

## Review Article

# Dependent Functional Data

**Piotr Kokoszka**

*Department of Statistics, Colorado State University, Fort Collins, CO 80523, USA*

Correspondence should be addressed to Piotr Kokoszka, [piotr.kokoszka@colostate.edu](mailto:piotr.kokoszka@colostate.edu)

Received 12 July 2012; Accepted 13 August 2012

Academic Editors: P. D'Urso and P. E. Jorgensen

Copyright © 2012 Piotr Kokoszka. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper reviews recent research on dependent functional data. After providing an introduction to functional data analysis, we focus on two types of dependent functional data structures: time series of curves and spatially distributed curves. We review statistical models, inferential methodology, and possible extensions. The paper is intended to provide a concise introduction to the subject with plentiful references.

## 1. Introduction

Functional data analysis (FDA) is a relatively new branch of statistics, going back to the early 1990s, but its mathematical foundations are rooted in much earlier developments in the theory of operators in a Hilbert space and the functional analysis. In the most basic setting, the sample consists of  $N$  curves  $X_1(t), X_2(t), \dots, X_N(t)$ ,  $t \in \mathcal{T}$ . The set  $\mathcal{T}$  is typically an interval of the line. In increasingly many applications, it is however a subset of the plane, or a sphere, or even a 3D region. In those cases, the data are surfaces over a region, or more general functions over some domain, hence the term functional data. This survey is concerned mostly with the analysis of curves, but some references to more general functions are given in Section 1.1.

Functional data are *high-dimensional data*, as, in a statistical model, each functions  $X_n$  consists of infinitely many values  $X_n(t)$ ,  $t \in \mathcal{T}$ . In traditional statistics, the data consist of a sample of scalars or vectors. For example, for each survey participant, we may record age, gender, income, and education level. The data point thus has dimension four; it is a vector with quantitative and categorical entries. High-dimensional data typically have dimension comparable to or larger than the sample size. As they are often analyzed using regression models in which the sample size is denoted by  $n$  and the number of explanatory variables by  $p$ , high-dimensional data often fall into the “large  $p$ , small  $n$ ” paradigm, but clearly they form a much broader class, with a great deal of work focusing on covariance matrices based

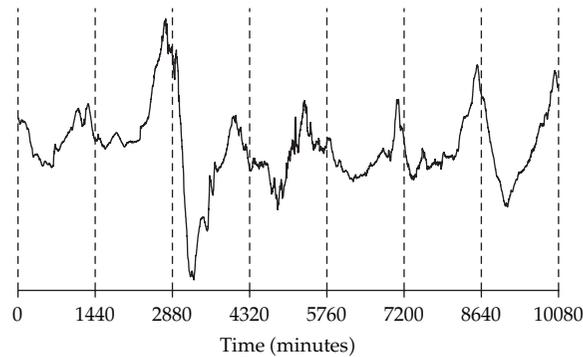
on a sample of  $np$ -dimensional vectors. A distinctive feature of functional data is that the curves or surfaces are assumed to be smooth in some sense; if  $t_1$  is close  $t_2$ , the values  $X_n(t_1)$  and  $X_n(t_2)$  should be similar. In the “large  $p$ , small  $n$ ” paradigm, there need not be any natural ordering of the covariates or any natural measure of distance between them. The analysis often focuses on the selection of a small number of relevant covariates (the variable selection problem). In the FDA, the analysis involves obtaining a smooth, low dimensional representation of each curve.

FDA views each curve in a sample as a separate statistical object. In this sense, FDA is part of the *object data analysis* in which data points are not scalars or vectors, but structures which are modeled by complex mathematical objects, for example, by graphs. Some references are given in Section 1.1.

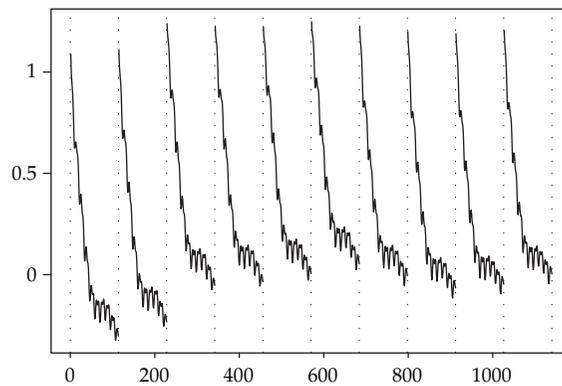
However, even curves are far more complicated structures than scalars or vectors. The curves are characterized not only by magnitude but also by shape. The shape of a random curve plays a role analogous to the dependence between the coordinates of a random vector. Human growth curves provide a well-known example. Suppose that there are  $N$  randomly selected subjects of the same gender. Let  $X_k(t_j)$  be the height of the  $k$ th subject measured at time  $t_j$  from birth. The points  $t_j$  are different for different subjects. Using methods of FDA, we can construct continuous and differentiable curves  $X_k(t)$ ,  $0 \leq t \leq T$ . The shapes and magnitudes of these curves give us an idea about the variability in the process of growth, rather just about the variability of the final height, which can be assessed using the scalars  $X_k(T)$ ,  $1 \leq k \leq N$ .

Some data can be very naturally viewed as curves. For example, if the height measurements are available at a fairly regular and sufficiently dense grid of times  $t_j$ , it is easy to visualize them as curves, even though it is not immediately obvious how to compute derivatives of such curves. In many situations, the points  $t_j$  are extremely dense. For example, physical instruments may return an observation every five seconds, so in a day, we will have 17,280 values  $t_j$ . A day is a natural time domain in many applications, and the problem is to replace the 17,280 values  $X_n(t_j)$  available in day  $n$  by a smaller more manageable set of numbers. This is generally possible due to the assumption of some smoothness. At the other extreme are sparse longitudinal data. Such data often arise in medical research. For example, a measurement can be made on a patient only several times during the course of treatment. Yet we know that the quantity that is measured exists at any time, so it is a curve that is observed only at a few sparse time points. References to the relevant functional methodology are given in Section 1.1.

Growth curves or sparse observations on a sample of patients can be viewed as independent curves drawn from a population of interest. A large body of research in FDA has been motivated by various problems arising in such a setting. At the same time, many functional data sets, most notably in physical and environmental sciences, arise from long records of observations. An example is presented in Figure 1 which shows seven consecutive functional observations (curves). These curves show a very rough periodic pattern, but modeling periodicity is difficult, as this pattern is, in fact, severely disturbed several times a month due to ionospheric storms. The 24h period must however enter into any statistical model as it is caused by the rotation of the Earth. It is thus natural in this context to treat the long continuous record as consisting of consecutive curves, each defined over a 24h time interval. Space physics researchers have long been associating the occur enhancements on a given day with physical phenomena in near Earth space. This gives additional support to treating these data as a time series of curves of evolving shape, which we will call a *functional time series*. Similar functional series arise, for example, in urban pollution monitoring studies.



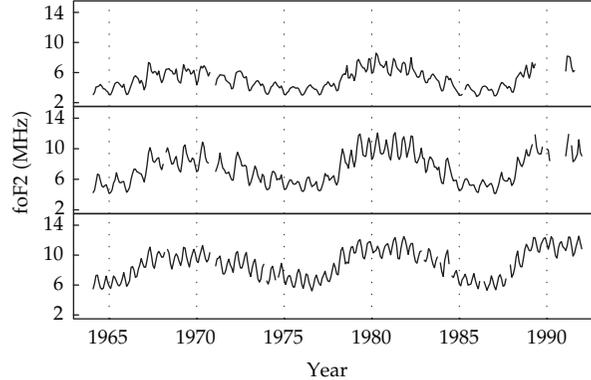
**Figure 1:** The horizontal component of the magnetic field measured in one minute resolution at Honolulu magnetic observatory from 1/1/2001 00:00 UT to 1/7/2001 24:00 UT.



**Figure 2:** Eurodollar futures curves on ten consecutive days.

A functional time series does not however need to arise from cutting a continuous time record into adjacent pieces of natural length. Figure 2 shows a functional time series of curves representing (centered) Eurodollar futures prices. For each day, the time is not time within that day, but a time to the expiration of a contract, more details are given in Chapter 14 of Horváth and Kokoszka [1]. The curves show how the prices of contracts with various expiration horizons evolve from day to day.

Functional time series focus on temporal or otherwise sequential dependence between curves. In many applications, the curves are available at points in space. Such spatially indexed functional data will be denoted  $X(\mathbf{s}_k, t)$ , where  $\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_N$  are locations in some region and  $X(\mathbf{s}_k, t)$  is the value of the function  $X(\mathbf{s}_k)$  at time  $t$ . To compare such data structures to functional time series, consider two ways of looking at pollution data. If we consider only one location and are interested in the day to day pattern of pollution, we may consider curves  $X_n(t)$ , where  $n$  is the day index and  $t$  is the time within the day. In many studies, we are however interested in long-term trends in pollution over a region, for example, a large city. In such a case, we may wish to smooth out the day to day variability on consider the curves  $X(\mathbf{s}_k, t)$ , where  $\mathbf{s}_k$  is the location of a measurement station within the city and  $t$  is time within a few decades. Such a data structure is difficult to show in one graph. The spatial location  $\mathbf{s}_k$  can be shown on a map. The curves  $X(\mathbf{s}_k, t)$  must be shown on a separate graph, as in Figure 3.



**Figure 3:** F2-layer critical frequency curves at three locations. Top to bottom (latitude in parentheses): Yakutsk (62.0), Yamagawa (31.2), and Manila (14.7). The functions exhibit a latitudinal trend in amplitude.

As for the functional time series, the main feature of such data is that there is dependence of the characteristics of the curves which depended on the distance between them.

Of course, one can consider a more complex structure in which one may study spatially indexed functional time series, with the data being  $X_n(\mathbf{s}_k, t)$ ; for a fixed  $\mathbf{s}_k$ , we have a functional time series, for a fixed  $n$ , we have a functional spatial random field. We briefly discuss such a setting in Section 5.

*The focus of this paper is on the recent developments in the research on dependent functional data.* Section 3 considers functional time series, and Section 4 spatially indexed functions. The required background is provided in Section 2. A summary, which includes the discussion of some problems of current interest, is given in Section 5.

### 1.1. Further Reading

In this section, we discuss several important topics mentioned above, mostly by directing the reader to appropriate references.

#### 1.1.1. Monographs on FDA

The two editions of the book of Ramsay and Silverman [2] have done a lot to introduce FDA to the statistics community and beyond. The monograph introduces most of the fundamental ideas of FDA including penalized smoothing, data registration, functional linear regression, functional principal components and functional analysis of derivatives. Ramsay and Silverman [3] elaborate on many data examples in Ramsay and Silverman [2] by providing detailed case studies, while Ramsay et al. [4] focus on the computational aspects of the analyses introduced in Ramsay and Silverman [2]. These books are concerned with iid samples of curves.

The book of Ferraty and Vieu [5] covers nonparametric methods with special emphasis on nonparametric prediction and classification. It has several chapters on dependent curves ( $\alpha$ -mixing). It studies both practical and abstract aspects of the problems. The collections Ferraty and Romain [6] and Ferraty [7] contain contributions from prominent researchers illustrating the increasingly many facets of FDA. Ferraty and Romain [6] has a smaller

number of detailed papers, while Ferraty [7] has a large number of short communications. Shi and Choi [8] consider Gaussian regression models.

The monograph of Bosq [9] is highly recommended to anyone seeking a fast and rigorous introduction to the theory of functional time series. It contains all practically necessary facts from the theory of probability in Hilbert and Banach spaces, and a great deal of relevant theory of linear functional time series. Bosq and Blanke [10] elaborate on some aspects of this theory in an abstract way.

To fully understand some theoretical aspects of FDA a good background in the theory of Hilbert spaces is needed. There are many introductory textbooks, my personal favorites are Riesz and Nagy [11], Akhiezer and Glazman [12], and Debnath and Mikusinski [13]. There are fewer books on probability in Hilbert and Banach spaces, I have benefited from the monographs of Araujo and Giné [14], Linde [15], and Vakhaniia et al. [16].

The monograph of Horváth and Kokoszka [1] elaborates on many topics mentioned in this paper, especially those related to the functional principal components.

### 1.1.2. Basis Expansions

In this paper, we focus on procedures for densely observed curves. For such data, functional objects needed to perform the calculations can be created using the R package `fda`. We will not address the details of calculations, but note that all necessary background is given in Ramsay et al. [4]. In the following, we often refer the choice of the a *basis* and the number of basis functions. We now briefly discuss this point.

Functions are observed at discrete sampling values  $t_j$ ,  $j = 1, \dots, J$ , which may or may not be equally spaced. We work with  $N$  functions. These data are converted to functional objects. In order to do this, we need to specify a basis. A basis is a system of basis functions, a linear combination of which defines the functional objects. The elements of a basis may or may not be orthogonal. We express a functional observation  $X_n$  as

$$X_n(t) \approx \sum_{k=1}^K c_{nk} \phi_k(t), \quad (1.1)$$

where the  $\phi_k$ ,  $k = 1, \dots, K$  are the basis functions. One of the advantages of this approach is that instead of storing all the data points, one stores the coefficients of the expansion, that is, the  $c_{nk}$ . This step thus involves an initial dimension reduction and some smoothing. All subsequent computations are performed on the matrices built from the coefficients  $c_{nk}$ . The number  $K$  of the basis functions impacts the performance of some procedures, but others are fairly insensitive to its choice. For the data studied in this paper, we generally choose  $K$  so that the plotted functional objects resemble original data with some smoothing that eliminates the most obvious noise. The two most commonly used basis systems are the *Fourier* and the *B-spline* bases. The Fourier basis is usually used for periodic, or nearly periodic functions with no strong local features and a roughly constant curvature. They are inappropriate for data with discontinuities in the function itself or in low order derivatives. The B-spline basis is typically used for nonperiodic locally smooth data. The B-spline basis functions are not orthogonal.

### 1.1.3. *Sparsely Observed Functions*

As noted above, in some applications, the data are available only at a few sparsely distributed points  $t_j$ , which may be different for different curves, and the data may exhibit nonnegligible measurement errors. Yao et al. [17] introduce methodology to deal with such data; smoothing with a basis expansion is at best inappropriate, and often not feasible. The essence of the alternative approach is explained in Müller [18]. While there are various elaborations, see for example, Hall et al. [19], the idea is that smoothing must be applied to all observations collectively, not to individual trajectories (which basically do not exist for sparse data). The mean function is first obtained by smoothing the mean of all trajectories. Then the functional principal components are calculated as the eigenfunctions of a covariance kernel obtained by surface smoothing. Extensions of this technique have been applied in many settings, see for example, Müller and Stadtmüller [20], Müller et al. [21], Yao and Müller [22], and Di and Cariniceanu [23].

### 1.1.4. *Software*

All procedures described in this paper can be implemented in readily available statistical software. Ramsay et al. [4] provide a solid introduction to computational issues for functional data, as well as numerous examples. Their book describes the R package `fda` and analogous Matlab code. Clarkson et al. [24] describe the implementation in S+. Many specialized R packages now exist. Techniques for sparse data are implemented in the Matlab package PACE developed at the University of California at Davis, available at <http://anson.ucdavis.edu/~mueller/data/software.html>, at the time of writing.

### 1.1.5. *Extensions beyond Curves*

Various brain scans can be viewed as functions over a spatial domain. Analysis of such data is discussed in several papers, see for example, Reiss and Ogden [25] and Zipunnikov et al. [26]. Guillas and Lai [27], consider functional regression models with surfaces as regressors, Morris et al. [28] propose a Bayesian framework for the analysis of images.

The study of the brain also provides motivation to study complex statistical objects. Aydin et al. [29] model blood vessels in the brain as trees. Kenobi et al. [30] and Crane and Patrangenu [31] provide references to data analysis on manifolds.

## 2. A Modeling Framework for Functional Data Analysis

This paper focuses on *inferential* procedures for functional data. To derive them, a statistical model for the data is needed. In particular, the curves must be viewed as elements of some space. Several modeling frameworks have been proposed, including the semimetric spaces, see Ferraty and Vieu [5], Sobolev spaces, see for example, Li and Hsing [32], and more general Besov spaces, see for example, Abramovich and Angelini [33] and Pensky and Sapatinas [34]. Sobolev spaces are Hilbert spaces in which derivatives enter into the definition of the inner product. They are used to study procedures which involve smoothing by penalizing functions with large integrated derivatives of higher order (typically order two). Besov spaces have been used to study wavelet expansions of functions, they are typically Banach spaces. In this

paper, we consider only the simplest framework in which the curves are assumed to belong to the space  $L^2(\mathcal{T})$  of square integrable functions on  $\mathcal{T}$ , and, to be more specific, we assume that  $\mathcal{T} = [0, 1]$ . All results we state in the following that involve inner products hold in an arbitrary separable Hilbert space. However, when we use integrals and functions with an argument like  $t$  or  $s$ , the formulas assume that the Hilbert space is  $L^2([0, 1])$ . This space is sufficient to describe procedures based on functional means and variances; most inferential procedures for scalar data use only means and variances. The corresponding functional framework is introduced in Section 2.1. In Section 2.2, we define the functional principal components which provide a dimension reduction beyond that obtained via the basis expansion (1.1);  $K$  is typically around one hundred, whereas in our applications the number of the functional principal components needed to effectively approximate the data is a single digit number, very often 2, 3, or 4.

In Sections 2.1 and 2.2, we define the required first- and second-order characteristics (functional parameters) of a random function  $X$ . We cannot start with definitions based on a random sample because for dependent data sample statistics often need to be defined in a different way to ensure that they estimate the corresponding population (theoretical) parameters.

### 2.1. The Hilbert Space Model

The space  $L^2 = L^2([0, 1])$  is the set of measurable real-valued functions  $x$  defined on  $[0, 1]$  satisfying  $\int_0^1 x^2(t)dt < \infty$ . The space  $L^2$  is a separable Hilbert space with the inner product

$$\langle x, y \rangle = \int x(t)y(t)dt. \quad (2.1)$$

An integral sign without the limits of integration is meant to denote the integral over the whole interval  $[0, 1]$ .

We view a random curve  $X = \{X(t), t \in [0, 1]\}$  as a random element of  $L^2$  equipped with the Borel  $\sigma$ -algebra. We say that  $X$  is integrable if  $E\|X\| = E[\int X^2(t)dt]^{1/2} < \infty$ . If  $X$  is integrable, there is a unique function  $\mu \in L^2$  such that  $E\langle y, X \rangle = \langle y, \mu \rangle$  for any  $y \in L^2$ . It follows that  $\mu(t) = E[X(t)]$  for almost all  $t \in [0, 1]$ . The expectation commutes with bounded operators, that is, if  $\Psi \in \mathcal{L}$  and  $X$  is integrable, then  $E\Psi(X) = \Psi(EX)$ .

If  $X$  is square integrable, that is,

$$E\|X\|^2 = E \int X^2(t)dt < \infty, \quad (2.2)$$

and  $EX = \mu$ , the covariance operator of  $X$  is defined by

$$C(y) = E[\langle (X - \mu), y \rangle (X - \mu)], \quad y \in L^2. \quad (2.3)$$

It is easy to see that

$$C(y)(t) = \int c(t, s)y(s)ds, \quad \text{where } c(t, s) = E[(X(t) - \mu(t))(X(s) - \mu(s))]. \quad (2.4)$$

One can show that a bounded operator  $C$  is a covariance operator of some square integrable random function taking values in  $L^2$  if and only if it is symmetric positive-definite and its eigenvalues satisfy  $\sum_{j=1}^{\infty} \lambda_j < \infty$ . To understand these statements, some background on operators in a Hilbert space is needed. The facts we now state will be used in the remainder of the paper.

### 2.1.1. Operators in a Hilbert Space

Consider an arbitrary separable Hilbert space  $H$  with inner product  $\langle \cdot, \cdot \rangle$  which generates the norm  $\| \cdot \|$ , and denote by  $\mathcal{L}$  the space of bounded (continuous) linear operators on  $H$  with the norm

$$\|\Psi\|_{\mathcal{L}} = \sup\{\|\Psi(x)\| : \|x\| \leq 1\}. \quad (2.5)$$

An operator  $\Psi \in \mathcal{L}$  is said to be *compact* if there exist two orthonormal bases  $\{v_j\}$  and  $\{f_j\}$ , and a real sequence  $\{\lambda_j\}$  converging to zero, such that

$$\Psi(x) = \sum_{j=1}^{\infty} \lambda_j \langle x, v_j \rangle f_j, \quad x \in H. \quad (2.6)$$

The  $\lambda_j$  may be assumed positive because one can replace  $f_j$  by  $-f_j$ , if needed. A compact operator is said to be a *Hilbert-Schmidt* operator if  $\sum_{j=1}^{\infty} \lambda_j^2 < \infty$ . The space  $\mathcal{S}$  of Hilbert-Schmidt operators is a separable Hilbert space with the scalar product

$$\langle \Psi_1, \Psi_2 \rangle_{\mathcal{S}} = \sum_{i=1}^{\infty} \langle \Psi_1(e_i), \Psi_2(e_i) \rangle, \quad (2.7)$$

where  $\{e_i\}$  is an arbitrary orthonormal base, the value of (2.7) does not depend on it. One can show that  $\|\Psi\|_{\mathcal{S}}^2 = \sum_{j \geq 1} \lambda_j^2$  and  $\|\Psi\|_{\mathcal{L}} \leq \|\Psi\|_{\mathcal{S}}$ .

An operator  $\Psi \in \mathcal{L}$  is said to be *symmetric* if

$$\langle \Psi(x), y \rangle = \langle x, \Psi(y) \rangle, \quad x, y \in H, \quad (2.8)$$

and positive-definite if

$$\langle \Psi(x), x \rangle \geq 0, \quad x \in H. \quad (2.9)$$

(An operator with the last property is sometimes called positive semidefinite, and the term positive-definite is used when  $\langle \Psi(x), x \rangle > 0$  for  $x \neq 0$ .)

A symmetric positive-definite Hilbert-Schmidt operator  $\Psi$  admits the decomposition

$$\Psi(x) = \sum_{j=1}^{\infty} \lambda_j \langle x, v_j \rangle v_j, \quad x \in H, \quad (2.10)$$

with orthonormal  $v_j$  which are the eigenfunctions of  $\Psi$ , that is,  $\Psi(v_j) = \lambda_j v_j$ . The  $v_j$  can be extended to a basis by adding a complete orthonormal system in the orthogonal complement

of the subspace spanned by the original  $v_j$ . The  $v_j$  in (2.10) can thus be assumed to form a basis, but some  $\lambda_j$  may be zero.

An important class of operators in  $L^2$  are the integral operators defined by

$$\Psi(x)(t) = \int \varphi(t, s)x(s)ds, \quad x \in L^2, \quad (2.11)$$

with the real kernel  $\varphi(\cdot, \cdot)$ . Such operators are Hilbert-Schmidt if and only if

$$\iint \varphi^2(t, s)dt ds < \infty, \quad (2.12)$$

in which case

$$\|\Psi\|_S^2 = \iint \varphi^2(t, s)dt ds. \quad (2.13)$$

If  $\varphi(s, t) = \varphi(t, s)$  and  $\iint \varphi(t, s)x(t)x(s)dt ds \geq 0$ , the integral operator  $\Psi$  is symmetric and positive-definite, and it follows from (2.10) that

$$\varphi(t, s) = \sum_{j=1}^{\infty} \lambda_j v_j(t)v_j(s) \quad \text{in } L^2([0, 1] \times [0, 1]). \quad (2.14)$$

## 2.2. Functional Principal Components

Suppose  $X$  is a square integrable random function in  $L^2$ . To lighten the notation, assume that  $\mu = 0$ . It is clear how to modify all formulas by subtracting the mean function  $\mu$  from  $X$ . The eigenfunctions of the covariance operator defined by (2.3) are called the *Functional Principal Components* (FPCs). The FPCs  $v_j$ ,  $j \geq 1$ , are thus defined by

$$C(v_j) = \lambda_j v_j. \quad (2.15)$$

The  $v_j$  are orthogonal, and we assume that they are normalized to unit norm. The  $v_j$  are defined only up to a sign.

The random function  $X$  admits the Karhunen-Loève expansion:

$$X(t) = \sum_{j=1}^{\infty} \xi_j v_j(t), \quad \xi_j = \langle X, v_j \rangle, \quad (2.16)$$

and the covariance kernel  $c(\cdot, \cdot)$  admits the expansion:

$$c(t, s) = \sum_{j=1}^{\infty} \lambda_j v_j(t)v_j(s), \quad 0 \leq t, s \leq 1. \quad (2.17)$$

The random variables  $\xi_j$  are called the scores of  $X$ .

An interpretation of the  $v_j$  is often based on the following decomposition of variance:

$$E\|X\|^2 = \sum_{j=1}^{\infty} E[\langle X, v_j \rangle^2] = \sum_{j=1}^{\infty} \langle Cv_j, v_j \rangle = \sum_{j=1}^{\infty} \lambda_j. \quad (2.18)$$

Another interpretation is the following. Fix  $p$  and consider the expected squared error  $E\|X - \sum_{j=1}^p a_j u_j\|^2$ , where  $u_1, u_2, \dots, u_p$  are orthonormal functions in  $L^2$ , and  $a_j$  are scalars. The above error is minimized if  $u_j = v_j$  and  $a_j = \langle X, v_j \rangle$ .

### 3. Functional Time Series

Functional time series (FTSs) were introduced in Section 1, were several examples were given. Using the framework of Section 2.1, we will understand by a FTS a sequence of curves  $X_n = \{X_n(t), 0 \leq t \leq 1\}$ . Notice that  $n$  is the time index, which typically refers to day or year, and  $t$  is the time within that unit. In most textbooks and papers on time series analysis, the index  $t$  denotes time, but in FDA it denotes the argument of a function, so we use  $n$  to denote time and  $N$  to denote the number of available curves;  $N$  corresponds to the length of a time series.

The main idea behind functional time series modeling is that in many situations the time record can be split into natural intervals, and instead of modeling periodicity, we treat the curve in each interval as a whole observational unit. There are proponents and opponents of this approach. In several applications, it has been shown that the functional approach yields superior results, see for example, Antoniadis et al. [35], a few examples are also given in Bosq [9]. But is clear that in many cases the more standard time series techniques are competitive, if not superior. Functional techniques can be expected to be definitely more useful if the daily or annual curves are not defined on a grid of equidistant time points. An interesting example of such a setting is studied by Liebl [36] who considers curves defined on different intervals on different days. The daily domain interval is defined by the minimum,  $m_n$ , and maximum,  $M_n$ , of the electricity demand on day  $n$ . For the demand  $h \in [m_n, M_n]$ ,  $X_n(h)$  is the price of electricity.

Most methods described in this section are based on the FPC. In certain applications, different orthonormal systems, similar in spirit to the FPCs but more optimal for a specific application can be used. In addition to the predictive factors of Kargin and Onatski [37], we note the factors introduced by Bathia et al. [38] who considered the problem of estimating the dimension  $d$  assuming that the functional time series takes values in a finite dimensional subspace of  $L^2$ .

The remainder of this section is organized as follows. We first review the autoregressive functional model which has been, by far, most extensively used and studied. We then turn to more general ways of describing temporal dependence between functions. We conclude with a discussion of some recent developments.

A reader seeking a good reference to the fundamental concepts of time series analysis is referred to Brockwell and Davis [39] or Shumway and Stoffer [40]. More advanced treatments of the theory of linear time series are given in Brockwell and Davis [41] and Anderson [42].

### 3.1. Functional Autoregressive Process

The theory of autoregressive and more general linear processes in Hilbert and Banach spaces is developed in the monograph of Bosq [9], so here we focus only on some recent research. A sequence  $\{X_n, -\infty < n < \infty\}$  of mean zero elements of  $L^2$  follows a functional AR(1) model if

$$X_n = \Psi(X_{n-1}) + \varepsilon_n, \quad (3.1)$$

where  $\Psi \in \mathcal{L}$  and  $\{\varepsilon_n, -\infty < n < \infty\}$  is a sequence of iid mean zero errors in  $L^2$  satisfying  $E\|\varepsilon_n\|^2 < \infty$ .

#### 3.1.1. Estimation

A popular approach to the estimation of the autoregressive operator  $\Psi$  involves the FPCs, see Section 2.2. Observe first that

$$E[\langle X_n, x \rangle X_{n-1}] = E[\langle \Psi(X_{n-1}), x \rangle X_{n-1}], \quad x \in L^2. \quad (3.2)$$

Define the lag-1 autocovariance operator by

$$C_1(x) = E[\langle X_n, x \rangle X_{n+1}] \quad (3.3)$$

and denote with superscript  $\cdot^T$  the adjoint operator. Then,  $C_1^T = C\Psi^T$  because, by a direct verification,  $C_1^T = E[\langle X_n, x \rangle X_{n-1}]$ , that is,

$$C_1 = \Psi C. \quad (3.4)$$

We would like to obtain an estimate of  $\Psi$  by using a finite sample version of the relation  $\Psi = C_1 C^{-1}$ . The operator  $C$  does not however have a bounded inverse on the whole of  $H$ . To see it, recall that  $C$  admits representation (2.10), which implies that  $C^{-1}(C(x)) = x$ , where

$$C^{-1}(y) = \sum_{j=1}^{\infty} \lambda_j^{-1} \langle y, v_j \rangle v_j. \quad (3.5)$$

This makes it difficult to estimate the bounded operator  $\Psi$  using the relation  $\Psi = C_1 C^{-1}$ . A practical solution is to use only the first  $p$  most important EFPC's  $\hat{v}_j$ , and to define

$$\widehat{IC}_p(x) = \sum_{j=1}^p \hat{\lambda}_j^{-1} \langle x, \hat{v}_j \rangle \hat{v}_j. \quad (3.6)$$

The operator  $\widehat{IC}_p$  is defined on the whole of  $L^2$ , and it is bounded if  $\hat{\lambda}_j > 0$  for  $j \leq p$ . By judiciously choosing  $p$ , we hope to find a balance between retaining the relevant information in the sample, and the danger of working with the reciprocals of small eigenvalues  $\hat{\lambda}_j$ . In

formula (3.6), the  $\hat{v}_j$  are the empirical (or sample) FPCs and the  $\hat{\lambda}_j$  are their eigenvalues, both defined by the relation  $\hat{C}(\hat{v}_j) = \hat{\lambda}_j \hat{v}_j$ , where

$$\hat{C}(x) = N^{-1} \sum_{i=1}^N \langle X_i - \bar{X}_N, x \rangle (X_i - \bar{X}_N), \quad x \in L^2, \quad (3.7)$$

defines the empirical covariance operator. Direct calculations, see Chapter 13 of Horváth and Kokoszka [1], lead to the estimator

$$\hat{\Psi}_p(x) = \frac{1}{N-1} \sum_{k=1}^{N-1} \sum_{j=1}^p \sum_{i=1}^p \hat{\lambda}_j^{-1} \langle x, \hat{v}_j \rangle \langle X_k, \hat{v}_j \rangle \langle X_{k+1}, \hat{v}_i \rangle \hat{v}_i. \quad (3.8)$$

If the operator  $\Psi$  is a kernel operator defined by  $\Psi(x)(t) = \int \psi(t, s)x(s)ds$ , then the estimator (3.8) is a kernel operator with the kernel

$$\hat{\psi}_p(t, s) = \frac{1}{N-1} \sum_{k=1}^{N-1} \sum_{j=1}^p \sum_{i=1}^p \hat{\lambda}_j^{-1} \langle X_k, \hat{v}_j \rangle \langle X_{k+1}, \hat{v}_i \rangle \hat{v}_j(s) \hat{v}_i(t). \quad (3.9)$$

In some simulated cases, the surfaces  $\hat{\psi}_p(t, s)$  may not closely resemble the true surface  $\psi(t, s)$ , see Chapter 13 of Horváth and Kokoszka [1].

### 3.1.2. Prediction

The most direct use of the functional AR(1) model is to predict the curve  $X_{n+1}$ , and the most obvious method is to use the predictor  $\hat{X}_{n+1} = \hat{\Psi}_p(X_n)$ , where  $\hat{\Psi}_p$  is the estimator (3.8). For kernel operators, this predictor is given explicitly by

$$\hat{X}_{n+1}(t) = \int \hat{\psi}_p(t, s) X_n(s) ds = \sum_{k=1}^p \left( \sum_{\ell=1}^p \hat{\psi}_{k\ell} \langle X_n, \hat{v}_\ell \rangle \right) \hat{v}_k(t), \quad (3.10)$$

where

$$\hat{\psi}_{ji} = \hat{\lambda}_i^{-1} (N-1)^{-1} \sum_{n=1}^{N-1} \langle X_n, \hat{v}_i \rangle \langle X_{n+1}, \hat{v}_j \rangle. \quad (3.11)$$

Kargin and Onatski [37] proposed a more sophisticated method that uses the so called predictive factors rather than the FPCs. Predictive factors are interesting in that they are functions which can be used to expand the curves, very much like the FPCs are used, but they define directions in the space  $L^2$  which are most relevant for prediction. Using a finite sample implementation, Didericksen et al. [43] found that it does not lead to more accurate predictions. In general, predicted curves that use formula (3.10) tend to be closer to the mean curve and smoother than the actual observations. These predictions are however significantly better than just using the mean curve. A more detailed study is given in Didericksen et al. [43] and in Chapter 13 of Horváth and Kokoszka [1].

### 3.1.3. Order Determination

A generalization of the functional AR(1) model (3.1) is the FAR( $p$ ) model

$$Z_n = \sum_{j=1}^p \Phi_j(Z_{n-j}) + \varepsilon_n. \quad (3.12)$$

We now use  $Z$  instead of  $X$ , because the letter  $X$  will be used below to denote a different quantity. The work of Kokoszka and Reimherr [44] shows how to determine the order  $p$ , and how to estimate the operators  $\Phi_j$ . The idea is to write (3.12) as a fully functional linear model. We start by expressing  $\Phi_j(Z_{i-j})$  as an integral over the interval  $((j-1)/p, j/p]$ . Setting  $x := (s + j - 1)/p$ , a change of variables yields

$$[\Phi_j(Z_{i-j})](t) = \int_0^1 \phi_j(t, s) Z_{i-j}(s) ds = \int_{(j-1)/p}^{j/p} \phi_j(t, xp - (j-1)) Z_{i-j}(xp - (j-1)) p dx. \quad (3.13)$$

Denoting by  $I_j$  the indicator function of the interval  $((j-1)/p, j/p]$ , we obtain

$$\sum_{j=1}^p [\Phi_j(Z_{i-j})](t) = \int_0^1 \sum_{j=1}^p I_j(x) \phi_j(t, xp - (j-1)) Z_{i-j}(xp - (j-1)) p dx. \quad (3.14)$$

Next we define

$$\begin{aligned} X_i(s) &= \sum_{j=1}^p Z_{i-j}(sp - (j-1)) I_j(s), \\ \psi(t, s) &= p \sum_{j=1}^p \phi_j(t, sp - (j-1)) I_j(s). \end{aligned} \quad (3.15)$$

Setting  $Y_i = Z_i$ , we have

$$Y_i = \Psi(X_i) + \varepsilon_i, \quad (3.16)$$

where  $\Psi$  is an integral Hilbert-Schmidt operator with the kernel  $\psi$ , that is,

$$Y_i(t) = \int \psi(t, s) X_i(s) ds + \varepsilon_i(t). \quad (3.17)$$

Thus, if we can estimate  $\Psi$ , then we can estimate each of the  $\Phi_j$ . An estimator of  $\Psi$  can be constructed by any method valid for the fully functional linear model, for example using the basis or the FPCs expansions, see Chapter 8 of Horváth and Kokoszka [1].

Determining the order  $p$  also uses representation (3.17). The FAR( $p-1$ ) model will be rejected in favor of FAR( $p$ ) if an estimate of  $\hat{\Phi}_p$  is large in an appropriate sense. Kokoszka and

Reimherr [44] developed a multistage procedures based on this idea, but it is too complex to be described here.

### 3.2. Approximable Functional Time Series

For many functional time series, it is not clear what specific model they follow, and for many statistical procedures, it is not necessary to assume a specific model. It is however important to know what the effect of temporal dependence on a given procedure is. In this section, we describe a very general notion of dependence, which is convenient to use in the framework of functional data. We restrict ourselves to this framework. Several related useful results, including a functional (in  $D([0, 1])$ ) central limit theorem, were established by Aue et al. [45, 46] for vector-valued time series. Sharp invariance principles and additional references are given in Berkes et al. [47]. The exposition that follows is based on Hörmann and Kokoszka [48].

For  $p \geq 1$ , we denote by  $L^p = L^p(\Omega, \mathcal{A}, P)$  the space of (classes of) real-valued random variables such that  $\|X\|_p = (E|X|^p)^{1/p} < \infty$ . Further, we let  $L_H^p = L_H^p(\Omega, \mathcal{A}, P)$  be the space of  $H = L^2$  valued random functions  $X$  such that

$$v_p(X) = (E\|X\|^p)^{1/p} = \left( E \int |X(t)|^p dt \right)^{1/p} < \infty. \quad (3.18)$$

In this section, we use  $H$  to denote the function space  $L^2 = L^2([0, 1])$  to avoid confusion with the space  $L^p$  of scalar random variables.

*Definition 3.1.* A sequence  $\{X_n\} \in L_H^p$  is called  $L^p$ - $m$ -approximable if each  $X_n$  admits the representation

$$X_n = f(\varepsilon_n, \varepsilon_{n-1}, \dots), \quad (3.19)$$

where the  $\varepsilon_i$  are iid elements taking values in a measurable space  $S$ , and  $f$  is a measurable function  $f : S^\infty \rightarrow H$ . Moreover we assume that if  $\{\varepsilon'_i\}$  is an independent copy of  $\{\varepsilon_i\}$  defined on the same probability space, then letting

$$X_n^{(m)} = f(\varepsilon_n, \varepsilon_{n-1}, \dots, \varepsilon_{n-m+1}, \varepsilon'_{n-m}, \varepsilon'_{n-m-1}, \dots), \quad (3.20)$$

we have

$$\sum_{m=1}^{\infty} v_p(X_m - X_m^{(m)}) < \infty. \quad (3.21)$$

Definition 3.1 implies that  $\{X_n\}$  is strictly stationary. It is clear from the representation of  $X_n$  and  $X_n^{(m)}$  that  $E\|X_m - X_m^{(m)}\|^p = E\|X_1 - X_1^{(m)}\|^p$ , so that condition (3.21) could be formulated solely in terms of  $X_1$  and the approximations  $X_1^{(m)}$ . Obviously the sequence  $\{X_n^{(m)}, n \in \mathbb{Z}\}$  as defined in (3.20) is *not*  $m$ -dependent. To this end we need to define for

each  $n$  an independent copy  $\{\varepsilon_k^{(n)}\}$  of  $\{\varepsilon_k\}$  (this can always be achieved by enlarging the probability space) which is then used instead of  $\{\varepsilon'_k\}$  to construct  $X_n^{(m)}$ , that is, we set

$$X_n^{(m)} = f\left(\varepsilon_n, \varepsilon_{n-1}, \dots, \varepsilon_{n-m+1}, \varepsilon_{n-m}^{(n)}, \varepsilon_{n-m-1}^{(n)}, \dots\right). \quad (3.22)$$

We call this method the *coupling construction*. Since this modification lets condition (3.21) unchanged, we will assume from now on that the  $X_n^{(m)}$  are defined by (3.22). Then, for each  $m \geq 1$ , the sequences  $\{X_n^{(m)}, n \in \mathbb{Z}\}$  are strictly stationary and  $m$ -dependent, and each  $X_n^{(m)}$  is equal in distribution to  $X_n$ .

The following example gives a good feel what Definition 3.1 means.

*Example 3.2* (functional autoregressive process). Suppose that  $\Psi \in \mathcal{L}$  satisfies  $\|\Psi\|_{\mathcal{L}} < 1$ . Let  $\varepsilon_n \in L_H^2$  be iid with mean zero. Then there is a unique stationary sequence of random elements  $X_n \in L_H^2$  such that

$$X_n(t) = \Psi(X_{n-1})(t) + \varepsilon_n(t). \quad (3.23)$$

The AR(1) sequence (3.23) admits the expansion  $X_n = \sum_{j=0}^{\infty} \Psi^j(\varepsilon_{n-j})$ , where  $\Psi^j$  is the  $j$ th iterate of the operator  $\Psi$ . We thus set  $X_n^{(m)} = \sum_{j=0}^{m-1} \Psi^j(\varepsilon_{n-j}) + \sum_{j=m}^{\infty} \Psi^j(\varepsilon_{n-j}^{(n)})$ . It is easy to verify that for every  $A$  in  $\mathcal{L}$ ,  $\nu_p(A(Y)) \leq \|A\|_{\mathcal{L}} \nu_p(Y)$ . Since  $X_m - X_m^{(m)} = \sum_{j=m}^{\infty} (\Psi^j(\varepsilon_{m-j}) - \Psi^j(\varepsilon_{m-j}^{(m)}))$ , it follows that  $\nu_p(X_m - X_m^{(m)}) \leq 2 \sum_{j=m}^{\infty} \|\Psi\|_{\mathcal{L}}^j \nu_p(\varepsilon_0) = O(1) \nu_p(\varepsilon_0) \|\Psi\|_{\mathcal{L}}^m$ . By assumption  $\nu_2(\varepsilon_0) < \infty$  and therefore  $\sum_{m=1}^{\infty} \nu_2(X_m - X_m^{(m)}) < \infty$ , so condition (3.21) holds with  $p \geq 2$ , as long as  $\nu_p(\varepsilon_0) < \infty$ .

Several other models are  $L^p$ - $m$ -approximable, examples are given in Chapter 16 of Horváth and Kokoszka [1]. But we emphasize that the main role of  $L^p$ - $m$ -approximability is to provide a convenient nonparametric framework for quantifying temporal dependence of curves. In this sense it is similar to various mixing conditions, but there is no obvious relationship between it and the traditional mixing conditions, as defined by Doukhan [49] or Bradley [50]. It is not related to the conditions of Doukhan and Louhichi [51] either. It is similar to the conditions of Wu [52, 53], in that it assumes that the observed functions are nonlinear moving averages (Bernoulli shifts) of iid unobservable errors.

Recently Lian [54] extended the concept of  $L^p$ - $m$ -approximability by replacing the  $L^p$  norm by a more general Orlicz norm. If  $\psi$  is a convex function on  $[0, \infty)$  with  $\psi(0) = 0$ , then we define

$$\nu_{\psi}(X) = \inf \left\{ C > 0, E \left[ \psi \left( \frac{\|X\|}{C} \right) \right] \leq 1 \right\}. \quad (3.24)$$

If  $\psi(x) = x^p$ , then  $\nu_{\psi} = \nu_p$ . The remaining elements of Definition 3.1 are extended without difficulty.

### 3.2.1. Convergence of the Eigenfunctions and Eigenvalues

An important result valid for  $L^4$ - $m$ -approximable FTSs is that the empirical eigenfunctions and eigenvalues converge at the same rate as for the iid functions. This property allows us to extend many results established for functional random samples to FTSs. We state this result in Theorem 3.3, in which  $\hat{v}_j$  and  $\hat{\lambda}_j$  are the eigenelements of the empirical covariance operator (3.7).

**Theorem 3.3.** *Suppose that  $\{X_n\} \in L^4_H$  is an  $L^4$ - $m$ -approximable sequence such that the largest  $d$  eigenvalues  $\lambda_j$  of its covariance operator are positive and distinct. Then, for  $1 \leq j \leq d$ ,*

$$\limsup_{N \rightarrow \infty} NE \left[ \left| \lambda_j - \hat{\lambda}_j \right|^2 \right] < \infty, \quad \limsup_{N \rightarrow \infty} NE \left[ \left\| \hat{c}_j \hat{v}_j - v_j \right\|^2 \right] < \infty. \quad (3.25)$$

Kokoszka and Reimherr [55] showed that under the assumptions of Theorem 3.3 the  $\hat{v}_j$  are asymptotically normal. For linear functional series this asymptotic normality is implicit in the work of Mas [56].

### 3.2.2. The Long-Run Variance

A central concept in time series analysis is the long run variance (LRV) which replaces the variance in many well-known formulas for valid for iid observations. Let  $\{X_n\}$  be a scalar stationary sequence. Its long run variance is defined as  $\sigma^2 = \sum_{j \in \mathbb{Z}} \gamma_j$ , where  $\gamma_j = \text{Cov}(X_0, X_j)$ , provided this series is absolutely convergent. In that case  $\text{Var}[\bar{X}_N] \sim \sigma^2/N$ . The concept of the LRV is extended to stationary random vectors. It turns out that  $L^2$ - $m$ -approximability implies that the LRV of exists and can be consistently estimated for vector-valued sequences. Details are given in Chapter 16 of Horváth and Kokoszka [1]. In the functional context, it is more natural to work with long-run variance kernels. The following results were established by Horváth et al. [57].

**Theorem 3.4.** *Suppose that  $\{X_n\}$  is an  $L^2$ - $m$ -approximable sequence of functions in  $L^2([0, 1])$  with  $\mu = EX_n$ . Then the infinite sum*

$$\sigma(t, s) = \sum_{h=-\infty}^{\infty} E[\{X_0(t) - \mu(t)\}\{X_h(s) - \mu(s)\}] \quad (3.26)$$

*converges in  $L^2([0, 1] \times [0, 1])$  (hence  $\sigma(\cdot, \cdot)$  is square integrable). Moreover,*

$$N^{-1/2} \sum_{n=1}^N [X_n - \mu] \xrightarrow{d} Z, \quad (3.27)$$

*where  $Z$  is a mean zero Gaussian element of  $L^2([0, 1])$  with the covariance function  $E[Z(t)Z(s)] = \sigma(t, s)$ .*

To apply Theorem 3.4, we must estimate the kernel  $\sigma(\cdot, \cdot)$ . We now turn to this objective. Let  $K$  be a kernel (weight function) defined on the line and satisfying the following conditions:

$$\begin{aligned} K(0) &= 1, \quad K \text{ is continuous and bounded,} \\ K(u) &= 0, \quad \text{if } |u| > c, \text{ for some } c > 0. \end{aligned} \quad (3.28)$$

Condition (3.28) is assumed only to simplify the proofs, a sufficiently fast decay could be assumed instead. Next we define the empirical (sample) correlation functions

$$\hat{\gamma}_i(t, s) = \frac{1}{N} \sum_{j=i+1}^N (X_j(t) - \bar{X}_N(t)) (X_{j-i}(s) - \bar{X}_N(s)), \quad 0 \leq i \leq N-1, \quad (3.29)$$

where  $\bar{X}_N(t) = N^{-1} \sum_{1 \leq i \leq N} X_i(t)$ . The estimator for  $\sigma(t, s)$  is given by

$$\hat{\sigma}_N(t, s) = \hat{\gamma}_0(t, s) + \sum_{i=1}^{N-1} K\left(\frac{i}{h}\right) (\hat{\gamma}_i(t, s) + \hat{\gamma}_i(s, t)), \quad (3.30)$$

where  $h = h(N)$  is the smoothing bandwidth satisfying

$$h(N) \longrightarrow \infty, \quad \frac{h(N)}{N} \longrightarrow 0, \quad \text{as } N \longrightarrow \infty. \quad (3.31)$$

**Theorem 3.5.** *Suppose that the functional time series  $\{X_n\}$  is  $L^2$ - $m$ -approximable and*

$$\lim_{m \rightarrow \infty} m \left[ E \int (X_n(t) - X_n^{(m)}(t))^2 dt \right]^{1/2} = 0. \quad (3.32)$$

*Then, under the conditions on the kernel  $K(\cdot)$  and the bandwidth  $h = h(N)$  stated above,*

$$\iint (\hat{\sigma}_N(t, s) - \sigma(t, s))^2 dt ds \xrightarrow{P} 0. \quad (3.33)$$

### 3.2.3. Self-Normalization

Estimation of the LRV, even for scalar time series, is difficult due to the selection of the bandwidth  $h$ . Data-driven methods exist, most notably those based on the theory of Andrews [58], but they do postulate a specific form of dependence. Moreover, some tests in which the LRV is estimated by such data driven procedures suffer from the problem of nonmonotonic power, see Shao and Zhang [59]. Block bootstrap and subsampling offer useful tools for solving many problems of time series analysis, but they again depend on bandwidth type block sizes whose selection is not obvious. Even though some recommendations for the selections of such block sizes exist, see for example, Bühlmann and Künsch [60], Politis et al.

[61], Bickel and Sakov [62], this remains a difficult practical problem. An improved approach has recently been proposed by Shao and Politis [63]. A different idea, the self-normalization, was proposed by Shao [64, 65]. We explain it now briefly using a simple example involving a scalar time series.

Suppose that  $\{X_n\}$  is a stationary time series such that

$$N^{-1/2} \sum_{1 \leq k \leq Nr} (X_k - \mu) \xrightarrow{d} \sigma W(r), \quad 0 \leq r \leq 1, \quad (3.34)$$

in the Skorokhod space. The parameter  $\sigma^2$  is the LRV:

$$\sigma^2 = \lim_{N \rightarrow \infty} N \text{Var}(\bar{X}_N) = \sum_h \gamma(h). \quad (3.35)$$

Set

$$D_N = N^{-2} \sum_{n=1}^N \left\{ \sum_{j=1}^n (X_j - \bar{X}_N) \right\}^2. \quad (3.36)$$

Then, (3.34) implies

$$\frac{N(\bar{X}_N - \mu)^2}{D_N} \xrightarrow{d} \frac{W^2(1)}{\int_0^1 B^2(r) dr}. \quad (3.37)$$

To see why (3.37) holds, set

$$S_N(r) = N^{-1/2} \sum_{1 \leq k \leq Nr} (X_k - \mu), \quad 0 \leq r \leq 1, \quad (3.38)$$

and observe that

$$S_N(1) = N^{1/2}(\bar{X}_N - \mu), \quad (3.39)$$

so that

$$N(\bar{X}_N - \mu)^2 = S_N^2(1) \xrightarrow{d} \sigma^2 W^2(1). \quad (3.40)$$

Next, observe that

$$\begin{aligned}\sum_{j=1}^n (X_j - \bar{X}_N) &= \sum_{j=1}^n (X_j - \mu) - n(\bar{X}_N - \mu); \\ N^{-1/2} \sum_{j=1}^n (X_j - \bar{X}_N) &= S_N\left(\frac{n}{N}\right) - \frac{n}{N} S_N(1).\end{aligned}\tag{3.41}$$

Consequently,

$$D_N = N^{-1} \sum_{n=1}^N \left\{ S_N\left(\frac{n}{N}\right) - \frac{n}{N} S_N(1) \right\}^2 \xrightarrow{d} \sigma^2 \int_0^1 \{W(r) - rW(1)\}^2 dr.\tag{3.42}$$

The convergences in (3.40) and (3.42) are joint, so (3.37) follows.

The key point is the cancelation of  $\sigma^2$  when (3.40) is divided by (3.42). Relation (3.42) shows that  $D_N$  is an inconsistent estimator of  $\sigma^2$ . The distribution of the right-hand side of (3.37) can however be simulated, and the critical values can be obtained with arbitrary precision. Relation (3.37) can be used to construct a confidence interval for  $\mu$  without estimating the LRV. Such a construction does not require the selection of a bandwidth parameter in the kernel estimates of  $\sigma^2$ .

Zhang et al. [66] extend the idea of self-normalization to the detection of change point in FTSs. A normalization analogous to  $D_N$  is not suitable for the change point problem, see Shao and Zhang [59]. It must be modified to take into account a possible change point. The details are too complex to be discussed here. The change point problem is discussed in Section 3.3.

### 3.3. Other Research Directions

#### 3.3.1. Two-Sample Problems

In a two-sample problem, we consider two samples  $X_1, \dots, X_N$  and  $X_1^*, \dots, X_M^*$ , which may be realizations of two functional time series, or a single series, but taken over different time intervals. The statistical problem consists in testing if some specified characteristics of the samples are equal. To illustrate, suppose we want to test if the mean functions are equal. We assume the model

$$\begin{aligned}X_i(t) &= \mu(t) + \varepsilon_i(t), \quad 1 \leq i \leq N, \\ X_i^*(t) &= \mu^*(t) + \varepsilon_i^*(t), \quad 1 \leq i \leq M.\end{aligned}\tag{3.43}$$

We wish to test the null hypothesis  $H_0 : \mu = \mu^*$  against the alternative that  $H_0$  is false. We can also test for the equality of the covariance operators, or any other quantities that might be of interest. In the FDA setting, testing for the equality of means, the FPCs, and the covariance operators has received a fair deal of attention.

An important contribution has been made by Benko et al. [67] who developed bootstrap procedures for testing the equality of mean functions, the FPCs, and the eigenspaces

spanned by them. Horváth et al. [57] developed asymptotic tests for testing the equality of mean functions. Laukaitis and Račkauskas [68] considered the model  $X_{g,i}(t) = \mu_g(t) + \varepsilon_{g,i}(t)$ ,  $g = 1, 2, \dots, G$ , with innovations  $\varepsilon_{g,i}$  and group means  $\mu_g$ , and test  $H_0 : \mu_1(t) = \dots = \mu_G(t)$ . Other related contributions are Cuevas et al. [69], Delicado [70], and Ferraty et al. [71].

Regarding testing the equality of the covariance operators, Panaretos et al. [72] developed a test assuming the data are iid and Gaussian. Their work was extended by Fremdt et al. [73] and Kraus and Panaretos [74] to non-Gaussian settings. Horváth et al. [75] studied two sample problems for functional regression models.

### 3.3.2. The Change Point Problem

When curves form a time series, it is typically assumed that, possibly after some transformation, they have the same distribution in  $L^2$ . If some aspects of this distribution change, inference will become invalid. For example, if a mean changes at some time point, the sample mean will not estimate any population parameter. Change point analysis is a very broad field of statistics with applications in industrial quality control, study of economic, financial and environmental time series, and many other areas. Many monographs are available, including Brodsky and Darkhovsky [76], Basseville and Nikifirov [77], Csörgő and Horváth [78], Chen and Gupta [79], and Basseville et al. [80].

The idea of change point analysis is simple, and we explain it in the functional context using the change in mean function as an example. We observe functions  $X_i$ ,  $1 \leq i \leq N$ , and we test the null hypothesis

$$H_0 : EX_1 = EX_2 = \dots = EX_N. \quad (3.44)$$

The simplest alternative is that of a single change point:

$$H_A : X_i(t) = \begin{cases} \mu_1(t) + Y_i(t), & 1 \leq i \leq k^*, \\ \mu_2(t) + Y_i(t), & k^* < i \leq N, \end{cases} \quad (3.45)$$

where the  $Y_i$  has the same distribution with *mean zero*, and  $\mu_1(\cdot) \neq \mu_2(\cdot)$ .

Note that under  $H_0$ , we do not specify the value of the common mean, and under  $H_A$  we do not specify  $\mu_1(\cdot)$  nor  $\mu_2(\cdot)$ . It is also important to distinguish between a change point problem and the problem of testing for the equality of means. In the latter setting, it is known which population or group each observation belongs to. In the change point setting, we do not have any a priori partition of the data into several sets with possibly different means. The change can occur at any point, and we want to test if it occurs or not, and, if it does, to estimate the point of change.

The simplest, and in many settings most effective, change point detection procedures are based on cumulative sums (CUSUM procedures). In the above setting, denote

$$\hat{\mu}_k(t) = \frac{1}{k} \sum_{1 \leq i \leq k} X_i(t), \quad \tilde{\mu}_k(t) = \frac{1}{N-k} \sum_{k < i \leq N} X_i(t). \quad (3.46)$$

If the mean is constant, the difference  $\Delta_k(t) = \hat{\mu}_k(t) - \tilde{\mu}_k(t)$  is small for all  $1 \leq k < N$  and all  $t \in [0, 1]$ . However,  $\Delta_k(t)$  can become large due to chance variability if  $k$  is close to 1 or to  $N$ . It is therefore usual to work with the sequence

$$P_k(t) = \sum_{1 \leq i \leq k} X_i(t) - \frac{k}{N} \sum_{1 \leq i \leq N} X_i(t) = \frac{k(N-k)}{N} [\hat{\mu}_k(t) - \tilde{\mu}_k(t)], \quad (3.47)$$

in which the variability at the end points is attenuated by a parabolic weight function. If the mean changes, the difference  $P_k(t)$  is large for some values of  $k$  and of  $t$ . Since the observations are in an infinite dimensional domain, projections on the FPCs are used to construct a test statistic and an estimator of a change point. The details are explained in Berkes et al. [81] who assume that the  $X_i$  are independent; an extension to approximable sequences is developed in Hörmann and Kokoszka [48]. Theoretical properties of the estimator of a change point are studied in Aue et al. [45, 46].

There have been several extensions to more complex settings. Horváth et al. [82] consider the change point in the covariance operator of the FAR(1) process. Zhang et al. [66] use self-normalization to deal with temporal dependence and study change points in mean and in the covariance structure. Aston and Kirch [83] consider change point detection in a dependent sequence of functions assuming that the alternative is not a single change point but a change interval, the so-called epidemic alternative. The problem of detecting a change in the mean of a sequence of Banach-space valued random elements under an epidemic alternative is theoretically studied by Račkauskas and Suquet [84] who propose a statistic based on increasingly fine dyadic partitions of the index interval, and derive its limit, which is nonstandard.

### 3.3.3. Other Temporal Dependence Structures

The most extensively used and studied model for FTS is the FAR(1) model. Some nonlinear ARCH-type models are introduced in Hörmann et al. [85]. Similar models are used by Kokoszka and Reimherr [86] in a simulation study. Kokoszka and Reimherr [86] study the predictability (not the prediction) of the cumulative intraday returns (CIDRs) defined by

$$R_n(t_j) = 100[\ln P_n(t_j) - \ln P_n(t_1)], \quad j = 2, \dots, m, \quad n = 1, \dots, N, \quad (3.48)$$

where  $P_n(t_j)$ ,  $n = 1, \dots, N$ ,  $j = 1, \dots, m$  is the price of a financial asset at time  $t_j$  on day  $n$ .

Kokoszka et al. [87], see also Kokoszka and Zhang [88], study the dependence of the CIDRs defined above on other factors, including the CIDRs on market indexes and energy futures. They consider a factor model of the form:

$$R_n(t) = \beta_0(t) + \sum_{j=1}^p \beta_j F_{nj}(t) + \varepsilon_n(t). \quad (3.49)$$

A different class of functional factor models was recently proposed by Hays et al. [89]. These models, introduced to forecast yield curves, say  $X_n$ , are of the form

$$X_n(t) = \sum_{k=1}^K \beta_{nk} F_k(t) + \varepsilon_n(t). \quad (3.50)$$

The factors  $F_k$  do not depend on  $n$  and are *orthonormal* functions to be estimated. The dynamics are in the coefficients  $\beta_{nk}$  which are assumed to follow Gaussian autoregressive processes (the  $\varepsilon_n$  are also Gaussian). Model (3.50) could be termed a statistical factor model. It is designed for temporal forecasting, while model (3.49) is designed for regression type prediction in which that correlation structure of factors plays a major role.

Gabrys et al. [90] developed a test to determine if the regressors  $\varepsilon_n$  in the fully functional linear model  $Y_n(t) = \int \varphi(t, s) X_n(s) ds + \varepsilon_n(t)$  are correlated. Benhenni et al. [91] consider a nonparametric regression  $Y_n = r(X_n) + \varepsilon_n$  in which  $X_n$  are functions and  $Y_n$  (and  $\varepsilon_n$ ) scalars. They allow the  $\varepsilon_n$  to have either short or long memory.

A different framework is developed by Battey and Sancetta [92] who assume that the FTS  $X_n$  is a Markov process and are concerned with the estimation of the conditional probabilities  $P(X_n(t) \leq x(t) \mid X_{n-1} = x_0)$ , where  $x$  and  $x_0$  are deterministic functions.

## 4. Geostatistical Functional Data

An interesting class of functional data are curves observed at several spatial locations, as already mentioned in Section 1. For example,  $X(\mathbf{s}_k; t)$  can be the concentration of a pollutant at location  $\mathbf{s}_k$  and measured at time  $t$ . In this section, we thus assume that the data consists of  $N$  curves  $X(\mathbf{s}_k; t)$ ,  $t \in [0, 1]$ ,  $1 \leq k \leq N$ . The analysis of such a data structure must draw heavily on the concepts and tools of spatial statistics. In fact, the random field  $X$  defined above is a special case of a spatiotemporal process. In this paper, we cannot provide the details of the tools of spatial statistics that we use. The required background, and much more, is given, for example, in Schabenberger and Gotway [93], Gelfand et al. [94], and Sherman [95]. We will however introduce the concepts to the extent needed to get a good idea of the main problems and solutions. This is done in Section 4.1. Then, in Section 4.2, we consider the fundamental problem of the estimation of the mean function. Section 4.3 points out to research on functional kriging. We note that there are functional data structures beyond those discussed here, we refer to Delicado et al. [96], who also discuss Bayesian approaches.

### 4.1. Basic Concepts of Spatial Statistics

A sample of spatial data is  $\{X(\mathbf{s}_k), \mathbf{s}_k \in \mathbf{S}, k = 1, 2, \dots, N\}$ . (The  $X(\mathbf{s}_k)$  are now scalars.) The spatial domain  $\mathbf{S}$  is typically a subset of the two-dimensional plane or sphere. The observed value  $X(\mathbf{s}_k)$  is viewed as a realization of a random variable. Just as in time series analysis, stationary random fields play a fundamental role in modeling spatial data. To define arbitrary shifts, we must assume that  $\mathbf{S}$  is either the whole Euclidean space  $\mathbb{R}^d$ , or the whole sphere. The random field  $\{X(\mathbf{s}), \mathbf{s} \in \mathbf{S}\}$  is then strictly stationary if

$$\{X(\mathbf{s}_1 + \mathbf{h}), X(\mathbf{s}_2 + \mathbf{h}), \dots, X(\mathbf{s}_m + \mathbf{h})\} \stackrel{d}{=} \{X(\mathbf{s}_1), X(\mathbf{s}_2), \dots, X(\mathbf{s}_m)\} \quad (4.1)$$

for any points  $\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_m \in \mathbf{S}$  and any shift  $\mathbf{h}$ . The covariance function is then defined by

$$C(\mathbf{h}) = \text{Cov}(X(\mathbf{s}), X(\mathbf{s} + \mathbf{h})). \quad (4.2)$$

If  $C(\mathbf{h})$  depends only on the length  $h$  of  $\mathbf{h}$ , we say that the random field is isotropic. The covariance function of an isotropic random field is typically parametrized as

$$C(h) = \sigma^2 \phi(h), \quad h \geq 0, \quad \phi(0) = 1. \quad (4.3)$$

The function  $\phi$  is called the correlation function. For example, the *exponential* correlation function is given by  $\phi(h) = \exp\{-(h/\rho)\}$ . The main idea of spatial modeling is that some sort of empirical estimate of  $C(h)$  is first obtained, which is typically a very irregular, noisy function. Then a suitable parametric covariance function  $C(h)$  is fitted to ensure that the resulting covariance matrix  $\{C(\|\mathbf{s}_k - \mathbf{s}_\ell\|), 1 \leq k, \ell \leq N\}$  is positive definite. There are many approaches and nontrivial issues involved in this process.

#### 4.2. Estimation of the Mean Function and of the FPCs

We assume that the functions  $X(\mathbf{s}_k, \cdot)$  are observations of a stationary and isotropic spatial random field taking values in the space  $L^2$ . In particular, the mean function  $\mu(t) = EX(\mathbf{s}; t)$  does not depend on the location  $\mathbf{s}$ . The estimation of this mean function is the most important first step of the inference for spatially distributed functional data. If the curves  $X_k$  are iid or form a time series, the usual sample mean is used. But if the curves are available at spatial locations, the curves at locations which are close to each other should be given smaller weights because they are very similar and contribute roughly the same information. This suggests that  $\mu$  could be estimated by

$$\hat{\mu}_N = \sum_{n=1}^N w_n X(\mathbf{s}_n), \quad (4.4)$$

with the weights  $w_k$  defined to minimize  $E\|\sum_{n=1}^N w_n X(\mathbf{s}_n) - \mu\|^2$  subject to the condition  $\sum_{n=1}^N w_n = 1$ . Using the method of the Lagrange multiplier, it is not difficult to find a closed form expression for the weights  $w_n$ . This expression involves unknown covariances

$$C_{k\ell} = E[\langle X(\mathbf{s}_k) - \mu, X(\mathbf{s}_\ell) - \mu \rangle]. \quad (4.5)$$

Gromenko et al. [97] proposed an iterative procedure for the estimation of the weights  $w_n$  and the covariances  $C_{k\ell}$ . They showed that the weighted average (4.4) is a significantly better estimator than the usual sample average. Hörmann and Kokoszka [98] showed that the sample average is not a consistent estimator of  $\mu$  if there are clusters of points. The clusters were defined using a modification of a metric introduced by Park et al. [99]. Gromenko and Kokoszka [100] refined the approach of Gromenko et al. [97] and applied it to testing the hypothesis that the means in the samples of curves over two disjoint regions are different.

The approaches outlined above use parametric model fitting to obtain the covariances  $C_{k\ell}$ . It is well-known that if the number of the spatial locations is small, less than 20 ~ 30, then

the optimization needed to fit a parametric model may fail or the fit may be poor. Gromenko and Kokoszka [101] addressed this issue by developing a nonparametric approach. They showed that in relevant settings (all with  $10 \leq N \leq 40$ ), the nonparametric approach is superior to the parametric approach even excluding the cases when the former does not converge. The method of Gromenko and Kokoszka [101] is a useable refinement and extension of the ideas of Hall et al. [102] and Hall and Patil [103].

#### 4.2.1. Estimation of the FPCs

If the random functions  $X(\mathbf{s}_k, \cdot)$  are realizations of a stationary  $L^2$ -valued random field, then they have the same FPCs  $v_j$ . Estimation of the  $v_j$  in the spatial setting involves issues similar to those related in the estimation of the mean function  $\mu$ , but the notation and approaches are more complex. They are discussed in Gromenko et al. [97] and Hörmann and Kokoszka [98].

### 4.3. Kriging and Other Research

A very important problem is to predict a curve at a specified location using the curves at available locations. This problem was addressed in Nerini et al. [104] and Giraldo [105]. Spatial prediction of this type is called kriging. Theoretical background to kriging of scalar fields is given in Stein [106]. The book of Wackernagel [107] gives an accessible introduction to kriging emphasizing multidimensional observations, a setting related to kriging of curves. Just as in the case of estimating the mean function, there are several approaches to kriging. The approach advocated by Giraldo et al. [105] is “most functional” in that it treats the curves as whole data objects. Similarly to (4.4), it predicts the unobserved function  $X(\mathbf{s}_0)$  by a linear combination

$$\hat{X}(\mathbf{s}_0) = \sum_{n=1}^N w_n X(\mathbf{s}_n), \quad \sum_{n=1}^N w_n = 1, \quad (4.6)$$

with the weights  $w_n$  minimizing

$$\int E \left[ X(\mathbf{s}_0; t) - \sum_{n=1}^N w_n X(\mathbf{s}_n) \right]^2 dt. \quad (4.7)$$

The estimation of the weights  $w_n$  requires some work.

Giraldo et al. [108] consider the problem of determining clusters in in spatially correlated functional data. Härdle and Osipienko [109] show how spatial analysis of functional data can be used to calculate more accurate prices of weather derivatives.

An emerging important class of functional models are those with hierarchical structure and correlation at some levels, similar to spatial correlations discussed in this Section. Such models have applications in the analysis of medical experiments in which tissue samples are

taken at several locations in an organ of a subject. Staicu et al. [110] propose fact methods for the estimation of such models.

## 5. Summary and Future Directions

We have reviewed recent developments in the analysis of dependent functions. We considered two data structures: functional time series and spatially distributed functions. Functional time series are collections of curves  $X_n$ , where  $n$  can be interpreted as time, most often day or year. The central point in the analysis and modeling of such data is to take into account the dependence of the curves  $X_n$ . Spatially distributed functions, or geostatistical functional data, consist of curves  $X(\mathbf{s}_k)$  available at locations  $\mathbf{s}_k$ . The main issue is to take into account the uneven distribution in space of the points  $\mathbf{s}_k$ .

Inference for data of both types assumes that the data are stationary, possibly after some transformation. At present there are no suitable tests of stationarity for such functional data. Second-order stationarity of FTS could potentially be tested using spectral methods. For scalar time series the relevant references are Grenander and Rosenblatt [111], Granger and Hatanaka [112], Priestley et al. [113], and Dwivedi and Subba Rao [114]. The recent work of Panaretos and Tavakoli [115] develops the fundamentals of the spectral theory of FTS.

In many settings, a hybrid data of the form  $X_n(\mathbf{s}_k, t)$  can be considered. It could, for example, be maximum temperature or the count of flu cases on day  $1 \leq t \leq 365$  of year  $n$  at location  $\mathbf{s}_k$ . Such a data structure could be useful to study long term trends or changes in the annual pattern of a variable of interest over a region. There has not been much work on functional data of this form; Odei et al. [116] consider Bayesian modeling of snow melt curves. In their setting,  $X_n(\mathbf{s}_k, t)$  is the amount of snow melt water on day  $1 \leq t \leq 365$  of year  $n$  at a location  $\mathbf{s}_k$  in Utah.

For long records of geophysical, weather, or environmental data, a serious problem are long segments of missing observations. For example, if  $X(\mathbf{s}_k, t)$  is the rainfall on day  $t$  at location  $\mathbf{s}_k$ , there may be whole years of missing data at certain locations (when a station is closed), and these segments will be different at different locations. An important challenge is to develop useful approaches that allow us to pool together information from curves at all locations when some of them have long missing segments. This problem is somewhat related to the problem of dealing with sparse functional data mentioned in Section 1.1, but it adds the complication of spatial dependence.

The functions discussed in this paper, whether those observed consecutively over time or at spatial locations, are assumed to be smooth, so that methods relying on basis and FPCs expansions can be applied. Some functions do not fall into this category and may exhibit sharp spikes and flat regions. There has not been much work on the time series or spatial fields of functions of this type. Timmermans and Von Sachs [117] propose an exploratory tool that aims at detecting the closeness of curves whose significant sharp features might not be well aligned.

The study of extremes involves work with point processes. For example  $P_n(\mathbf{s}_k)$  could be a point process of threshold exceedances in year  $n$  at location  $\mathbf{s}_k$ . No systematic modeling framework for such point process valued time series is available. The estimation of exceedance probabilities is studied in Draghicescu and Ignaccolo [118].

In summary, the study of dependent functional data has reached a level of maturity that makes it a useful subfield of FDA, but many important problems remain to be addressed. It is hoped that this paper has provided a useful introduction into this area.

## Acknowledgments

This work was partially completed at the Institute for Mathematical Sciences, National University of Singapore, 2012. The research was partially supported by the NSF Grant DMS-0931948.

## References

- [1] L. Horváth and P. Kokoszka, *Inference for Functional Data with Applications*, Springer, New York, NY, USA, 2012.
- [2] J. O. Ramsay and B. W. Silverman, *Functional Data Analysis*, Springer, New York, NY, USA, 2005.
- [3] J. O. Ramsay and B. W. Silverman, *Applied Functional Data Analysis*, Springer, New York, NY, USA, 2002.
- [4] J. Ramsay, G. Hooker, and S. Graves, *Functional Data Analysis with R and MATLAB*, Springer, New York, NY, USA, 2009.
- [5] F. Ferraty and P. Vieu, *Nonparametric Functional Data Analysis: Theory and Practice*, Springer, New York, NY, USA, 2006.
- [6] F. Ferraty and Y. Romain, Eds., *The Oxford Handbook of Functional Data Analysis*, Oxford University Press, New York, NY, USA, 2011.
- [7] F. Ferraty, Ed., *Recent Advances in Functional Data Analysis and Related Topics*, Physica, 2011.
- [8] J. Q. Shi and T. Choi, *Gaussian Process Regression Analysis for Functional Data*, CRC Press, New York, NY, USA, 2011.
- [9] D. Bosq, *Linear Processes in Function Spaces*, Springer, New York, NY, USA, 2000.
- [10] D. Bosq and D. Blanke, *Inference and Prediction in Large Dimensions*, John Wiley & Sons, New York, NY, USA, 2007.
- [11] F. Riesz and B. S. Nagy, *Functional Analysis*, Dover, New York, NY, USA, 1990.
- [12] N. I. Akhiezer and I. M. Glazman, *Theory of Linear Operators in Hilbert Space*, Dover, New York, NY, USA, 1993.
- [13] L. Debnath and P. Mikusinski, *Introduction to Hilbert Spaces with Applications*, Elsevier, New York, NY, USA, 2005.
- [14] A. Araujo and E. Giné, *The Central Limit Theorem for Real and Banach Valued Random Variables*, John Wiley & Sons, New York, NY, USA, 1980.
- [15] W. Linde, *Probability in Banach Spaces—Stable and Infinitely Divisible Distributions*, John Wiley & Sons, New York, NY, USA, 1986.
- [16] N. N. Vakhaniia, V. I. Tarieladze, and S. A. Chobanian, *Probability Distributions on Banach Spaces*, Springer, New York, NY, USA, 1987.
- [17] F. Yao, H. G. Müller, and J. L. Wang, "Functional data analysis for sparse longitudinal data," *Journal of the American Statistical Association*, vol. 100, no. 470, pp. 577–590, 2005.
- [18] H. G. Müller, "Functional modeling of longitudinal data," in *Longitudinal Data Analysis*, G. Fitzmaurice, M. Davidian, G. Verbeke, and G. Molenberghs, Eds., pp. 223–252, John Wiley & Sons, New York, NY, USA, 2009.
- [19] P. Hall, H. G. Müller, and F. Yao, "Modelling sparse generalized longitudinal observations with latent Gaussian processes," *Journal of the Royal Statistical Society. Series B*, vol. 70, no. 4, pp. 703–723, 2008.
- [20] H. G. Müller and U. Stadtmüller, "Generalized functional linear models," *Annals of Statistics*, vol. 33, no. 2, pp. 774–805, 2005.
- [21] H. G. Müller, U. Stadtmüller, and F. Yao, "Functional variance processes," *Journal of the American Statistical Association*, vol. 101, no. 475, pp. 1007–1018, 2006.
- [22] F. Yao and H. G. Müller, "Functional quadratic regression," *Biometrika*, vol. 97, no. 1, pp. 49–64, 2010.
- [23] C. Z. Di and C. M. Crainiceanu, "Multilevel sparse functional principal component analysis," Tech. Rep., Johns Hopkins University, 2010.
- [24] D. B. Clarkson, C. Fraley, C. Gu, and J. O. Ramsay, *S+Functional Data Analysis*, Springer, New York, NY, USA, 2005.
- [25] P. T. Reiss and R. T. Ogden, "Functional generalized linear models with images as predictors," *Biometrics*, vol. 66, no. 1, pp. 61–69, 2010.
- [26] V. Zippunikov, B. Caffo, D. M. Yousem, C. Davatzikos, B. S. Schwartz, and C. Crainiceanu, "Functional principal component model for high-dimensional brain imaging," *Neuroimage*, vol. 58, no. 3, pp. 772–784, 2011.

- [27] S. Guillas and M. J. Lai, "Bivariate splines for spatial functional regression models," *Journal of Nonparametric Statistics*, vol. 22, no. 4, pp. 477–497, 2010.
- [28] J. S. Morris, V. Baladandayuthapani, R. C. Herrick, P. Sanna, and H. Gutstein, "Automated analysis of quantitative image data using isomorphic functional mixed models, with application to proteomics data," *The Annals of Applied Statistics*, vol. 5, no. 2, pp. 894–923, 2012.
- [29] B. Aydin, G. Pataki, H. Wang, A. Ladha, E. Bullitt, and J. S. Marron, "Visualizing the structure of large trees," *Electronic Journal of Statistics*, vol. 5, pp. 405–420, 2011.
- [30] K. Kenobi, I. L. Dryden, and H. Le, "Shape curves and geodesic modelling," *Biometrika*, vol. 97, no. 3, pp. 567–584, 2010.
- [31] M. Crane and V. Patrangenaru, "Random change on a Lie group and mean glaucomatous projective shape change detection from stereo pair images," *Journal of Multivariate Analysis*, vol. 102, no. 2, pp. 225–237, 2011.
- [32] Y. Li and T. Hsing, "On rates of convergence in functional linear regression," *Journal of Multivariate Analysis*, vol. 98, no. 9, pp. 1782–1804, 2007.
- [33] F. Abramovich and C. Angelini, "Testing in mixed-effects FANOVA models," *Journal of Statistical Planning and Inference*, vol. 136, no. 12, pp. 4326–4348, 2006.
- [34] M. Pensky and T. Sapatinas, "Functional deconvolution in a periodic setting: uniform case," *Annals of Statistics*, vol. 37, no. 1, pp. 73–104, 2009.
- [35] A. Antoniadis, E. Paparoditis, and T. Sapatinas, "A functional wavelet-kernel approach for time series prediction," *Journal of the Royal Statistical Society. Series B*, vol. 68, no. 5, pp. 837–857, 2006.
- [36] D. Liebl, "Modeling hourly electricity spot market prices as nonstationary functional times series," Tech. Rep., University of Cologne, 2012.
- [37] V. Kargin and A. Onatski, "Curve forecasting by functional autoregression," *Journal of Multivariate Analysis*, vol. 99, no. 10, pp. 2508–2526, 2008.
- [38] N. Bathia, Q. Yao, and F. Ziegelmann, "Identifying the finite dimensionality of curve time series," *Annals of Statistics*, vol. 38, no. 6, pp. 3352–3386, 2010.
- [39] P. J. Brockwell and R. A. Davis, *Introduction to Time Series and Forecasting*, Springer, New York, NY, USA, 2nd edition, 2002.
- [40] R. H. Shumway and D. S. Stoffer, *Time Series Analysis and Its Applications with R Examples*, Springer, New York, NY, USA, 2011.
- [41] P. J. Brockwell and R. A. Davis, *Time Series: Theory and Methods*, Springer, New York, NY, USA, 1991.
- [42] T. W. Anderson, *The Statistical Analysis of Time Series*, John Wiley & Sons, New York, NY, USA, 1994.
- [43] D. Didericksen, P. Kokoszka, and X. Zhang, "Empirical properties of forecasts with the functional autoregressive model," *Computational Statistics*, vol. 27, no. 2, pp. 285–298, 2012.
- [44] P. Kokoszka and M. Reimherr, "Determining the order of the functional autoregressive model," *Journal of Time Series Analysis*. In press.
- [45] A. Aue, R. Gabrys, L. Horváth, and P. Kokoszka, "Estimation of a change-point in the mean function of functional data," *Journal of Multivariate Analysis*, vol. 100, no. 10, pp. 2254–2269, 2009.
- [46] A. Aue, S. Hörmann, L. Horváth, and M. Reimherr, "Break detection in the covariance structure of multivariate time series models," *Annals of Statistics*, vol. 37, no. 6, pp. 4046–4087, 2009.
- [47] I. Berkes, S. Hörmann, and J. Schauer, "Split invariance principles for stationary processes," *The Annals of Probability*, vol. 39, pp. 2441–2473, 2011.
- [48] S. Hörmann and P. Kokoszka, "Weakly dependent functional data," *The Annals of Statistics*, vol. 38, no. 3, pp. 1845–1884, 2010.
- [49] P. Doukhan, *Mixing: Properties and Examples*, Lecture Notes in Statistics, Springer, New York, NY, USA, 1994.
- [50] R. C. Bradley, *Introduction to Strong Mixing Conditions*, vol. 1–3, Kendrick Press, 2007.
- [51] P. Doukhan and S. Louhichi, "A new weak dependence condition and applications to moment inequalities," *Stochastic Processes and Their Applications*, vol. 84, no. 2, pp. 313–343, 1999.
- [52] W. B. Wu, "Nonlinear system theory: another look at dependence," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 40, pp. 14150–14154, 2005.
- [53] W. B. Wu, "Strong invariance principles for dependent random variables," *Annals of Probability*, vol. 35, no. 6, pp. 2294–2320, 2007.
- [54] H. Lian, "Convergence of functional responses," Tech. Rep., Nanyang Technological University, Singapore, 2012.

- [55] P. Kokoszka and M. Reimherr, "Asymptotic normality of the principal components of functional time series," Tech. Rep., University of Chicago, 2012.
- [56] A. Mas, "Weak convergence for the covariance operators of a Hilbertian linear process," *Stochastic Processes and Their Applications*, vol. 99, no. 1, pp. 117–135, 2002.
- [57] L. Horváth, P. Kokoszka, and R. Reeder, "Estimation of the mean of functional time series and a two sample problem," *Journal of the Royal Statistical Society. Series B*. In press.
- [58] D. W. K. Andrews, "Heteroskedasticity and autocorrelation consistent covariance matrix estimation," *Econometrica*, vol. 59, no. 3, pp. 817–858, 1991.
- [59] X. Shao and X. Zhang, "Testing for change points in time series," *Journal of the American Statistical Association*, vol. 105, no. 491, pp. 1228–1240, 2010.
- [60] P. Bühlmann and H. R. Künsch, "Block length selection in the bootstrap for time series," *Computational Statistics and Data Analysis*, vol. 31, no. 3, pp. 295–310, 1999.
- [61] D. N. Politis, J. P. Romano, and M. Wolf, *Subsampling*, Springer, New York, NY, USA, 1999.
- [62] P. J. Bickel and A. Sakov, "On the choice of  $m$  in the  $m$  out of  $n$  bootstrap and confidence bounds for extrema," *Statistica Sinica*, vol. 18, no. 3, pp. 967–985, 2008.
- [63] X. Shao and D. Politis, "Fixed- $b$  subsampling and block bootstrap: improved confidence sets based on  $p$ -value calibration," *Journal of the Royal Statistical Society. Series B*. In press.
- [64] X. Shao, "A self-normalized approach to confidence interval construction in time series," *Journal of the Royal Statistical Society. Series B*, vol. 72, no. 3, pp. 343–366, 2010.
- [65] X. Shao, "Corrigendum: a self-normalized approach to confidence interval construction in time series," *Journal of the Royal Statistical Society. Series B*, vol. 72, no. 5, pp. 695–696, 2010.
- [66] X. Zhang, X. Shao, K. Hayhoe, and D. Wuebbles, "Testing the structural stability of temporally dependent functional observations and application to climate projections," *Electronic Journal of Statistics*, vol. 5, pp. 1765–1796, 2011.
- [67] M. Benko, W. Härdle, and A. Kneip, "Common functional principal components," *Annals of Statistics*, vol. 37, no. 1, pp. 1–34, 2009.
- [68] A. Laukaitis and A. Račkauskas, "Functional data analysis for clients segmentation tasks," *European Journal of Operational Research*, vol. 163, no. 1, pp. 210–216, 2005.
- [69] A. Cuevas, M. Febrero, and R. Fraiman, "An anova test for functional data," *Computational Statistics and Data Analysis*, vol. 47, no. 1, pp. 111–122, 2004.
- [70] P. Delicado, "Functional  $k$ -sample problem when data are density functions," *Computational Statistics*, vol. 22, no. 3, pp. 391–410, 2007.
- [71] F. Ferraty, P. Vieu, and S. Viguier-Pla, "Factor-based comparison of groups of curves," *Computational Statistics and Data Analysis*, vol. 51, no. 10, pp. 4903–4910, 2007.
- [72] V. M. Panaretos, D. Kraus, and J. H. Maddocks, "Second-order comparison of gaussian random functions and the geometry of DNA minicircles," *Journal of the American Statistical Association*, vol. 105, no. 490, pp. 670–682, 2010.
- [73] S. Fremdt, L. Horváth, P. Kokoszka, and J. Steinebach, "Testing the equality of covariance operators in functional samples," *Scandinavian Journal of Statistics*. In press.
- [74] D. Kraus and V. M. Panaretos, "Dispersion operators and resistant second-order analysis of functional data," *Biometrika*. In press.
- [75] L. Horváth, P. Kokoszka, and M. Reimherr, "Two sample inference in functional linear models," *Canadian Journal of Statistics*, vol. 37, no. 4, pp. 571–591, 2009.
- [76] B. E. Brodsky and B. S. Darkhovsky, *Nonparametric Methods in Change-Point Problems*, Kluwer, 1993.
- [77] M. Basseville and I. V. Nikifirov, *Detection of Abrupt Changes: Theory and Applications*, Prentice Hall, Upper Saddle River, NJ, USA, 1993.
- [78] M. Csörgő and L. Horváth, *Limit Theorems in Change-Point Analysis*, John Wiley & Sons, New York, NY, USA, 1997.
- [79] J. Chen and A. K. Gupta, *Parametric Statistical Change Point Analysis: With Applications to Genetics, Medicine, and Finance*, Birkhäuser, Boston, Mass, USA, 2011.
- [80] M. Basseville, I. V. Nikifirov, and A. Tartakovsky, *Sequential Analysis: Hypothesis Testing and Change-Point Detection*, Chapman & Hall/CRC, 2012.
- [81] I. Berkes, R. Gabrys, L. Horváth, and P. Kokoszka, "Detecting changes in the mean of functional observations," *Journal of the Royal Statistical Society. Series B*, vol. 71, no. 5, pp. 927–946, 2009.
- [82] L. Horváth, M. Hušková, and P. Kokoszka, "Testing the stability of the functional autoregressive process," *Journal of Multivariate Analysis*, vol. 101, no. 2, pp. 352–367, 2010.
- [83] J. A. D. Aston and C. Kirch, "Estimation of the distribution of change-points with application to fMRI data," *The Annals of Applied Statistics*. In press.

- [84] A. Račkauskas and C. Suquet, "Testing epidemic changes of infinite dimensional parameters," *Statistical Inference for Stochastic Processes*, vol. 9, no. 2, pp. 111–134, 2006.
- [85] S. Hörmann, L. Horváth, and R. Reeder, "A functional version of the ARCH model," *Econometric Theory*. In press.
- [86] P. Kokoszka and M. Reimherr, "Predictability of shapes of intra day price curves," Tech. Rep., University of Chicago, 2012.
- [87] P. Kokoszka, H. Miao, and X. Zhang, "A factor model for intraday cumulative returns," Tech. Rep., Colorado State University, 2012.
- [88] P. Kokoszka and X. Zhang, "Functional prediction of intraday cumulative returns," *Statistical Modelling*, vol. 12, no. 4, pp. 377–398, 2012.
- [89] S. Hays, H. Shen, and J. Z. Huang, "Functional dynamic factor models with application to yield curve forecasting," *The Annals of Applied Statistics*. In press.
- [90] R. Gabrys, L. Horváth, and P. Kokoszka, "Tests for error correlation in the functional linear model," *Journal of the American Statistical Association*, vol. 105, no. 491, pp. 1113–1125, 2010.
- [91] K. Benhenni, S. Hedli-Griche, and M. Rachdi, "Estimation of the regression operator from functional fixed-design with correlated errors," *Journal of Multivariate Analysis*, vol. 101, no. 2, pp. 476–490, 2010.
- [92] H. Battey and A. Sancetta, "Conditional estimation for dependent functional data," Tech. Rep., University of Bristol, 2012.
- [93] O. Schabenberger and C. A. Gotway, *Statistical Methods for Spatial Data Analysis*, Chapman & Hall/CRC, 2005.
- [94] A. E. Gelfand, P. J. Diggle, M. Fuentes, and P. Guttorp, Eds., *Handbook of Spatial Statistics*, CRC Press, New York, NY, USA, 2010.
- [95] M. Sherman, *Spatial Statistics and Spatio-Temporal Data: Covariance Functions and Directional Properties*, John Wiley & Sons, New York, NY, USA, 2011.
- [96] P. Delicado, R. Giraldo, C. Comas, and J. Mateu, "Statistics for spatial functional data: some recent contributions," *Environmetrics*, vol. 21, no. 3-4, pp. 224–239, 2010.
- [97] O. Gromenko, P. Kokoszka, L. Zhu, and J. Sojka, "Estimation and testing for spatially indexed curves with application to ionospheric and magnetic field trends," *The Annals of Applied Statistics*, vol. 6, no. 2, pp. 669–696, 2012.
- [98] S. Hörmann and P. Kokoszka, "Consistency of the mean and the principal components of spatially distributed functional data," *Bernoulli*. In press.
- [99] B. U. Park, T. Y. Kim, J. S. Park, and S. Y. Hwang, "Practically applicable central limit theorem for spatial statistics," *Mathematical Geosciences*, vol. 41, no. 5, pp. 555–569, 2009.
- [100] O. Gromenko and P. Kokoszka, "Testing the equality of mean functions of spatially distributed curves," *Journal of the Royal Statistical Society. Series C*. In press.
- [101] O. Gromenko and P. Kokoszka, "Nonparametric estimation in small data sets of spatially indexed curves with application to temporal trend determination," Tech. Rep., Utah State University, 2012.
- [102] P. Hall, N. I. Fisher, and B. Hoffmann, "On the nonparametric estimation of covariance functions," *The Annals of Statistics*, vol. 22, no. 4, pp. 2115–2134, 1994.
- [103] P. Hall and P. Patil, "Properties of nonparametric estimators of autocovariance for stationary random fields," *Probability Theory and Related Fields*, vol. 99, no. 3, pp. 399–424, 1994.
- [104] D. Nerini, P. Monestiez, and C. Manté, "Cokriging for spatial functional data," *Journal of Multivariate Analysis*, vol. 101, no. 2, pp. 409–418, 2010.
- [105] R. Giraldo, P. Delicado, and J. Mateu, "Ordinary kriging for function-valued spatial data," *Environmental and Ecological Statistics*, vol. 18, no. 3, pp. 411–426, 2011.
- [106] M. L. Stein, *Interpolation of Spatial Data: Some Theory for Krigging*, Springer, New York, NY, USA, 1999.
- [107] H. Wackernagel, *Multivariate Geostatistics*, Springer, New York, NY, USA, 2003.
- [108] R. Giraldo, P. Delicado, and J. Mateu, "Hierarchical clustering of spatially correlated functional data," *Statistica Neerlandica*. In press.
- [109] W. K. Härdle and M. Osipenko, "Spatial risk premium on weather derivatives and hedging weather exposure in electricity," *Energy Journal*, vol. 33, no. 2, pp. 149–170, 2012.
- [110] A. M. Staicu, C. M. Crainiceanu, and R. J. Carroll, "Fast methods for spatially correlated multilevel functional data," *Biostatistics*, vol. 11, no. 2, pp. 177–194, 2010.
- [111] U. Grenander and M. Rosenblatt, *Statistical Analysis of Stationary Time Series*, John Wiley & Sons, New York, NY, USA, 1957.
- [112] C. W. J. Granger and M. Hatanaka, *Spectral Analysis of Economic Time Series*, Princeton University Press, 1964.

- [113] M. B. Priestley, Rao, and T. Subba, "A test for non-stationarity of time-series," *Journal of the Royal Statistical Society. Series B*, vol. 31, no. 1, pp. 140–149, 1969.
- [114] Y. Dwivedi and S. Subba Rao, "A test for second-order stationarity of a time series based on the discrete Fourier transform," *Journal of Time Series Analysis*, vol. 32, no. 1, pp. 68–91, 2011.
- [115] V. M. Panaretos and S. Tavakoli, "Fourier analysis of stationary time series in function space," Tech. Rep., EPFL, 2012.
- [116] J. B. Odei, M. B. Hooten, and J. Jin, "Hierarchical spatio-temporal models for intermountain snow water storage," in *JSM Proceedings, Section on Statistics and the Environment*, pp. 870–878, American Statistical Association, 2009.
- [117] C. Timmermans and R. von Sachs, "BAGIDIS, a new method for statistical analysis of differences between curves with sharp discontinuities," Tech. Rep., UC Louvain, 2010.
- [118] D. Draghicescu and R. Ignaccolo, "Modeling threshold exceedance probabilities of spatially correlated time series," *Electronic Journal of Statistics*, vol. 3, pp. 149–164, 2009.