \|C_{\ell-k}\|_{\mathcal{L}}\|\phi_{m\ell}\|\|\phi_{mk}\|.$$

Parts (b) and (c) are immediate consequences of Proposition 1 (and Corollary 1 for the general setup). □

Proof of Propositions 4 and 5. Assume we have filter coefficients $\mathbf{\Psi} = (\Psi_k \colon k \in \mathbb{Z})$ and $\mathbf{\Upsilon} = (\Upsilon_k \colon k \in \mathbb{Z})$ where $\Psi_k : H \to \mathbb{C}^p$ and $\Upsilon_k : \mathbb{C}^p \to H$. If (X_t) and (Y_t) are H-valued and $\mathbb{C}^p$-valued processes, respectively, then there exist elements ψ_{mk} and υ_{mk} in H, such that

$$\Psi(B)(X_t) = \sum_{k\in\mathbb{Z}} \left(\langle X_{t-k}, \psi_{1k}\rangle, \dots, \langle X_{t-k}, \psi_{pk}\rangle\right)'$$

and

$$\Upsilon(B)(Y_t) = \sum_{\ell\in\mathbb{Z}}\sum_{m=1}^{p} Y_{t+\ell,m}\upsilon_{m\ell}.$$

Hence, the p-dimensional reconstruction of X_t in Proposition 5 is of the form

$$\sum_{m=1}^{p} \tilde{X}_{mt} = \Upsilon(B)[\Psi(B)X_t] =: \Upsilon\Psi(B)X_t.$$

Letting $\psi_m(\theta) = \sum_{k\in\mathbb{Z}} \psi_{mk}e^{\mathrm{i}k\theta}$ and $\upsilon_m(\theta) = \sum_{\ell\in\mathbb{Z}} \upsilon_{m\ell}e^{\mathrm{i}\ell\theta}$, we obtain, for $x \in H$ and $y = (y_1, \dots, y_m)' \in \mathbb{C}^p$, that the frequency response functions Ψ_θ and Υ_θ satisfy

$$\Psi_\theta(x) = \sum_{k\in\mathbb{Z}} \left(\langle x, \psi_{1k}\rangle, \dots, \langle x, \psi_{pk}\rangle\right)' e^{-\mathrm{i}k\theta} = \left(\langle x, \psi_1(\theta)\rangle, \dots, \langle x, \psi_p(\theta)\rangle\right)'$$

and

$$\Upsilon_\theta(y) = \sum_{\ell\in\mathbb{Z}}\sum_{m=1}^{p} y_m\upsilon_{m\ell}e^{-\mathrm{i}\ell\theta} = \sum_{m=1}^{p} y_m\upsilon_m(-\theta).$$

Consequently,

$$\Upsilon_\theta\Psi_\theta = \sum_{m=1}^{p} \upsilon_m(-\theta) \otimes \psi_m(\theta). \tag{24}$$

Now, using Proposition 10, it can be readily verified that, for $Z_t := X_t - \Upsilon\Psi(B)X_t$, we obtain the spectral density operator

$$\begin{aligned}\mathcal{F}_\theta^Z &= \mathcal{F}_\theta^X + \Upsilon_\theta\Psi_\theta\mathcal{F}_\theta^X\Psi_\theta^*\Upsilon_\theta^* - \Upsilon_\theta\Psi_\theta\mathcal{F}_\theta^X - \mathcal{F}_\theta^X\Psi_\theta^*\Upsilon_\theta^* \qquad (25)\\ &= \left(\tilde{\mathcal{F}}_\theta^X - \Upsilon_\theta\Psi_\theta\tilde{\mathcal{F}}_\theta^X\right)\left(\tilde{\mathcal{F}}_\theta^X - \tilde{\mathcal{F}}_\theta^X\Psi_\theta^*\Upsilon_\theta^*\right),\end{aligned}$$

where $\tilde{\mathcal{F}}_\theta^X$ is such that $\tilde{\mathcal{F}}_\theta^X \tilde{\mathcal{F}}_\theta^X = \mathcal{F}_\theta^X$. Also, from Proposition 10 it follows that the autocovariances of (Z_t) are summable in Hilbert-Schmidt norm, that is, $\sum_{h\in\mathbb{Z}} \|C_h^Z\|_\mathcal{S} < \infty$. We can conclude that the integral $\int_{-\pi}^{\pi} \mathcal{F}_\theta^Z d\theta$ exists and is equal to $EZ_t \otimes Z_t$. Therefore,

$$\begin{aligned} E\|X_t - \Upsilon\Psi(B)X_t\|^2 &= \operatorname{tr}\left(E[Z_t \otimes Z_t]\right) \\ &= \operatorname{tr}\left(\int_{-\pi}^{\pi} \mathcal{F}_\theta^Z d\theta\right) = \int_{-\pi}^{\pi} \operatorname{tr}(\mathcal{F}_\theta^Z)\, d\theta \\ &= \int_{-\pi}^{\pi} \left\|\tilde{\mathcal{F}}_\theta^X - \Upsilon_\theta \Psi_\theta \tilde{\mathcal{F}}_\theta^X\right\|_\mathcal{S}^2 d\theta. \end{aligned} \tag{26}$$

For the sake of rigor, let us justify that we can interchange above the trace and the integral. To this end, note that $\int_{-\pi}^{\pi} \mathcal{F}_\theta^Z d\theta = \mathrm{I}\mathcal{F}^Z$ if and only if

$$\langle \mathrm{I}\mathcal{F}^Z, V\rangle_\mathcal{S} = \int_{-\pi}^{\pi} \langle \mathcal{F}_\theta^Z, V\rangle_\mathcal{S}\, d\theta, \tag{27}$$

for all V in the space of Hilbert-Schmidt operators on H. From some ONB (v_k) define $V_N = \sum_{k=1}^N v_k \otimes v_k$. Then (27) implies that

$$\begin{aligned} \operatorname{tr}(\mathrm{I}\mathcal{F}^Z) &= \lim_{N\to\infty} \sum_{k=1}^N \langle \mathrm{I}\mathcal{F}^Z(v_k), v_k\rangle \\ &= \lim_{N\to\infty} \langle \mathrm{I}\mathcal{F}^Z, V_n\rangle_\mathcal{S} = \lim_{N\to\infty} \int_{-\pi}^{\pi} \langle \mathcal{F}_\theta^Z, V_n\rangle_\mathcal{S} d\theta \\ &= \lim_{N\to\infty} \int_{-\pi}^{\pi} \sum_{k=1}^N \langle \mathcal{F}_\theta^Z(v_k), v_k\rangle\, d\theta. \end{aligned}$$

Since $\mathcal{F}_\theta^Z$ is non-negative definite for any θ, the monotone convergence theorem allows to interchange the limit with the integral.

Now, (26) is minimized if we minimize the integrand for every fixed θ under the constraint that $\Upsilon_\theta \Psi_\theta$ is of the form (24). Employing the eigendecomposition

$$\mathcal{F}_\theta^X = \sum_{m\geq 1} \lambda_m(\theta) \varphi_m(\theta) \otimes \varphi_m(\theta),$$

we infer that

$$\tilde{\mathcal{F}}_\theta^X = \sum_{m\geq 1} \sqrt{\lambda_m(\theta)} \varphi_m(\theta) \otimes \varphi_m(\theta).$$

The best approximating operator of rank p to $\tilde{\mathcal{F}}_\theta^X$ is the operator

$$\tilde{\mathcal{F}}_\theta^X(p) = \sum_{m=1}^p \sqrt{\lambda_m(\theta)} \varphi_m(\theta) \otimes \varphi_m(\theta).$$

It is obtained if we choose $\Upsilon_\theta \Psi_\theta = \sum_{m=1}^p \varphi_m(\theta) \otimes \varphi_m(\theta)$ and hence

$$\psi_m(\theta) = \varphi_m(\theta) \quad \text{and} \quad \upsilon_m(\theta) = \varphi_m(-\theta).$$

Consequently, by Proposition 6 we get

$$\psi_{mk} = \frac{1}{2\pi}\int_{-\pi}^{\pi} \varphi_m(s)e^{-\mathrm{i}ks}ds \quad \text{and} \quad \upsilon_{mk} = \frac{1}{2\pi}\int_{-\pi}^{\pi} \varphi_m(-s)e^{-\mathrm{i}ks}ds = \psi_{m,-k}.$$

With this choice, it is clear that $\Upsilon\Psi(B)X_t = \sum_{m=1}^{p} X_{mt}$. Condition (8) assures that the involved series are mean square convergent. Hence

$$E\|X_t - \sum_{m=1}^{p} X_{tm}\|^2 = \int_{-\pi}^{\pi} \left\|\tilde{\mathcal{F}}_\theta^X - \tilde{\mathcal{F}}_\theta^X(p)\right\|_{\mathcal{S}}^2 d\theta = \int_{-\pi}^{\pi} \sum_{m>p} \lambda_m(\theta)d\theta;$$

the proof of Proposition 5 follows.

Turning to Proposition 4, observe that by the monotone convergence theorem, the last integral tends to zero if $p \to \infty$, which gives the proof of Proposition 4. □

B. Large Sample Properties

In this appendix, we study the consistency of the estimated dynamic FPC scores. For the sake of a neat and compact theory, we shall put aside the computational aspects treated in Section 3. More precisely, we assume that we have *fully observed functional data* and that all complicated computation (like integration, eigendecomposition, etc.) can be performed with arbitrary precision. As we already did throughout this article, we suppose that $(X_t : t \in \mathbb{Z})$ is a weakly stationary zero mean time series such that (4) holds. Then, the natural estimator for Y_{mt} is

$$\hat{Y}_{mt} := \sum_{\ell=-L}^{L} \langle X_{t-\ell}, \hat{\phi}_{m\ell}\rangle, \quad m = 1, \dots, p \quad \text{and} \quad t = L+1, \dots n-L, \tag{28}$$

where L is some integer and $\hat{\phi}_{m\ell}$ are retrieved from the estimated spectral density operator $\hat{\mathcal{F}}_\theta^X$. Note that, by a slight abuse of notation, $\hat{\mathcal{F}}_\theta^X$ and $\hat{\phi}_{m\ell}$ are based on fully observed and not on approximated data, as introduced in Section 3. We impose the following assumption.

Assumption B.1 The estimator $\hat{\mathcal{F}}_\theta^X$ is consistent in integrated mean square, i.e. we have

$$\int_{-\pi}^{\pi} E\|\mathcal{F}_\theta^X - \hat{\mathcal{F}}_\theta^X\|_{\mathcal{S}}^2 \, d\theta \to 0 \quad (n \to \infty). \tag{29}$$

Panaretos and Tavakoli [26] present an estimator which satisfies (29) under some functional cumulant conditions. Below we will establish an alternative sufficient condition, involving very mild technical conditions, such that Assumption B.1. holds. By stating (29) as an assumption, we intend to keep the theory more widely applicable.

Since our method requires the estimation of eigenvectors of the spectral density operator, we need to introduce the following identifiability assumption.

Assumption B.2 Define $\alpha_1(\theta) := \lambda_1(\theta) - \lambda_2(\theta)$ and $\alpha_m(\theta) := \min\{\lambda_{m-1}(\theta) - \lambda_m(\theta), \lambda_m(\theta) - \lambda_{m+1}(\theta)\}$ for $m > 1$, where $\lambda_i(\theta)$ is the i-th largest eigenvalue of the spectral density operator evaluated in θ. Then $\alpha_m(\theta)$ has at most finitely many zeros.

Theorem 1. *Let $\hat{Y}_{mt}$ be the random variable defined by (28). If $L = L(n) \to \infty$ sufficiently slowly, then, under Assumptions B.1 and B.2, we have $Y_{mt} \xrightarrow{\mathcal{P}} \hat{Y}_{mt}$ as $n \to \infty$.*

Remark 3. *This result does not provide any guidelines how to choose the truncation level L. This is a common problem for infinite dimensional data and usually can only be overcome by imposing a number of additional technical assumptions, which cannot be verified in practice. We refer to Hörmann and Kidziński [16] for a similar problem in the context of functional regression. A solution would be to follow their approach, and develop a data-driven algorithm for the choice of L. This, however, goes far beyond the scope of this article.*

For the proof of Theorem 1, we show that $E|Y_{mt} - \hat{Y}_{mt}| \to 0$. Since

$$E|Y_{mt} - \hat{Y}_{mt}| \le E\bigg|\sum_{j\in\mathbb{Z}} \langle X_{t-j}, \phi_{mj}\rangle - \sum_{j=-L}^{L} \langle X_{t-j}, \hat{\phi}_{mj}\rangle\bigg|$$

$$\le E\bigg|\sum_{j=-L}^{L} \langle X_{t-j}, \phi_{mj} - \hat{\phi}_{mj}\rangle\bigg| + E\bigg|\sum_{|j|>L} \langle X_{t-j}, \phi_{mj}\rangle\bigg|, \tag{30}$$

the result follows if each summand in (30) converges to zero. This will be proven in the two subsequent lemmas.

Lemma 2. *If $L = L(n) \to \infty$ sufficiently slowly, then, under Assumptions B.1 and B.2, we have that*

$$\bigg|\sum_{|j|\le L} \langle X_{k-j}, \phi_{mj} - \hat{\phi}_{mj}\rangle\bigg| = o_P(1) \quad (n \to \infty).$$

Proof. The triangle inequality and the Cauchy-Schwarz inequality yield

$$\bigg|\sum_{|j|\le L} \langle X_{k-j}, \phi_{mj} - \hat{\phi}_{mj}\rangle\bigg| \le \sum_{j=-L}^{L} \|X_{k-j}\| \|\phi_{mj} - \hat{\phi}_{mj}\|$$

$$\le \max_{j\in\mathbb{Z}} \|\phi_{mj} - \hat{\phi}_{mj}\| \sum_{j=-L}^{L} \|X_{k-j}\|.$$

Furthermore, Jensen's inequality and Lemma 3.2 in [17] imply that, for any $j \in \mathbb{Z}$,

$$\begin{aligned} 2\pi\|\phi_{mj} - \hat{\phi}_{mj}\| &= \left\| \int_{-\pi}^{\pi} (\varphi_m(\theta) - \hat{\varphi}_m(\theta)) e^{\mathrm{i}j\theta} d\theta \right\| \\ &\le \int_{-\pi}^{\pi} \|\varphi_m(\theta) - \hat{\varphi}_m(\theta)\| d\theta \\ &\le \int_{-\pi}^{\pi} \frac{8}{|\alpha_m(\theta)|^2} \|\Gamma_\theta^X - \hat{\Gamma}_\theta^X\|_{\mathcal{S}} \wedge 2 \; d\theta. \end{aligned}$$

By Assumption B.2, $\alpha_m(\theta)$ has only finitely many zeros, $\theta_1, \ldots, \theta_K$, say. Let now $\delta_\varepsilon(\theta) = [\theta - \varepsilon, \theta + \varepsilon]$ and $A(m, \varepsilon) = \bigcup_{i=1}^{K} \delta_\varepsilon(\theta_i)$. By definition, the length of this set is $|A(m, \varepsilon)| \le 2K\varepsilon$. Now define M_ε such that $M_\varepsilon^{-1} = \min\{\alpha_m(\theta) \mid \theta \in [-\pi, \pi] \backslash A(m, \varepsilon)\}$. By continuity of $\alpha_m(\theta)$ (see Proposition 8), we have $M_\varepsilon < \infty$, and thus

$$\int_{-\pi}^{\pi} \frac{8}{|\alpha_m(\theta)|^2} \|\Gamma_\theta^X - \hat{\Gamma}_\theta^X\| \wedge 2 d\theta \le 4K\varepsilon + 8M_\varepsilon^2 \int_{-\pi}^{\pi} \|\Gamma_\theta^X - \hat{\Gamma}_\theta^X\| d\theta =: B_{n,\varepsilon}.$$

By Assumption B.1, there exists a sequence $\varepsilon_n \to 0$ such that $B_{n,\varepsilon_n} \to 0$ in probability. Thus, if we choose $L(n)$ such that $L = L(n) \to \infty$ and $LB_{n,\varepsilon_n} = o_P(1)$, then

$$\left| \sum_{|j| \le L} \langle X_{k-j}, \phi_{mj} - \hat{\phi}_{mj} \rangle \right| \le LB_{n,\varepsilon_n} \left(L^{-1} \sum_{j=-L}^{L} \|X_{k-j}\| \right). \tag{31}$$

It remains to show that $L^{-1} \sum\limits_{j=-L}^{L} \|X_{k-j}\| = O_P(1)$. By the imposed weak stationarity we have $E\|X_k\|^2 = E\|X_1\|^2$, and hence for any $R > 0$

$$P\left(L^{-1} \sum_{j=-L}^{L} \|X_{k-j}\| > R \right) \le \frac{\sum_{k=-L}^{L} E\|X_k\|}{LR} \le \frac{3\sqrt{E\|X_1\|^2}}{R}.$$

□

Lemma 3. *Let $L = L(n) \to \infty$. Then, under condition* (4)*, we have*

$$\left| \sum_{|j| > L} \langle X_{k-j}, \phi_{mj} \rangle \right| = o_P(1) \quad (n \to \infty).$$

Proof. The triangle and the Cauchy-Schwarz inequalities, and elementary algebraic transformations give

$$\begin{aligned} E\left| \sum_{|j| > L} \langle X_{k-j}, \phi_{mj} \rangle \right|^2 &\le \sum_{|k| > L} \sum_{|l| > L} \|C_{k-l}\|_{\mathcal{L}} \|\phi_{km}\| \|\phi_{lm}\| \\ &= \sum_{k \in \mathbb{Z}} \sum_{l \in \mathbb{Z}} \|C_{k-l}\|_{\mathcal{L}} \|\phi_{km}\| \|\phi_{lm}\| I\{|k| > L\} I\{|l| > L\}. \end{aligned}$$

Now, setting $h = k - l$, we get

$$
\begin{aligned}
&E\left|\sum_{|j|>L}\langle X_{k-j}, \phi_{mj}\rangle\right|^2 \\
&\quad\leq \sum_{k\in\mathbb{Z}}\sum_{h\in\mathbb{Z}}\|C_h\|_{\mathcal{L}}\|\phi_{km}\|\|\phi_{(k-h)m}\|I\{|k|>L\}I\{|h-k|>L\} \\
&\quad\leq \sum_{h\in\mathbb{Z}}\|C_h\|_{\mathcal{L}}\sum_{k\in\mathbb{Z}}\|\phi_{km}\|\|\phi_{(k-h)m}\|I\{|k|>L\} \\
&\quad\leq \sum_{h\in\mathbb{Z}}\|C_h\|_{\mathcal{L}}\left(\sum_{k\in\mathbb{Z}}\|\phi_{km}\|^2 I\{|k|>L\}\right)^{1/2}\left(\sum_{k\in\mathbb{Z}}\|\phi_{km}\|^2\right)^{1/2}.
\end{aligned}
$$

The proof follows now from condition (4) and $\sum_{k\in\mathbb{Z}}\|\phi_{km}\|^2 = 1$. □

We have stated the consistency result under the assumption of weak stationarity and Assumptions B.1 and B.2. The following proposition shows that Assumption B.1 holds under L^4-m-approximability. We use the estimator

$$
\hat{\mathcal{F}}_\theta^X = \sum_{|h|\leq q}\left(1 - \frac{|h|}{q}\right)\hat{C}_h^X e^{-\mathrm{i}h\theta}, \quad 0 < q < n.
$$

Proposition 11. *Let $(X_t : t \in \mathbb{Z})$ be an L^4-m-approximable series and let $q = q(n) \to \infty$ such that $q^3 = o(n)$. Then Assumption B.1 holds.*

For the proof, we need the following lemma, which is extending a consistency result from [17] for the empirical covariance operator to lag h autocovariance operators. We define, for $|h| < n$,

$$
\hat{C}_h = \frac{1}{n}\sum_{k=1}^{n-h} X_{k+h}\otimes X_k, \quad h \geq 0, \quad \text{and} \quad \hat{C}_h = \hat{C}_{-h}, \quad h < 0.
$$

Lemma 4. *Assume that $(X_t : t \in \mathbb{Z})$ is an L^4-m-approximabable series. Then for all $|h| < n$ we have $E\|\hat{C}_h - C_h\|_{\mathcal{S}} \leq U\sqrt{\frac{|h|\vee 1}{n}}$, where the constant U does neither depend on n nor on h.*

Proof. Let us only consider the case $h \geq 0$. Define $X_n^{(r)}$ as the r-dependent approximation of (X_n) provided by Definition 3. We observe that

$$
nE\big\|\hat{C}_h - C_h\big\|_{\mathcal{S}}^2 = nE\left\|\frac{1}{n}\sum_{k=1}^{n-h} Z_k\right\|_{\mathcal{S}}^2,
$$

where $Z_k = X_{k+h}\otimes X_k - C_h$. Set $Z_k^{(r)} = X_{k+h}^{(r)}\otimes X_k^{(r)} - C_h$. Using the stationarity of the sequence (Z_k) we obtain

$$nE\left\|\frac{1}{n}\sum_{k=1}^{n-h} Z_k\right\|_{\mathcal{S}}^2 = \sum_{|r|<n-h}\left(1-\frac{|r|}{n}\right)E\langle Z_0, Z_r\rangle_{\mathcal{S}}$$
$$\leq \sum_{r=-h}^{h}|E\langle Z_0, Z_r\rangle_{\mathcal{S}}| + 2\sum_{r=h+1}^{\infty}|E\langle Z_0, Z_r\rangle_{\mathcal{S}}|, \tag{32}$$

and using the Cauchy-Schwarz inequality gives

$$|E\langle Z_0, Z_r\rangle_{\mathcal{S}}| \leq E|\langle Z_0, Z_r\rangle_{\mathcal{S}}| \leq \sqrt{E\|Z_0\|_{\mathcal{S}}^2 E\|Z_r\|_{\mathcal{S}}^2} = E\|Z_0\|_{\mathcal{S}}^2.$$

Furthermore, from $\|X_h \otimes X_0\| = \|X_h\|\|X_0\|$, we deduce

$$E\|Z_0\|_{\mathcal{S}}^2 = E\|X_0\|^2\|X_h\|^2 \leq \left(E\|X_0\|^4\right)^{1/2} < \infty.$$

Consequently, we can bound the first sum in (32) by $(2h+1)\left(E\|X_0\|^4\right)^{1/2}$. For the summands of the second term in (32) we obtain by independence of $Z_r^{(r-h)}$ and Z_0 that

$$|E\langle Z_0, Z_r\rangle_{\mathcal{S}}| = |E\langle Z_0, Z_r - Z_r^{(r-h)}\rangle_{\mathcal{S}}| \leq (E\|Z_0\|_{\mathcal{S}}^2)^{1/2}(E\|Z_r - Z_r^{(r-h)}\|_{\mathcal{S}}^2)^{1/2}.$$

To conclude, it suffices to show that $\sum_{r=1}^{\infty}(E\|Z_r - Z_r^{(r-h)}\|_{\mathcal{S}}^2)^{1/2} \leq M < \infty$, where the bound M is independent of h. Using an inequality of the type $|ab-cd|^2 \leq 2|a|^2|b-d|^2 + 2|d|^2|a-c|^2$, we obtain

$$\begin{aligned}
E\|Z_r - Z_r^{(r-h)}\|_{\mathcal{S}}^2 &= E\|X_r \otimes X_{r+h} - X_r^{(r-h)} \otimes X_{r+h}^{(r-h)}\|_{\mathcal{S}}^2 \\
&\leq 2E\|X_r\|^2\|X_{r+h} - X_{r+h}^{(r-h)}\|^2 + 2E\|X_{r+h}^{(r-h)}\|^2\|X_r - X_r^{(r-h)}\|^2 \\
&\leq 2(E\|X_r\|^4)^{1/2}(E\|X_{r+h} - X_{r+h}^{(r-h)}\|^4)^{1/2} \\
&\quad + 2(E\|X_{r+h}^{(r-h)}\|^4)^{1/2}(E\|X_r - X_r^{(r-h)}\|^4)^{1/2}.
\end{aligned}$$

Note that $E\|X_r\|^4 = E\|X_{r+h}^{(r-h)}\|^4 = E\|X_0\|^4$ and that $E\|X_{r+h} - X_{r+h}^{(r-h)}\|^4 = E\|X_r - X_r^{(r-h)}\|^4 = E\|X_0 - X_0^{(r-h)}\|^4$. Altogether we get

$$E\|Z_r - Z_r^{(r-h)}\|_{\mathcal{S}}^2 \leq 4(E\|X_0\|^4)^{1/2}(E\|X_0 - X_0^{(r-h)}\|^4)^{1/2}.$$

Hence, L^4-m-approximability implies that $\sum_{r=h+1}^{\infty}|E\langle Z_0, Z_r\rangle_{\mathcal{S}}|$ converges and is uniformly bounded over $0 \leq h < n$. □

Proof of Proposition 11. By the triangle inequality,

$$\begin{aligned}
2\pi\|\mathcal{F}_\theta^X - \hat{\mathcal{F}}_\theta^X\|_\mathcal{S} &= \left\| \sum_{k\in\mathbb{Z}} C_h e^{-\mathrm{i}h\theta} - \sum_{h=-q}^{q} \left(1 - \frac{|h|}{q}\right) \hat{C}_h e^{-\mathrm{i}h\theta} \right\|_\mathcal{S} \\
&\leq \left\| \sum_{h=-q}^{q} \left(1 - \frac{|h|}{q}\right) (C_h - \hat{C}_h) e^{-\mathrm{i}h\theta} \right\|_\mathcal{S} \\
&\quad + \left\| \frac{1}{q} \sum_{h=-q}^{q} |h| C_h e^{-\mathrm{i}h\theta} \right\|_\mathcal{S} + \left\| \sum_{|h|>q} C_h e^{-\mathrm{i}h\theta} \right\|_\mathcal{S} \\
&\leq \sum_{h=-q}^{q} \left(1 - \frac{|h|}{q}\right) \|C_h - \hat{C}_h\|_\mathcal{S} + \frac{1}{q} \sum_{h=-q}^{q} |h| \|C_h\|_\mathcal{S} + \sum_{|h|>q} \|C_h\|_\mathcal{S}.
\end{aligned}$$

The last two terms tend to 0 by condition (4) and Kronecker's lemma. For the first term we may use Lemma 4. By taking the expectation, we obtain for some U_1 that

$$\sum_{h=-q}^{q} \left(1 - \frac{|h|}{q}\right) E\|C_h - \hat{C}_h\|_\mathcal{S} \leq U_1 \frac{q^{3/2}}{\sqrt{n}}.$$

Note that the bound does not depend on θ, hence $q^3 = o(n)$ and the condition (4) imply that $\sup_{\theta\in[-\pi,\pi]} E\|\mathcal{F}_\theta^X - \hat{\mathcal{F}}_\theta^X\|_\mathcal{S} \to 0$ as $n \to \infty$. □

Acknowledgement

The research of Siegfried Hörmann and Łukasz Kidziński was supported by the Communauté française de Belgique – Actions de Recherche Concertées (2010–2015) and the Belgian Science Policy Office – Interuniversity attraction poles (2012–2017). The research of Marc Hallin was supported by the Sonderforschungsbereich "Statistical modeling of nonlinear dynamic processes" (SFB823) of the Deutsche Forschungsgemeinschaft and the Belgian Science Policy Office – Interuniversity attraction poles (2012–2017).

References

[1] Aue, A., Dubart Norinho, D. and Hörmann, S. (2012+), *On the prediction of functional time series,* Submitted for publication.

[2] Aston, JAD and Kirch, C. (2011), *Estimation of the distribution of changepoints with appliction to fMRI data,* Technical Report, University of Warwick, Centre for Research in Statistical Methodology, 2011.

[3] Benko, M., Härdle, W. and Kneip, A. (2009), *Common functional principal components,* The Annals of Statistics **37**, 1–34.

[4] Besse, P. and Ramsay, J. O. (1986), *Principal components analysis of sampled functions,* Psychometrika **51**, 285–311.

[5] Brillinger, D. R. (1981), *Time Series: Data Analysis and Theory,* Holden Day, San Francisco.

[6] Brockwell, P. J. and Davis, R. A. (1981), *Time Series: Theory and Methods,* Springer, New York.

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

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

[9] 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**, 136–154.

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

[11] Gervini, D. (2007), *Robust functional estimation using the median and spherical principal components,* Biometrika **95**, 587–600.

[12] Gohberg, I., Goldberg, S. and Kaashoek, M. A. (2003), *Basic Classes of Linear Operators,* Birkhäuser.

[13] Gokulakrishnan, P., Lawrence, P. D., McLellan, P. J. and Grandmaison, E. W. (2006), *A functional-PCA approach for analyzing and reducing complex chemical mechanisms,* Computers and Chemical Engineering **30**, 1093–1101.

[14] Hall, P. and Hosseini-Nasab, M. (2006), *On properties of functional principal components analysis,* J. Roy. Statist. Soc. Ser. B **68**, 109–126.

[15] Hall, P., Müller, H.-G. and Wang, J.-L. (2006), *Properties of principal component methods for functional and longitudinal data analysis,* The Annals of Statistics **34**, 1493–1517.

[16] Hörmann, S. and Kidziński, Ł. (2012), *A note on estimation in Hilbertian linear models,* Technical Report, Université libre de Bruxelles, Department of Mathematics, May 2012.

[17] Hörmann, S. and Kokoszka, P. (2010), *Weakly dependent functional data,* The Annals of Statistics **38**, 1845–1884.

[18] Hörmann, S. and Kokoszka, P. (2012), *Functional Time Series,* Handbook of Statistics: Time Series Analysis-Methods and Applications. Vol. 30, 157–186.

[19] Hyndman, R. J. and Ullah, M. S. (2007), *Robust forecasting of mortality and fertility rates: a functional data approach,* Computational Statistics & Data Analysis **51**, 4942–4956.

[20] James, G.M., Hastie T. J. and Sugar, C. A. (2000), *Principal Component Models for Sparse Functional Data,* Biometrika **87**, 587–602.

[21] Jolliffe, J. (2005), *Principal Component Analysis,* Wiley & Sons.

[22] Karhunen, K. (1947), *Über lineare Methoden in der Wahrscheinlichkeitsrechnung,* Ann. Acad. Sci. Fennicae. Ser. A. I. Math.-Phys. **37**, 79.

[23] Kneip, A. and Utikal, K. (2001), *Inference for density families using functional principal components analysis,* J. Amer. Statist. Assoc. **96**, 519–531.

[24] Locantore, N., Marron, J. S., Simpson, D. G., Tripoli, N., Zhang, J. T. and Cohen, K. L. (1999), *Robust principal component analysis for functional data*, Test **8**, 1–73.

[25] Loève, M. (1946), *Fonctions aleatoires de second ordre,* Revue Sci. **84**, 195–206.

[26] Panaretos, V. M. and Tavakoli, S. (2012), *Fourier analysis of stationary time series in function space,* Technical Report #01-12, Chair of Mathematical Statistics, EPFL, April 2012 (submitted to the Annals of Statistics).

[27] Politis, D. N. (2011), *Higher-order accurate, positive semi-definite estimation of large-sample covariance and spectral density matrices,* Econometric Theory **27**, 703–744.

[28] Ramsay, J. O. and Dalzell, C. J. (1991), *Some tools for functional data analysis (with discussion),* Journal of the Royal Statistical Society: Series B **53**, 539–572.

[29] Ramsay, J. and Silverman, B. (2002), *Applied Functional Data Analysis,* Springer, New York.

[30] Ramsay, J. and Silverman, B. (2005), *Functional Data Analysis (2nd ed.),* Springer, New York.

[31] Reiss, P. T. and Ogden, R. T. *Functional principal component regression and functional partial least squares,* J. Amer. Statist. Assoc. **102**, 984–996.

[32] Silverman, B. (1996), *Smoothed functional principal components analysis by choice of norm,* The Annals of Statistics **24**, 1–24.

[33] Stadlober, E., Hörmann, S., and Pfleiler, B. (2008), *Quality and performance of a PM10 daily forecasting model,* Atmospheric Environment **42**, 1098–1109.

[34] Viviani, R., Grön, G. and Spitzer, M. (2005), *Functional principal component analysis of fMRI data,* Human Brain Mapping **24**, 109–129.