ANNUAL REVIEWS

Second-Generation Functional Data

Salil Koner[1] and Ana-Maria Staicu[2]

[1]Department of Biostatistics, Duke University, Durham, North Carolina, USA;
email: Salil.Koner@duke.edu

[2]Department of Statistics, North Carolina State University, Raleigh, North Carolina, USA;
email: astaicu@ncsu.edu

ANNUAL REVIEWS CONNECT

www.annualreviews.org

- Download figures
- Navigate cited references
- Keyword search
- Explore related articles
- Share via email or social media

OPEN ACCESS

Keywords

functional principal component analysis, spatial functional data, longitudinal functional data, functional time series, multivariate functional data

Abstract

Modern studies from a variety of fields record multiple functional observations according to either multivariate, longitudinal, spatial, or time series designs. We refer to such data as second-generation functional data because their analysis—unlike typical functional data analysis, which assumes independence of the functions—accounts for the complex dependence between the functional observations and requires more advanced methods. In this article, we provide an overview of the techniques for analyzing second-generation functional data with a focus on highlighting the key methodological intricacies that stem from the need for modeling complex dependence, compared with independent functional data. For each of the four types of second-generation functional data presented—multivariate functional data, longitudinal functional data, functional time series and spatially functional data—we discuss how the widely popular functional principal component analysis can be extended to these settings to define, identify main directions of variation, and describe dependence among the functions. In addition to modeling, we also discuss prediction, statistical inference, and application to clustering. We close by discussing future directions in this area.

1. INTRODUCTION

Classical functional data analysis (FDA) studies independent random functions, which are sometimes referred to as first-generation functional data (Wang et al. 2016). This area has seen an explosion of development over the past few decades, and numerous journal publications, books, and review articles have been written on this topic. In particular, Morris (2015) discussed regression analysis when the response and/or the predictors are functional observations, and Wang et al. (2016) provided an overview of the common techniques for modeling, prediction, clustering, and classification.

Modern scientific applications now routinely collect functional data, but as a basic measurement in a longitudinal, multivariate, spatial, time series, or multilevel design. For example, in an animal study of osteoarthritis (OA), minute-by-minute daily physical activity of indoor cats was recorded for many days during a 20-day study period (Koner et al. 2022). **Figure 1** shows the daily activity profiles, obtained by connecting the activity measurements (cumulative, log-scale) recorded in a day, for three of the cats and highlights the activity profiles observed on the 4th, 8th, 12th, and 16th days since the beginning of the study. This is an example of functional data recorded in a longitudinal design, because each cat (subject) is monitored repeatedly over many days, and at each day, a curve is measured. This example is further discussed below and in the **Supplemental Appendix**. Other examples include medical studies, such as the case of multiple five-years profiles of Parkinson's disease markers recorded for many patients, and a plethora of emerging studies where a continuous data collection regime is present, such as electronic health records–based studies such as the National Health and Nutrition Examination Survey; manufacturing systems, neuroimaging, and environmental applications; growth studies; and traffic monitoring systems, to list a few. We refer to these data as second-generation functional data because not only are functional data recorded at a second layer (multivariate, longitudinal design, and so on), but also, their analysis needs to account both for the complex dependence between the functional observations
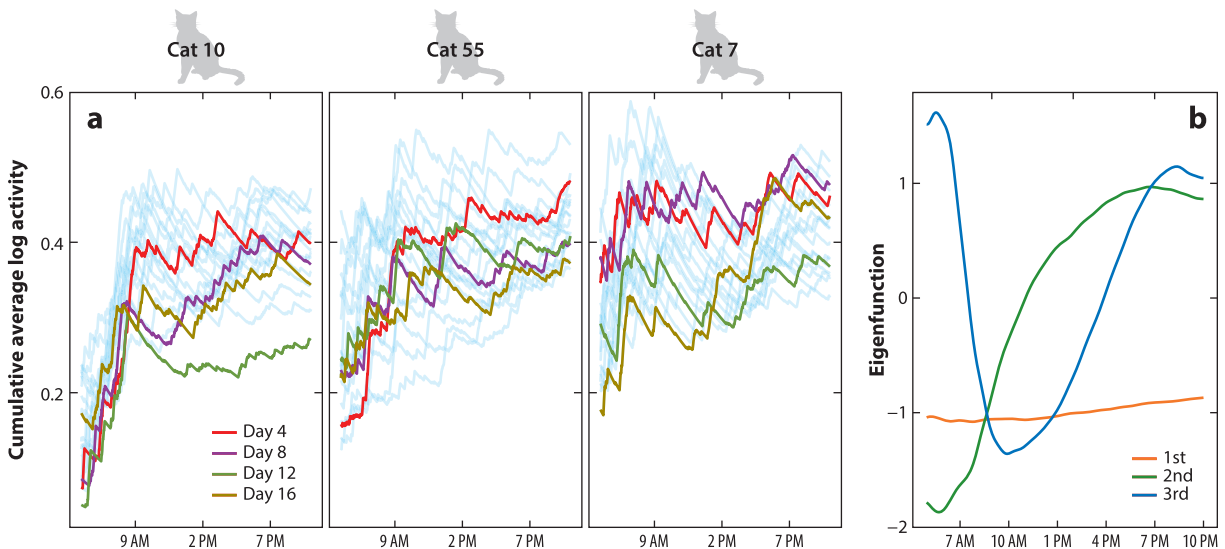
Figure 1

(*a*) Spaghetti plot of the cumulative average daily activity between 5 AM and 10 PM for three indoor cats observed over a 20-day study period (*light sky blue*). Highlighted in different colors are their activity profiles for four days since the start of the period. (*b*) Top leading eigenfunctions using the model in Equation 3 and setting the percentage of variation explained to 95%. Data from **https://repository.lib.ncsu.edu/handle/1840.20/39047**.

and for the specific design, while keeping the computations feasible. Analyzing these data requires more advanced methods.

This review considers second-generation functional data acquired in a multivariate, longitudinal, time series, or spatial design. For each setting in turn, we describe the general statistical framework and review the methodologies for estimation, prediction, and inference, with a focus on the interplay between the FDA techniques and the ones from classical statistics for multivariate, longitudinal, time series, and spatial data. The primary focus is modeling and methods and rather than theory and computation, although we do highlight existing R packages that implement the reviewed approaches. While the multivariate and longitudinal parts equally review approaches adequate for functional data obtained from sparse and dense grids of points, the sections on time series and spatial designs concern functional data observed at fine grids, where the underlying trajectories can be recovered with negligible error.

Throughout the article, we assume that the real-valued latent functions are in a Hilbert space, such as $L^2(\mathcal{D})$, that is equipped with the inner product $\langle f, g \rangle = \int_{u \in \mathcal{D}} f(u)g(u)\mathrm{d}u$. Although research exists on functional data in Banach space and other non-Euclidean spaces, we do not discuss it in this review. We based the review on a selection of articles in this fast-moving research area, limited by journal space constraints. As some of the topics have been researched for nearly 30 years, and all of them are very active areas of current research, we apologize for any omissions of relevant work.

2. MULTIVARIATE FUNCTIONAL DATA ANALYSIS

2.1. Statistical Framework

Multivariate functional data (MVFD) are one of the most commonly observed types of next-generation functional data, where the trajectories observed for each subject are multidimensional. Take, for example, the growth study of infants described by (Han et al. 2018): Each infant is followed for a set period of time after birth, and three physical outcomes—head circumference, body weight, and body length—are measured at different ages. More generally, consider that each subject i is observed at many time points in a time interval $\mathcal{D} \subset \mathbb{R}$, and at each time point, denoted generically by d_{ij}, q outcomes $Y_{ij}^1, \ldots, Y_{ij}^q$ are recorded. We assume that the m_i observations for each outcome, say $\{Y_{i1}^\ell, \ldots, Y_{im_i}^\ell\}$, are noisy realizations of a smooth latent curve, $X_i^\ell(\cdot)$, defined on $\mathcal{D}$ and evaluated at a finite grid $\{d_{i1}, \ldots, d_{im_i}\}$, like $Y_{ij}^\ell = X_i^\ell(d_{ij}) + \epsilon_{ij}^\ell$ for measurement error ϵ_{ij}^ℓ, where $\ell = 1, \ldots, q$. The time points d_{ij} at which the outcomes Y_{ij}^ℓ are measured are assumed the same across ℓ; later, we discuss the case when they are outcome specific, introducing the notation d_{ij}^ℓ. Denote by the bold symbol $\mathbf{X}_i(d) = (X_i^1(d), \ldots, X_i^q(d))^\top$ the vector of q latent functions, and assume that $\mathbf{X}_i(\cdot)$s are independent and identically distributed (i.i.d.) copies of a smooth process $\mathbf{X}(\cdot)$, which lives in $\mathcal{L}_q^2(\mathcal{D})$—the space of square-integrable q-dimensional functions on $\mathcal{D}$, associated with the inner product $\langle \mathbf{f}, \mathbf{g} \rangle = \sum_{\ell=1}^q \langle f^\ell, g^\ell \rangle$. The measurement errors are assumed independent, with $\epsilon_{ij}^\ell \sim (0, \tau_\ell^2)$ for all i, j, and ℓ; $(0, \tau_\ell^2)$ stands for mean zero and variance τ_ℓ^2 distribution. Denote the mean of $\mathbf{X}(d)$ by $\boldsymbol{\mu}(d)$, with $\mu^\ell(\cdot) = \mathbb{E}[X^\ell(\cdot)]$, $\ell = 1, \ldots, q$, and the covariance operator by $\boldsymbol{\Xi}$, where $(\boldsymbol{\Xi}\,\mathbf{f})(d)$ is a $q \times 1$ vector with elements $(\boldsymbol{\Xi}\,\mathbf{f})^\ell(d) = \sum_{\ell'=1}^q \langle \Sigma_{\ell\ell'}(d, \cdot), f^{\ell'} \rangle$, where $\Sigma_{\ell\ell'}(d, d') = \mathrm{cov}\{X^\ell(d), X^{\ell'}(d')\}$ for $1 \leq \ell \leq q$. It is assumed that the means $\mu^\ell(\cdot)$ and the covariances $\Sigma_{\ell\ell'}(\cdot, \cdot)$ are smooth functions. Typical goals for MVFD include (a) characterizing dependence between the components of $\mathbf{X}(d)$, (b) visualizing and detecting outliers, (c) developing models that account for additional covariates, and (d) detecting clustering.

2.2. Modeling the Multivariate Functional Dependence

The first methods developed for the analysis of MVFD are based on decomposing each latent function using predefined bases. Guo (2004) used smoothing splines, and Morris & Carroll (2006)

considered discrete wavelet transform and modeled the coefficients with a Bayesian functional mixed model (FMM) framework. This is particularly useful for data exhibiting local features. A detailed review of these functional regression models is provided by Morris (2015).

2.2.1. Multivariate Karhunen-Loéve expansion. One of the most widely used techniques in FDA is functional principal component analysis (FPCA), which allows us to represent i.i.d. curves using a few leading eigenfunctions. A direct application of FPCA to each component $X_i^\ell(\cdot)$ leads to the Karhunen-Loéve (KL) representation as $X_i^\ell(d) = \mu^\ell(d) + \sum_{k=1}^\infty \xi_{ik}^\ell \phi_k^\ell(d)$, where $\{\phi_k^\ell(\cdot), \lambda_k^\ell\}_{k\geq 1}$ are the eigenelements of $\Sigma_{\ell\ell}(d, d')$, with $\lambda_1^\ell \geq \lambda_2^\ell \geq \cdots \geq 0$ and orthogonal functions $\phi_1^\ell(\cdot), \phi_2^\ell(\cdot), \ldots$. Here ξ_{ik}^ℓ are functional principal component (FPC) scores $\xi_{ik}^\ell = \langle X_i^\ell - \mu^\ell, \phi_k^\ell \rangle$, such that $\xi_{ik}^\ell \sim (0, \lambda_k^\ell)$ and ξ_{ik}^ℓ are uncorrelated across k. A naïve FPCA approach is one that ignores the correlation of FPCs across components and assumes that ξ_{ik}^ℓ are also uncorrelated over ℓ.

Multivariate FPCA (MVFPCA) captures the joint variation among the vector function components. Since the covariance operator $\boldsymbol{\Xi}$ is positive, compact, and self-adjoint in $\mathcal{D}$, there exists a set of orthonormal basis functions $\{\boldsymbol{\psi}_k(\cdot)\}_{k\geq 1}$, with $\langle \boldsymbol{\psi}_k, \boldsymbol{\psi}_{k'} \rangle = \mathbb{I}(k = k')$ such that $\boldsymbol{\Xi}\, \boldsymbol{\psi}_k = \theta_k \boldsymbol{\psi}_k$, where $\theta_1 \geq \theta_2 \cdots \geq 0$; $\mathbb{I}(k = k')$ is the common indicator function. Then $\mathbf{X}_i(\cdot)$ can be decomposed using the multivariate KL expansion

$$\mathbf{X}_i(d) = \boldsymbol{\mu}(d) + \sum_{r=1}^\infty \zeta_{ir} \boldsymbol{\psi}_r(d), \qquad i = 1, \ldots, n,$$

where $\zeta_{ik} = \langle \mathbf{X}_i - \boldsymbol{\mu}, \boldsymbol{\psi}_k \rangle$ are scalar multivariate FPC (MVFPC) scores, and $\zeta_{ik} \sim (0, \theta_k)$ are uncorrelated across k and independent over i. The convergence holds uniformly in d with respect to the norm induced by $\langle \cdot, \cdot \rangle$. The vector $\boldsymbol{\zeta}_i$ of the leading FPC scores extracts the main features of $\mathbf{X}_i(\cdot)$. It can be shown that the MVFPC components are a weighted average of the counterparts obtained from the naïve application of univariate FPCA, where the weights are a function of the covariance between $(\xi_{ik}^\ell, \xi_{ik'}^{\ell'})$ for $\ell \neq \ell'$ and $k \neq k'$.

When the time domain is component specific, $\mathcal{D}^\ell$ (a setting called heterogeneous domain) and thus the observed time points d_{ij}^ℓ are also component specific, Happ & Greven (2018) conceptualized the FPCA for MVFD by establishing the theoretical basis of the multivariate KL theorem. A detailed description of the notation, set-up, and theoretical framework required by a heterogeneous domain is provided in the **Supplemental Appendix**.

2.2.2. Adjusting for uneven variability among components. The eigenfunctions $\boldsymbol{\psi}_k(\cdot)$ in the MVFPCA discussed above maximize the total variance defined as $\sum_{\ell=1}^q \int \mathrm{var}\{X^\ell(u)\}\mathrm{d}u$, which equals the trace of $\boldsymbol{\Xi}$, and moreover, $\sum_{k=1}^\infty \theta_k$. Thus, MVFPCA is meaningful when all the component functions of the response are measured on the same scale, or when they have a similar range of variability. When the components are observed over different scales, normalization is required to handle the heteroscedasticity of their variance. Normalized MVFPCA first scales the functions pointwise, $\mathbf{X}_w(d) = \mathbf{W}(d)^{-1}\{\mathbf{X}(d) - \boldsymbol{\mu}(d)\}$, and then represents the weighted functions using multivariate KL as $\mathbf{X}_w(d) = \sum_{k=1}^\infty \zeta_{w,k} \boldsymbol{\psi}_{w,k}(d)$, where $\langle \boldsymbol{\psi}_{w,k}, \boldsymbol{\psi}_{w,k'} \rangle = \mathbb{I}(k = k')$ and $\zeta_{w,k} = \langle \mathbf{X}_w, \boldsymbol{\psi}_{w,k'} \rangle$. Jacques & Preda (2014) used $\mathbf{W}(d) = \boldsymbol{\Sigma}(d, d)^{1/2}$, the square root of pointwise $q \times q$ covariance, and Chiou et al. (2014) considered $\mathbf{W}(d) = \mathrm{diag}\{\Sigma_{11}(d, d), \ldots, \Sigma_{qq}(d, d)\}^{1/2}$. However, pointwise normalization fails to account for different sizes of the domains of heterogeneous MVFD, and it downscales the regions with stronger variation. An alternative is to redefine the inner product as $\langle \mathbf{f}, \mathbf{g} \rangle^w = \sum_{\ell=1}^q w_\ell \langle f^\ell, g^\ell \rangle$, where the weights w_ℓ reflect the adjustment made for the ℓth component; the new inner product is equivalent to using $\mathbf{X}_w(d)$ with $\mathbf{W}(d) = \mathrm{diag}(w_1^{1/2}, \ldots, w_q^{1/2})$ and the norm induced by $\langle \cdot, \cdot \rangle$. Happ & Greven (2018) consider $w_\ell = \int \Sigma_{\ell\ell}(u, u)\mathrm{d}u$.

2.2.3. Alternative additive models for multivariate functional data.

Another approach to model MVFD is by describing univariate models for each individual component in a manner that captures the dependence among the components. Chiou & Müller (2016) introduced the so-called pairwise interaction model, which represents each functional component of the vector response as the sum of independent processes that quantify the interaction between the component and all the other components. Specifically, for each $\ell = 1, \ldots, q$, it is posited $X^\ell(d) = \mu^\ell(d) + \sum_{\ell \neq \ell'} Z^{\ell\ell'}(d) + V^\ell(d)$, where $Z^{\ell\ell'}(d) = Z^{\ell'\ell}(d)$ are mean zero pairwise interaction processes with covariance $G_{\ell\ell'}(d, d')$ and $V^\ell(d)$ is a mean zero residual process with variance $H_\ell(d, d')$ that describes the remaining variance of $X^\ell(d)$; all processes are assumed independent. The identifiability of this model is ensured by setting $G_{\ell\ell'}(d, d') = \Sigma_{\ell\ell'}(d, d')$ for all $\ell \neq \ell'$ and $H_\ell(d, d') = \Sigma_{\ell\ell}(d, d') - \sum_{\ell' \neq \ell}^q G_{\ell\ell'}(d, d')$. One appeal of this approach is that the ratio $\rho_{\ell\ell'}(d) = G_{\ell\ell'}(d, d)/\Sigma_{\ell\ell}(d, d)$ can be interpreted as a numerical measure of percentage of variation explained (PVE) by the interaction between $X^\ell(d)$ and $X^{\ell'}(d)$. However, $\rho_{\ell\ell'}$ should not be interpreted as a measure of correlation since it is not symmetric.

2.2.4. Cross-component dependence-based alternatives.

Earlier approaches to develop FPCA for MVFD include pointwise application of the classical PCA to the q-dimensional vector $\mathbf{X}(d)$ for each $d \in \mathcal{D}$. This implies finding directions $\mathbf{a}_k(d) = (a_k^1(d), \ldots, a_k^q(d))^\top$ such that $\int \mathbf{a}_k^\top(u)\Sigma(u, u)\mathbf{a}_k(u)\mathrm{d}u$ is maximized subject to $\|\mathbf{a}_k(d)\|_2 = 1$ for all $d \in \mathcal{D}$ and $\mathbf{a}_k^\top(d)\mathbf{a}_{k'}(d) = 0$ for all $k \neq k'$, $k = 1, \ldots, q$. The solutions $\{\mathbf{a}_k(d)\}_{k=1}^q$ for the above constrained maximization are the eigenvectors of $\Sigma(d, d)$ associated with the ordered eigenvalues $\{\lambda_k(d)\}_{k=1}^q$. Berrendero et al. (2011) considered pointwise PCA separately, for every d, and showed that smoothness of $\mathbf{a}_k(d)$ requires absolute continuity of the covariance function. The optimal number of principal components also varies across d, and to bypass this challenge different criteria for choosing $K(d)$, uniformly over d, have been proposed. One choice is to use the minimum value such that the percentage of integrated variation $\sum_{k=1}^K \int \lambda_k(u)du / \sum_{\ell=1}^q \int \lambda_\ell(u)\mathrm{d}u$ is greater than some prespecified large number. Although $\{\mathbf{a}_k(d)\}_{k=1}^q$ are orthonormal vectors in $\mathbb{R}^q$ for each d, the set $a_k(\cdot)_k$ does not form an orthonormal basis with respect to the functional norm induced by $\langle, \rangle$.

Related to this direction is the linear manifold learning approach (Chiou & Müller 2014) that seeks to extract the time-varying linear combinations $\{\mathbf{b}_k^\top(d)(\mathbf{X}(d) - \boldsymbol{\mu}(d))\}_{k=1}^K$ with zero integrated variance, $\eta_k = \int \mathrm{var}\{\mathbf{b}_k^\top(u)(\mathbf{X}(u) - \boldsymbol{\mu}(u))\}\mathrm{d}u$, across $k = 1, \ldots, K$ subject to $\langle \mathbf{b}_k, \mathbf{b}_{k'} \rangle = \mathbb{I}(k = k')$. Since $\eta_k = 0$ means $\sum_{\ell=1}^q b_k^\ell(d)(X^\ell(d) - \mu^\ell(d)) \equiv 0$ for all $d \in \mathcal{D}$, it follows that unlike pointwise PCA, which aims at maximal variation, manifold learning attempts to reduce the dimension by identifying a stable linear dependence between the component functions of the vector response.

Another way to sparsely characterize the dependence between the components, which is appealing when the number of functions, q, is large, is functional graphical models. When $\mathbf{X}(\cdot)$ follows a multivariate Gaussian process, $X^\ell(d)$ and $X^{\ell'}(d)$ are conditionally independent if and only if the conditional cross-covariance $C_{\ell\ell'}(d, d') = \mathrm{cov}(X^\ell(d), X^{\ell'}(d') \mid \{X^r(\cdot), r \neq \ell, \ell'\}) = 0$ for all $d, d' \in \mathcal{D}$ for $\ell \neq \ell'$. In the notion of graphical models, two nodes ℓ and ℓ' are connected via an edge if $C_{\ell\ell'}(d, d') \neq 0$ for some d, d'. Qiao et al. (2019) reduced the dimension of each component through truncated KL expansion to rewrite the edges by means of finite-dimensional precision matrix of the FPC scores, and they further implemented the tools available in the multivariate graphical literature to estimate the network. Zhu et al. (2016) provided a Bayesian version of the graphical model.

2.3. Estimation and Prediction

2.3.1. Estimation.
The mean function is estimated separately for each component using smoothing techniques and, typically, a working independence assumption (Wang et al. 2016).

Let $\widehat{\boldsymbol{\mu}}(d_{ij})$ be a vector of estimated mean functions; denote the residuals by $\widetilde{\mathbf{E}}_{ij} = \mathbf{Y}_{ij} - \widehat{\boldsymbol{\mu}}(d_{ij})$. For a dense design, for each ℓ, we estimate the latent processes indexed by i by smoothing the data $\{(\widetilde{E}_{ij}^{\ell}, d_{ij}) : j\}$; the sample covariance of the vectors of latent processes estimates the big covariance matrix $\boldsymbol{\Sigma}(d, d')$. Under a sparse design, one can pool the data for all subjects to obtain an estimator of the covariance. Specifically, the cross-covariance function is modeled as a tensor product of univariate bases, $\boldsymbol{\Sigma}_{\ell\ell'}(d, d') = \mathbf{B}^{\top}(d)\boldsymbol{\Gamma}_{\ell\ell'}\mathbf{B}(d')$, where $\mathbf{B}(d)$ is the vector of basis functions and $\boldsymbol{\Gamma}_{\ell\ell'}$ is a matrix of coefficients; $\boldsymbol{\Gamma}_{\ell\ell'}$ is estimated by penalized regression using the raw covariances $\{\widetilde{E}_{ij}^{\ell}\widetilde{E}_{ij'}^{\ell'}\}_{i,j}$ as pseudo responses for all $1 \le \ell \le \ell' \le q$ and setting $\boldsymbol{\Gamma}_{\ell\ell'} = \boldsymbol{\Gamma}_{\ell'\ell}^{\top}$. To bypass working with a large covariance dimension, Li et al. (2020) proposed an efficient way to extract the eigenfunctions by $\boldsymbol{\psi}_k(d) = (\mathbf{I}_q \otimes \mathbf{B}^{\top}(d)\mathbf{S}^{-\frac{1}{2}})\mathbf{V}_k$, where $\mathbf{S} = \int \mathbf{B}(u)\mathbf{B}^{\top}(u)\mathrm{d}u$ and $\{\mathbf{V}_k\}_k$ are eigenvectors of the block matrix with elements $\mathbf{S}^{\frac{1}{2}}\boldsymbol{\Gamma}_{\ell\ell'}\mathbf{S}^{\frac{1}{2}}$, for $1 \le \ell, \ell' \le q$. The positive definiteness of the estimated covariance is obtained by truncating the negative eigenvalues. The optimal number of eigenfunctions, K, is chosen by the Akaike information criterion (AIC), the Bayesian information criterion (BIC), or a prespecified PVE.

Under the heterogeneous domain, the eigencomponents are obtained from a three-stage algorithm (Happ & Greven 2018): (*a*) conduct univariate FPCA separately for each component, (*b*) estimate the cross-covariance between the FPC vectors, and (*c*) use these covariance estimates to appropriately weight the eigenfunctions/eigenvalues from univariate FPCA to extract $\boldsymbol{\psi}_k(d)$ and also obtain the FPC scores. This method is only suitable for densely observed MVFD because the univariate scores under a sparse design are shrunk toward zero, subsequently failing to capture the cross-correlations (Li et al. 2020).

2.3.2. Prediction. A major advantage of modeling the dependence in MVFD is that by using parsimonious models, such as MVFPCA, one can recover the entire vector of trajectories for a subject who may have observations for all the vector components. Once the mean function and the eigencomponents are estimated, prediction of the latent multivariate process entails prediction of the FPC scores. If all the profiles are observed over the common grid of d_{ij}s and the grid is fine, then $\{\zeta_{ik}\}$ can be estimated by numerical integration. In the case of a sparse design, or when subjects miss one or more profiles entirely, $\widehat{\zeta}_{ik}$ are obtained by best linear unbiased prediction under the mixed model $\mathbf{Y}_i = \boldsymbol{\mu}_i + \sum_{k=1}^{K} \widehat{\boldsymbol{\psi}}_{i,k}\zeta_{ik} + \boldsymbol{\epsilon}_i$ and under a Gaussian assumption, where $\mathbf{Y}_i$ is the long vector of responses obtained by stacking Y_{ij}^{ℓ} over first j and then ℓ, $\boldsymbol{\mu}_i$ is the mean vector, $\{\widehat{\boldsymbol{\psi}}_{i,k}\}$ are the estimated leading directions evaluated at the corresponding time points, and $\boldsymbol{\epsilon}_i$ is the vector of measurement error. Volkmann et al. (2021) proposed this framework to jointly estimate the mean and the scores. The latent vector process is predicted as $\widehat{\mathbf{X}}_i(d) = \widehat{\boldsymbol{\mu}}(d) + \sum_{k=1}^{K} \widehat{\zeta}_{ik}\widehat{\boldsymbol{\psi}}_k(d)$, with K chosen as above.

The pairwise interaction model (Chiou & Müller 2016) can also be used to reconstruct the trajectories by representing the interaction processes via truncated FPCA.

2.4. Depth Measures and Functional Boxplots

In many applications, functional data are often preprocessed before they enter the analysis. Detection of potential outliers accompanied by an informative visualization tool is crucial for robust inference. A functional observation can be identified as an outlier in terms of its shift, amplitude, or shape, or a combination of any of these. The depth function, initially developed for multivariate data (Zuo & Serfling 2000), offers an objective measure of centrality toward the deepest observation and outlyingness of a function in the sample with respect to the underlying distribution of curves. Let $F_{\mathbf{X}}(\cdot)$ be the cumulative distribution function of $\mathbf{X}(\cdot)$, and let $\mathbf{D}(\cdot, F_{\mathbf{X}}) : \mathbb{R}^q \to [0, 1]$ be a statistical depth function at each $d \in \mathcal{D}$. An example of such depth function for the univariate case is the halfspace depth, i.e., $D(X(u), F_X(u)) = 2\min\{F_X(u), 1 - F_X(u)\}$. Multivariate functional

depth (MVFDP) (Claeskens et al. 2014) is defined as $\text{MVFDP}(\mathbf{X}, F_{\mathbf{X}}) = \int_0^1 \mathbf{D}(\mathbf{X}(u), F_{\mathbf{X}}(u)) w(u) \mathrm{d}u$ for some unit norm weight function $w(\cdot)$. The MVFDP has a lower value for functions that are away from the center. Using this measure, a multivariate functional median is defined as the curve at which $\text{MVFDP}(\mathbf{X}, F_{\mathbf{X}})$ is maximized. A functional boxplot can be obtained by finding the band demarcated by the curves with the highest 25% and 75% depth measure. Hubert et al. (2015) provided some comprehensive data examples and choices of depth functions. Besides $\text{MVFDP}(\mathbf{X}, F_{\mathbf{X}})$, which measures only the magnitude of outlyingness, Dai & Genton (2019) developed a generalized depth function that quantifies also the directional outlyingness of a function. This is implemented in the generalized functional boxplot (Dai & Genton 2018b), and can detect departures from centrality with respect to both position and the shape.

2.5. Accounting for Covariate Information

In many applications, multiple covariates $\mathbf{W}_i = (W_{i1}, \ldots, W_{ip})^{\top}$ are also measured, so that the observed data are $[\{\mathbf{Y}_{ij}, d_{ij}\}, \mathbf{W}_i]_{i=1}^n$. When the objective is to study their association with the response, it is common to model the covariates' effect on the mean response. For scalar covariates, the multivariate varying coefficient model (Zhu et al. 2012) describes the mean as $\mu^\ell(\mathbf{W}_i, d) = \sum_{k=1}^p W_{ik} \beta_k^\ell(d)$ for $\ell = 1, \ldots, q$, where $\beta_k^\ell(\cdot)$ is a smooth effect of the kth covariate on the ℓth component. Estimation of $\{\boldsymbol{\beta}^\ell(d)\}_{\ell=1}^q$ is done separately for each ℓ by either local smoothing (Zhu et al. 2012) or a discrete wavelet transform FMM (Zhu et al. 2017). The multivariate functional additive mixed model uses a nonparametric mean of the form $\mu^\ell(\mathbf{W}_i, d) = \sum_{k=1}^p f_k^\ell(W_{ik}, d)$, for some smooth $f_k^\ell(W_{ik}, \cdot)$ (Volkmann et al. 2021). Functional predictors $\{W_{ik}(\cdot)\}_{k=1}^p$ can be accounted for as $\mu^\ell(\mathbf{W}_i, d) = \sum_{k=1}^p \int W_{ik}(s) \eta_k^\ell(s, d) \mathrm{d}s$, where $\eta_k^\ell(s, d)$ is a smooth effect of the kth predictor at the ℓth component. Like in the univariate functional data, this model is estimated by projecting both $W_{ik}(s)$ and $\eta_k^\ell(s, d)$ onto a common set of specified basis functions to convert it to a functional concurrent model (Liu et al. 2022).

2.6. Clustering Methods

Cluster analysis seeks to group the data so that objects within a group are more comparable with each other than objects from different groups, in terms of a specified measure. For example, in clinical electrocardiographs, multiple lead electrocardiograph signals are collected for the identification of cardiovascular ischemic diseases. Clustering these signals can be helpful for early diagnosis (Ieva & Paganoni 2013). Clustering algorithms for MVFD partition the n signals $\{\mathbf{X}_i(\cdot)\}_{i=1}^n$ into a set of disjoint homogeneous clusters and estimate the cluster membership of each curve from the data, as we review below.

2.6.1. Functional distance-based approach. Assuming that there are G (typically unknown) clusters, nonparametric distance-based clustering methods aim to find the cluster centroids $\{\mathbf{M}_g(\cdot)\}_{g=1}^G$ that minimize the weighted average distance between $\mathbf{M}_g(\cdot)$ and $\mathbf{X}_i(\cdot)$, i.e., $\sum_{i=1}^n w_i \|\mathbf{X}_i - \mathbf{M}_g\|$ using distance $\|\cdot\|$ for all possible choices of g and $\mathbf{M}_g$ and some suitable weights w_i. The object $\mathbf{M}_g$ can be interpreted as the center of data associated with the gth cluster, i.e., $\{\mathbf{X}_i(\cdot) : C_i = g\}$, with respect to $\|\cdot\|$, where $C_i \in \{1, \ldots, G\}$ denotes the cluster membership of $\mathbf{X}_i$. Ieva et al. (2013) considered the L_2 distance between the functions and/or the distance between the first derivative of the functions to develop a functional analogue to the k-means procedure. Tokushige et al. (2007) proposed a pointwise approach to estimate the cluster membership through an integrated version of the pointwise k-means.

Functional subspace-projected clustering (Chiou & Li 2007) assumes that the cluster centroids $\mathbf{M}_g$ admit a cluster-specific KL representation, i.e., $\mathbf{M}_g(d) = \boldsymbol{\mu}_g(d) + \sum_{k=1}^\infty \zeta_{g,k} \boldsymbol{\psi}_{g,k}(d)$, for

$g = 1, \ldots, G$; Park & Ahn (2017) showed its application to clustering warped MVFD. Dai & Genton (2018a) used the magnitude and shape of outlyingness (Section 2.4) combined with Mahalanobis distance as a robust measure of distance to assign cluster membership. Chiou (2012) developed a multinomial logistic regression based on the weighted distance between $\mathbf{M}_g$ and $\mathbf{X}_i$ to estimate the cluster membership probabilistically.

2.6.2. Basis representation.

One common approach is to represent the curves using preset basis functions $X_i^\ell(d) = \sum_{k=1}^K c_{ik}^\ell \phi_k^\ell(d)$, for basis $\{\phi_k^\ell(d)\}_{k \geq 1}$, $\ell = 1, \ldots, q$, and apply the traditional multivariate clustering methods to the stacked projections $\mathbf{c}_i = \{c_{ik}^\ell : k = 1, \ldots, K\}_{\ell=1}^q$. For example, a B-spline basis can be used in conjunction with k-means clustering of the coefficients or Gaussian basis combined with fitting a neural network to the coefficients (Kayano et al. 2010). Other approaches use MVFPC expansion (Section 2.2.1). Peng & Müller (2008) defined the L_2 distances between two curves as a function of the distances between the MVFPC scores in the truncated KL expansion and applied k-means to the finite dimensional scores to determine the cluster membership. Jacques & Preda (2014) posed a Gaussian latent mixture model on the cluster-specific MVFPC scores to determine the cluster probability assuming that the eigenfunctions $\boldsymbol{\psi}_g(\cdot)$ are common to all clusters.

Optimal selection of both the number of clusters and the finite truncation of the preset basis greatly affects the quality of clustering. The AIC, BIC, and integrated complete likelihood (Biernacki et al. 2000) are typically used for selection of G. Under the Bayesian framework, a standard way to estimate G is by maximum a posteriori estimation under a uniform prior.

2.7. Available Statistical Software

Notable software for MVFD include R packages `mulitfamm` (Volkmann 2021) to implement the multivariate functional additive mixed model discussed in Section 2.5, `MFPCA` (Happ-Kurz 2021) for dense MVFPCA under heterogeneous domain, and `mfaces` (Li & Xiao 2021) for MVFPCA under sparse design.

3. LONGITUDINAL FUNCTIONAL DATA ANALYSIS

3.1. Statistical Framework

Many longitudinal studies routinely collect data where the basic measurement is a function or a surface that can be repeatedly observed. In the study of cats with OA, introduced in Section 1, minute-by-minute daily physical activity of indoor cats was recorded repeatedly during a 20-day study period (Koner et al. 2022); **Figure 1** shows the daily physical activity profiles between 5 AM and 10 PM for three cats that were observed every day during this study period. The main study objective is broader; here, we consider how the daily physical activity changes during 20 days in household cats with OA. We refer to functional data collected in a longitudinal design as longitudinal functional data (LFD). In this setting, for every subject $i = 1, \ldots, n$ multiple functional data are observed corresponding to different time points, say $\{(d_{ijr}, Y_{ijr})_{r=1}^{R_{ij}}, \ t_{ij} : j = 1 \ldots, m_i\}$, where $(d_{ijr}, Y_{ijr})_{r=1}^{R_{ij}}$ are the functional data observed at time t_{ij}. We assume that $d_{ijr} \in \mathcal{D}$, the grids $\{d_{ij1}, \ldots, d_{ijR_{ij}}\}$ are dense in $\mathcal{D}$, and R_{ij} is large for each i, j. As in the classical longitudinal set-up, the number of repeated time points m_i is moderately large or small, but the set of time points $\{t_{ij} : i, j\}$ is dense in the compact interval $\mathcal{T}$. Without loss of generality, assume $d_{ijr} = d_r$ and $R_{ij} = R$.

We assume that the functional data $\{Y_{ij1}, \ldots, Y_{ijR}\}$ corresponding to time t_{ij} are error contaminated realizations of some latent bivariate process observed at time t_{ij}, $X_i(\cdot, t_{ij})$, evaluated at grid $\{d_{ij1}, \ldots, d_{ijR}\}$. We write $Y_{ijr} = X_i(d_r, t_{ij}) + \epsilon_{ijr}$, where $\{X_i(\cdot, \cdot)\}_{i=1}^n$ are i.i.d. copies of a latent

bivariate process $X(\cdot, \cdot)$ and ϵ_{ijr} are i.i.d. (across i, j and r) measurement errors. Define the mean $\mu(d, t) = \mathbb{E}[X(d, t)]$ and the cross-covariance operator $\Xi(t, t')$ with the associated cross-covariance kernel $\Sigma(d, d'; t, t') = \text{cov}\{X(d, t), X(d', t')\}$. One of the key focuses in the analysis of LFD is describing the longitudinal dynamics of the subject-level trajectories, which allows for full trajectory prediction at any time $t \in \mathcal{T}$, for a subject previously observed. Generally, the temporal dependence is modeled using approaches inspired by either the longitudinal data analysis or sparse FDA literature.

3.2. Modeling Mean Function

The specification of the mean model can incorporate any available information. For example, Zipunnikov et al. (2014) assume that the mean function is longitudinally invariant, i.e., $\mu(\cdot, t) = \eta(\cdot)$ for all $t \in \mathcal{T}$ for a smooth univariate function $\eta(\cdot)$; Koner et al. (2021) discussed statistical testing when such structure is doubtful. Other examples are (a) $\mu(d, t) = \mu_0(d) + \mu_1(d) t$ for smooth univariate functions $\mu_0(d)$ and $\mu_1(d)$, implying that the mean function is linear in t for every $d \in \mathcal{D}$, and (b) separable function of d and t, i.e., $\mu(d, t) = \mu_0(d) + \eta(t)$, where $\eta(t)$ is smooth function of $t \in \mathcal{T}$. Lacking any additional information, the mean function $\mu(d, t)$ is modeled as a smooth bivariate function. The mean function is represented using basis function representation, $\mu(d, t_{ij}) = \sum_{k=1}^{K} B_k(d) \alpha_k(t_{ij})$, for some suitable choice of basis $\{B_k(\cdot)\}_{k=1}^{K}$ in $\mathcal{L}^2(\mathcal{D})$, where $\alpha_k(t_{ij})$ are basis coefficients. Yuan et al. (2014) modeled the coefficients $\alpha_k(t_{ij})$ parametrically, while Park et al. (2017) proposed semiparametric models as $\alpha_k(t_{ij}) = \sum_{\ell=1}^{L_k} A_\ell(t_{ij}) \delta_{k\ell}$, for some basis $\{A_\ell(\cdot)\}$ in $\mathcal{L}^2(\mathcal{T})$. The latter one is equivalent to modeling $\mu(d, t)$ as a tensor product of two univariate bases.

3.3. Covariance Models for Longitudinal-Functional Dependence

Two main approaches are used to model dependence in longitudinal functional data. The first approach extends the common random effects model, from longitudinal data analysis, by replacing the usual random effects with processes. The second approach extends the functional principal component framework, from FDA, by allowing the scores to be time varying.

3.3.1. Longitudinally varying functional mixed model. Let $\widetilde{X}(d, t) = X(d, t) - \mu(d, t)$ be the mean zero process. Greven et al. (2011) modeled $\widetilde{X}(d, t)$ using an FMM framework as

$$\widetilde{X}_i(d, t_{ij}) = V_{0i}(d) + t_{ij} V_{1i}(d) + U_{ij}(d), \quad i = 1, \dots, n, j = 1, \dots, m_i, \qquad 1.$$

where the subject-level bivariate mean zero process $\mathbf{V}_i(d) = (V_{0i}(d), V_{1i}(d))^\top$ has covariance function $\mathbf{G}(d, d')$ and the residual process $U_{ij}(d)$ has covariance $H(d, d')$ and accounts for the deviation specific to jth measurement; $\mathbf{V}_i(\cdot)$ and $U_{ij}(\cdot)$ are assumed mutually independent. The linear dependence in t can be easily extended to accommodate higher-order polynomial terms in the time t. The model is identifiable as $\text{cov}\{X(d, t_{ij}), X(d', t_{ij'})\} = (1, t_{ij})\mathbf{G}(d, d')(1, t_{ij'})^\top + H(d, d')\mathbb{I}(j \neq j')$, where the first part quantifies the interfunction dependence longitudinally as a function of t.

Other structural assumptions, like separable additive or multiplicative structure, can be incorporated in the representation of $\widetilde{X}(d, t)$ (Huang et al. 2017). A common strategy is to represent $X_i(d, t)$ via preset basis functions, such as smoothing splines (Guo 2002), B-splines, or a wavelet basis (Morris et al. 2011), and convert the original model to the FMM framework. Shamshoian et al. (2022) used the tensor product of univariate bases and estimated the coefficients via a nonparametric Bayes approach based on a Gaussian assumption.

3.3.2. Longitudinal functional principal component analysis. Three different approaches can be recognized to perform longitudinal FPCA, a term coined by Greven et al. (2011) in

reference to the model in Equation 1 and the representation obtained by using KL expansion for the two processes $\mathbf{V}_i(\cdot)$ and $U_{ij}(\cdot)$. Specifically, assuming smooth covariances of both $\mathbf{G}(d, d')$ and $H(d, d')$, the two latent processes can be KL expanded as $\mathbf{V}_i(d) = \sum_{k=1}^{\infty} \xi_{ik}^V \boldsymbol{\phi}_k^V(d)$ and $U_{ij}(d) = \sum_{l=1}^{\infty} \xi_{ijl}^U \phi_l^U(d)$, where $\{\boldsymbol{\phi}_k^V(d), \lambda_k^V\}_k$ and $\{\phi_l^U(d), \lambda_l^U\}_l$ are the eigenfunction and eigenvalue pairs of the respective covariance operators, and $\xi_{ik}^V = \langle V_{0i}, \phi_{1k}^V \rangle + \langle V_{1i}, \phi_{2k}^V \rangle \sim (0, \lambda_k^V)$ and $\xi_{ijl}^U = \langle U_{ij}, \phi_l^U \rangle \sim (0, \lambda_l^U)$ are the associated FPC scores. Here $\boldsymbol{\phi}_k^V(d) = (\phi_{1k}^V(d), \phi_{2k}^V(d))^\top$ is the vector eigenfunction; the model in Equation 1 can be approximated by the finite KL truncation as

$$\widetilde{X}_i(d, t_{ij}) \approx \sum_{k=1}^{K_v} \xi_{ik}^V \left\{ \phi_{1k}^V(d) + t_{ij} \phi_{2k}^V(d) \right\} + \sum_{k=1}^{K_u} \xi_{ijk}^U \phi_k^U(d). \qquad 2.$$

The second approach, called double FPCA (Chen & Müller 2012), expands $\widetilde{X}_i(d, t)$ using two layers of KL expansions. Specifically, for each t, $\{\phi_k(\cdot|t)\}_k$ are the eigenfunctions of the covariance operator $\Xi(t, t)$ and represent the residual process at time t using KL as $\widetilde{X}_i(d, t) = \sum_{k=1}^{\infty} \xi_{ik}(t) \phi_k(d|t)$, where $\xi_{ik}(t) = \langle \widetilde{X}_i(\cdot, t), \phi_k(\cdot|t) \rangle$ are the time-varying basis coefficients. A second layer of KL expansion is done separately for each coefficient $\xi_{ik}(t)$ by representing $\xi_{ik}(t) = \sum_{p=1}^{\infty} \zeta_{ipk} \psi_{pk}(t)$, for $t \in \mathcal{T}$, where $\{\psi_{pk}(t)\}_{p \geq 1}$ are eigenfunctions of the covariance operator induced by $\text{cov}\{\xi_{ik}(t), \xi_{ik}(t')\}$, for each $k \geq 1$. This method requires simultaneous smoothing over d, d' and t to estimate $\Sigma(d, d'; t, t)$ and thus $\phi_k(d|t)$, and then of the covariance of the scores $\xi_{ik}(t)$ for every $t \in \mathcal{T}$, which is computationally intensive.

The computational bottleneck of double FPCA is overcome by projecting $\widetilde{X}_i(d, t)$ onto a set of time-invariant orthonormal basis $\{\phi_k(\cdot)\}_k$ in $\mathcal{L}^2(\mathcal{D})$. Park & Staicu (2015) proposed to select the basis of the spectral decomposition of the covariance operator induced by so-called marginal covariance kernel $\Sigma_{\mathcal{T}}(d, d') = \int_{\mathcal{T}} \Sigma(d, d', t) dF_{\mathcal{T}}(t)$, where $F_{\mathcal{T}}(t)$ is the sampling distribution of the longitudinal time points in $\mathcal{T}$ (see also Chen et al. 2017). The truncated KL expansion leads to

$$X_i(d, t_{ij}) \approx \mu(d, t_{ij}) + \sum_{k=1}^{K} \zeta_{ik}(t_{ij}) \phi_k(d), \qquad 3.$$

where $\zeta_{ik}(t_{ij}) = \langle \widetilde{X}_i(\cdot, t_{ij}), \phi_k \rangle$ are the marginal FPC scores and K is the truncation parameter. The longitudinal dynamics, captured by $\zeta_{ik}(t_{ij})$, can be flexibly characterized through the rich models for longitudinal data analysis, such as (*a*) by using the mixed model, $\zeta_{ik}(t_{ij}) = b_{0,ik} + b_{1,ik} t_{ij}$ for some coefficients $(b_{0,ik}, b_{1,ik})$; (*b*) by assuming a temporal isotropic covariance structure $\text{cov}\{\zeta_{ik}(t_{ij}), \zeta_{ik}(t_{ij'})\} = \lambda_\ell \rho_\ell(|t_{ij} - t_{ij'}|)$, for some known autocorrelation function $\rho_k(\cdot)$ with $\rho_k(0) = 1$; or (*c*) by nonparametric modeling $\zeta_{ik}(t) = \sum_{p=1}^{\infty} \eta_{ipk} \gamma_{pk}(t)$ for some known or data-driven basis. While the scores $\{\eta_{ipk}\}_{p \geq 1}$ are mutually uncorrelated for a fixed k, they may be correlated for different k. When the design for $\{t_{i1}, \ldots, t_{im_i}\}$ is dense in $\mathcal{T}$, then a more parsimonious, yet nonparametric, model can be used to describe the variation of $\zeta_{ik}(t)$. Chen et al. (2017) introduced the eigenbasis $\{\psi_p(\cdot)\}_p$ from the spectral decomposition of the marginal covariance over $\mathcal{D}$ induced by $\Xi_{\mathcal{D}}(t, t') = \int_{\mathcal{D}} \text{cov}\{X(u, t), X(u, t')\} du$. The use of this basis leads to $\widetilde{X}_i(d, t) = \sum_{k=1}^{\infty} \sum_{p=1}^{\infty} \eta_{ipk} \psi_k(d) \psi_p(t)$, where the scores $\{\eta_{ipk}\}_{p,k}$ are uncorrelated. The model in Equation 3 applied to the physical activity variation in cats helps uncover three main directions: a vertical shift, the contrast in activity between the first half and the second half of the day, and the contrast between the activity when their owner is at work versus at home (see **Figure 1*b***).

3.4. Estimation and Prediction of Full Trajectory

Let $\widehat{\mu}(d, t)$ be a smoothed mean estimator obtained by modeling the mean, as in the Section 3.2, and fitting a penalized criterion based on independence of the responses Y_{ijr}. Let the residuals be

$E_{ijr} = Y_{ijr} - \widehat{\mu}(d_r, t_{ij})$ for all i, j, and r. To estimate the model dependence, the residuals are pooled together and smoothing is done differently, corresponding to each approach in part. Greven et al. (2011) performed linear regressions of the raw covariances $E_{ijr}E_{ij'r'}$ onto $(1, t_{ij}, t_{ij'}, t_{ij}t_{ij'})$ for $j \neq j'$. Chen & Müller (2012) used three-dimensional smoothing of $E_{ijr}E_{ij'r}$ to estimate covariance surface $\Xi(d, d', t)$ at every d and t; Park & Staicu (2015) perform bivariate smoothing of the pooled sample covariances to construct a smooth estimator of $\widehat{\Sigma}_{\mathcal{T}}(d, d')$. The estimated eigenfunctions $\{\widehat{\phi}_\ell(\cdot)\}$ and eigenvalues $\widehat{\lambda}_\ell$ are obtained through the spectral decomposition of the estimated covariance.

Modeling the dependence allows us to make a prediction of the full trajectory $X(\cdot, t)$ at any time $t \in \mathcal{T}$ by means of Equation 2 or 3. Greven et al. (2011) predicted the full trajectory by $\widehat{X}(\cdot, t) = \widehat{\mu}(\cdot, t) + \sum_{k=1}^{K_v} \widehat{\xi}_{ik}^V \{\widehat{\phi}_{1k}^V(\cdot) + t\widehat{\phi}_{2k}^V(\cdot)\}$, where ξ_{ik}^V, along with ξ_{ijk}^U, are estimated from the mixed model $Y_{ijr} = \widehat{\mu}(d_{ijr}, t_{ij}) + \sum_{k=1}^{K_v} \xi_{ik}^V \{\widehat{\phi}_{1k}^V(d_{ijr}) + t_{ij}\widehat{\phi}_{2k}^V(d_{ijr})\} + \sum_{k=1}^{K_u} \xi_{ijk}^U \widehat{\phi}_k^U(d_{ijr}) + \epsilon_{ijr}$, for all $j = 1, \ldots, m_i$ and r. The computational complexity is $O(\max\{nRK_v, nK_v^3\})$. Here the truncations are estimated from the data. Park & Staicu (2015) first obtained the raw scores $\widetilde{\xi}_{ik}(t_{ij})$ by projecting the residuals $\{E_{ijr}\}_r$ onto the estimated $\widehat{\phi}_k(\cdot)$ and then fitting either a parametric model or a nonparametric model as described in Section 3.3.2.

Consistency rates of the estimator of covariance and the eigenfunctions depend on the sampling design at which the functional response is recorded—dense, ultradense, or sparse—and the smoothing techniques employed. Since nonparametric estimation of the mean in LFD requires smoothing along both d and t that is commonly observed on sparse grids, the optimal L_∞ rate of convergence for $\mu(d, t)$ is slower than the parametric rate (Chen & Müller 2012, Xiao 2020), and it corresponds to the optimal rate at which a two-dimensional covariance surface can be estimated. The same rate is for estimation of pointwise covariance $\Sigma(d, d', t, t)$ (Chen & Müller 2012). Zhu et al. (2019) derived the almost parametric $(\sqrt{n})$ L_2 rate of the covariance under the FMM framework, since the estimation does not require smoothing over t. If estimation of $\mu(d, t)$ does not require smoothing over t, [e.g., under a special structural assumption of $\mu(d, t)$, as in Greven et al. 2011], an almost parametric L_∞ rate of convergence for $\Sigma_{\mathcal{T}}(d, d')$ can be achieved under a dense design for d, or else the convergence rate is affected by the error of estimation of the mean (Koner et al. 2021). The estimation of eigenfunctions and eigenvalues can be done at an almost parametric rate if the covariance is estimated at the same rate. For covariance estimated at a slower rate, the eigenfunctions are estimated at a rate that corresponds to the optimal rate for estimating a univariate smooth function.

3.5. Available Statistical Software

The R function `fpca.lfda` in `refund` (Goldsmith et al. 2021) and the interactive visualization using `plot.shiny` (Wrobel et al. 2016) can be used for longitudinal FPCA.

4. FUNCTIONAL TIME SERIES DATA ANALYSIS

4.1. Set-Up and Assumptions

Suppose now that functional observations are collected sequentially over time, such as daily price curves of a financial stock over consecutive days or annual fertility rates as a function of a woman's age for every year since 1970, described in the R package `ftsa` (see also **Figure 2**). Mathematically, we conceptualize the problem by assuming that the observations are a noisy realization of some latent process $\{X_n(d), n \geq 1 : d \in \mathcal{D}\}$ observed at finite grids, for a compact interval $\mathcal{D}$, and such that for fixed d, the sequence $\{X_n(d), n \geq 1\}$ is a time series. Such a situation can arise naturally, as in the above examples, where the days/years are used to index the sequential time, n, or because
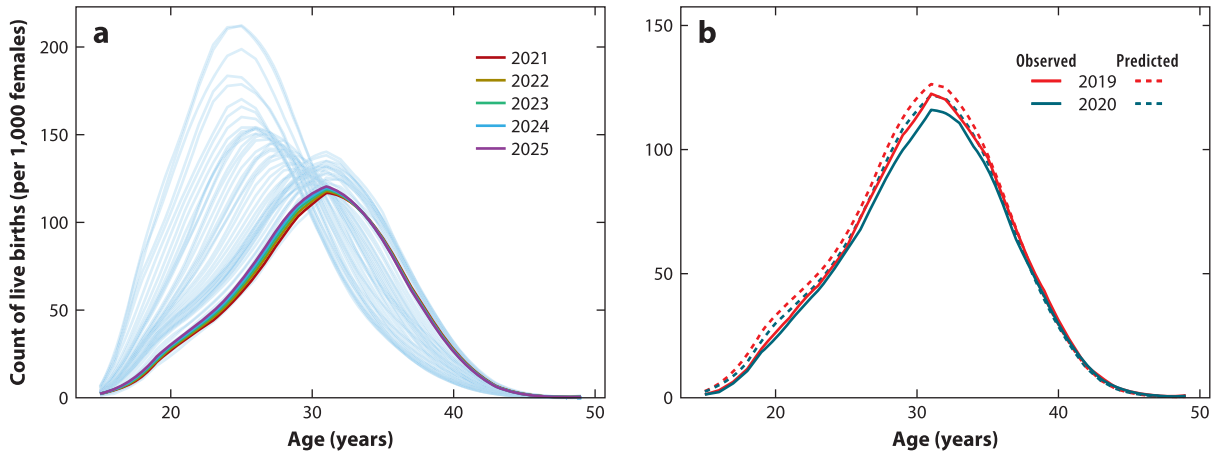
Figure 2

(*a*) Fertility rates in Australia as a function of age (1970–2020) with one to five years ahead forecasts of 2021–2025. (*b*) Observed and predicted one-year ahead fertility rates. Model based on *FAR*(1) and fitted via `ftsm` in R package `ftsa`.

the response is recorded as a long time series that can be partitioned into multiple shorter series corresponding to the same interval window; for example, $X_n(d)$ would correspond to all the observations recorded at time points of the form $n\delta + d$, for $d \in [0, \delta]$, by taking $\mathcal{D} = [0, \delta]$, or because a long time series is artificially divided into multiple shorter series corresponding to overlapping time intervals, which can be aligned to correspond to the same interval window $\mathcal{D}$. Unlike LFD analysis (Section 3), the time domain, indexed by n, is unbounded, $n \in Z_+$, and the entire series is observed once, thus without replications. Typical scientific goals include forecasting the trajectory behavior at future times, quantifying the dynamics of the process, and formally identifying changes in its mean, covariance, or other characteristics.

In the following, we denote the observed data by $\{(Y_{nj}, d_{nj}), j = 1, \ldots, m_n\}_{n=1,\ldots,N}$ and assume that $Y_{nj} = X_n(d_{nj}) + \epsilon_{nj}$, where $X_n(\cdot)$ is a stochastic process defined on $\mathcal{D}$ and $\epsilon_{nj} \sim (0, \tau^2)$ are i.i.d. measurement errors for each $n \geq 1$. Functional time series (FTS) analysis was popularized by Bosq (1991), who introduced the autoregressive Hilbert-valued processes of finite order. Since then, there have been extensive contributions to the literature in this area, including the comprehensive monographs of Bosq (2000) and Horváth & Kokoszka (2012). This review focuses on Hilbertian square integrable processes, $X_n \in L^2(\mathcal{D})$ for all $n \geq 1$, that are weakly (temporal) stationary.

An FTS is weakly (temporal) stationary if the mean and covariance of the series do not depend on the time at which the series is observed. Let $\mu : \mathcal{D} \to \mathrm{R}$ be the mean, $\mu(d) = E[X_n(d)]$, and denote by Ξ the simultaneous (or lag-zero) covariance operator, $\Xi(f)(d) = \int_{\mathcal{D}} \Sigma(d, u) f(u) du$, $f \in L^2(\mathcal{D})$, where $\Sigma(d, d') = \mathrm{cov}\{X_n(d), X_n(d')\}$ is the simultaneous covariance kernel function. For serially dependent data, the lag covariances characterize the serial dependence; define as Ξ_h the lag-h covariance operator and as $\Sigma_h(d, d') = \mathrm{cov}[X_{n+h}(d), X_n(d')]$ its associated lag h covariance kernel, $\Xi_h(f)(d) = \int_{\mathcal{D}} \Sigma_h(d, u) f(u) du$, $f \in L^2(\mathcal{D})$. An important quantity that aggregates the lag covariances of the series and is used to quantify the total serial dependence is the long-run covariance operator, defined as $\Xi_{\mathrm{long}} = \sum_{h \in \mathbb{Z}} \Xi_h$, provided the series is absolutely convergent. The long-run covariance, similar to the simultaneous one, is a symmetric and positive definite Hilbert–Schmidt operator. For independent functions X_n, the long-run covariance coincides with the simultaneous covariance and describes the functional dependence solely.

4.2. Modeling Framework

We start with reviewing some models that extend directly from time series literature and then consider models based on KL-inspired representations.

4.2.1. Functional autoregressive models.

A popular model for multi-step ahead predictions is the functional autoregressive (FAR) model. The first-order FAR model, or *FAR*(1), is defined by

$$X_n(d) - \mu(d) = \Psi_1(X_{n-1} - \mu)(d) + \varepsilon_n(d),\ d \in \mathcal{D},$$

where Ψ_1 is a bounded linear operator, $\Psi_1 : L^2(\mathcal{D}) \to L^2(\mathcal{D})$, with $\|\Psi_1\| < 1$, and ε_n are i.i.d. zero mean square integrable processes on $\mathcal{D}$. The operator inequality ensures that there exists a unique strictly stationary and causal solution in $L^2(\mathcal{D})$, and it can be relaxed (Horváth & Kokoszka 2012). Here we use the standard operator norm $\|\Psi\| = \sup_{f, \|f\|=1} \|\Psi f\|$.

Bosq (2000) studied *FAR*(1) by extending the Yule–Walker equations, known to lead to optimal estimation in classical time series (Brockwell & Davis 1991), to the functional setting. This approach expresses the solution as $\Psi_1 = \Xi_1 \Xi^{-1}$ and requires inversion of the simultaneous covariance operator Ξ. Several alternatives have been proposed to handle the challenge posed by the unbounded inverse operator, Ξ^{-1} (see, e.g., Bosq 1991, 2000; Besse et al. 2000; Antoniadis et al. 2006). Didericksen et al. (2012) reported that prediction using *FAR*(1), as discussed by Bosq (2000), shows superior performance in finite samples when compared with other alternatives. Martínez-Hernández et al. (2019) considered robust methods for parameter estimation, and Damon & Guillas (2002) extended the framework to accommodate exogeneous variables.

FAR(1) is generalized to a higher order of dependence, *FAR*(*p*), using the recursion $X_n(d) - \mu(d) = \sum_{l=1}^{p} \Psi_l(X_{n-l} - \mu)(d) + \varepsilon_n(d)$, where the operators Ψ_ls are linear such that Ψ_p is not the zero operator and $\sum_{l=1}^{p} \|\Psi_l\| < 1$. To determine the order p, Kokoszka & Reimherr (2013b) proposed a sequential testing procedure of hypotheses of the form $H_0^{(p)} : \{X_n\}$ are *FAR*(*p*), versus the alternative $H_A^{(p+1)} : \{X_n\}$ are *FAR*(*p* + 1); the starting null hypothesis is taken to be that $H_0^{(0)} : \{X_n\}$ are i.i.d. The test statistic uses a projection-based estimation of the unknown linear operator and is shown to have an approximately null chi-square distribution with degrees of freedom based on the projection space dimension.

4.2.2. Functional autoregressive moving average.

A class of models for which predictors cannot be directly derived are functional autoregressive moving average (FARMA) models (Bosq 2014). The FARMA(*p*, *q*) process is defined by

$$X_n(d) = \mu(d) + \sum_{l=1}^{p} \Psi_l(X_{n-l} - \mu)(d) + \sum_{\ell=1}^{q} \Theta_\ell(\varepsilon_{n-\ell})(d) + \varepsilon_n(d), d \in \mathcal{D}, \qquad 4.$$

where ε_n are i.i.d. zero mean square integrable random functions with finite fourth moments, and Ψ_l and Θ_ℓ are linear operators. Inspired by Brockwell & Davis (1991) for the ARMA models in classical time series, Klepsch et al. (2017) provided sufficient conditions on the operators to ensure a stationary and causal solution of Equation 4. The authors study theoretically the *h*-step prediction and prediction bounds by using the projection of this model to a suitable lower-dimensional space and prove that it approaches the best linear predictor (Bosq 2014) as the dimension of the projection space diverges. Fitting of this model is carried out using the KL framework presented in the next section and requires selection of the projection space dimension and the two tuning parameters, *p* and *q*; in practice, all parameters are selected using cross-validation (CV). Klepsch

et al. (2017) applied the FARMA(p,q) methodology to model and perform next-day prediction for traffic velocity data.

4.2.3. Functional autoregressive conditional heteroskedasticity.

For modeling financial time series data, such as daily monitoring of stocks indexes or exchange rates over long periods, nonlinear models are more adequate. Hörmann et al. (2013) defined FAR conditional heteroskedasticity (FARCH) as follows:

$$X_n(d) = \varepsilon_n(d)\sigma_n(d), \qquad \sigma_n^2(d) = \delta(d) + \Upsilon(X_{n-1}^2)(d), \quad \forall d \in \mathcal{D}, n \geq 1, \qquad 5.$$

where ε_n are i.i.d. square integrable functions in $L^2(\mathcal{D})$, δ is an unknown nonnegative function in $L^2(\mathcal{D})$, and Υ is an unknown nonnegative and bounded operator. The magnitude of the conditional covariance is modeled by the function $\sigma(\cdot)$ and the conditional correlation is entirely described by ε_n. Under mild conditions there exists a unique strictly stationary solution of Equation 5. Recently, Aue et al. (2017) and Cerovecki et al. (2019) developed the functional generalized autoregressive conditional heteroskedastic model and proposed a consistent estimation approach.

4.2.4. Karhunen-Loéve representation for functional time series.

An entirely different perspective is to use basis function representation. Let $\{\phi_1, \phi_2, \ldots\}$ be a unit norm orthogonal basis in $L^2(\mathcal{D})$, and consider the expansion of the FTS X_n in this basis:

$$X_n(d) = \mu(d) + \sum_{k \geq 1} \xi_{nk}\phi_k(d), \quad d \in \mathcal{D}, \qquad 6.$$

where the basis coefficients $\xi_{nk} = <X_n - \mu, \phi_k>$ capture the dependence structure of the series. To avoid using a preset basis, the functions ϕ_k are selected based on the spectral decomposition of the simultaneous covariance operator $\Xi = \sum_{k \geq 1} \lambda_k \phi_k \otimes \phi_k$, assuming strictly decreasing eigenvalues λ_k, where $\otimes$ is defined as $(\phi_k \otimes \phi_k)f(d) = <\phi_k, f> \phi_k(d)$. Using the eigenbasis of Ξ, Equation 6 is known as KL representation; Hyndman & Ullah (2007) discussed KL representation to describe an *FAR*(1) dependence, where the h-step ahead curve prediction is approximated by using a finite truncation of Equation 6 with estimated eigenfunctions and with coefficients that are predicted separately, using classical AR(1) models. Aue et al. (2015) studied prediction via KL representation (Equation 6) using a multivariate autoregressive model for the basis coefficients and showed that it is theoretically optimal and accommodates a variety of dependence structures, such as multivariate AR(p), MA(p) (moving average), and ARMA(p, q). For example, in the case of *FAR*(1), this would entail fitting the multivariate time series model $\boldsymbol{\xi}_n^K = B_K \boldsymbol{\xi}_{n-1}^K + \boldsymbol{\delta}_n^K$, where in bold we have the K-dimensional vectors of basis coefficients $\boldsymbol{\xi}_n^K$ and vector of innovations $\boldsymbol{\delta}_n^K$, with K being the truncation, and B^K is the $K \times K$ unknown matrix. The approach is readily implemented in R. We modeled the fertility rate data using *FAR*(1), which is fitted using this approach: **Figure 2a** shows one to five years ahead for 2021–2025, and **Figure 2b** shows the comparison between the one-year ahead predictions for 2019 and 2020, indicating little impact of COVID-19 on the fertility rates in Australia.

One criticism of this basis choice is that, while it captures the within curve dynamics, it fails to capture the serial dependence. An alternative option that addresses this matter with limited success is the eigenbasis of the long-run covariance operator, Ξ_{long}.

4.2.5. Dynamic Karhunen-Loéve representation for functional time series.

Panaretos & Tavakoli (2013) and Hörmann et al. (2015) proposed to study the series' dynamics using a frequency domain framework by means of the spectral density operator of the complete set of

lag autocovariance operators. The idea is to aggregate the lag covariance operators, but in the frequency domain, and construct functional filters later used to expand the time series.

Let $\omega \in [-\pi, \pi]$ index the frequencies, and define the spectral density operator at frequency ω by $\mathcal{F}_\omega^X = (2\pi)^{-1} \sum_{b \in Z} \Xi_b \exp(-ib\omega)$, presuming it exists, where i denotes the imaginary unit. As this is a nonnegative, self-adjoint Hilbert–Schmidt operator, for each ω it admits a spectral decomposition, $\mathcal{F}_\omega^X = \sum_{m \geq 1} \lambda_m(\omega) \varphi_m(\omega) \otimes \varphi_m(\omega)$, where $\{\lambda_m(\omega), \varphi_m(\omega)\}_m$ are the pairs of eigenvalue/eigenvector. The eigenvectors can be intuitively transformed back to the time domain and obtain so-called functional filters $\phi_{ml}(\omega) = (2\pi)^{-1} \int_{-\pi}^\pi \varphi_m(\omega|u) \exp(-ilu) du$, $l \in Z$ and $m \geq 1$, which are the building blocks for the dynamic representation of the functional series. Hörmann et al. (2015) introduced the dynamic KL expansion of X_n (convergence is in mean square),

$$X_n(d) = \mu(d) + \sum_{m \geq 1} \widetilde{X}_{mn}(d), \qquad \widetilde{X}_{mn}(d) = \sum_{l \in Z} \xi_{m,n+l} \phi_{ml}(d),$$

where $\xi_{m,n} = \sum_{l \in Z} < X_{n-l} - \mu, \phi_{ml} >$ is the mth dynamic FPC score (series convergences in mean square); the scores $\xi_{m,n}$ are uncorrelated across m, n. The number of dynamic components $\widetilde{X}_{mn}(d)$ is determined using a criterion analogous to the PVE for i.i.d. curves. The consistency of the dynamic scores relies on using a consistent spectral density operator estimator (Panaretos & Tavakoli 2013). This approach is useful in modeling the series; as the estimation requires data at future times, it is not clear how to use it for prediction.

4.3. Estimation of Mean and Covariance

Most methods assume the functional observations are measured at fine grids of points, possibly irregular, in $\mathcal{D}$. Common smoothing techniques (Ramsay & Silverman 2005, Wood 2006) can be applied to recover each smooth trajectory with negligible error (Zhang & Chen 2007); alternatively, the curve-level data can be smoothed by using a criterion based on truncated basis representation of the curves, which can be combined with an autoregressive dependence (Besse & Cardot 1996). The empirical mean and the empirical lag-h covariances are calculated as

$$\widehat{\mu}(d) = \frac{1}{N} \sum_{n=1}^N X_n(d); \qquad \widehat{\Sigma}_h(d, d') = \frac{1}{N} \sum_{n=\max(1, 1-h)}^{\min(N-h, N)} \{X_{n+h}(d) - \widehat{\mu}(d)\} \{X_n(d') - \widehat{\mu}(d')\} \qquad 7.$$

and are used to estimate the mean μ and the lag-h covariance functions Σ_h, respectively, with $\widehat{\Sigma} = \widehat{\Sigma}_0$ estimating the simultaneous autocovariance function Σ.

Under analogous conditions to the classical time series counterpart (Brockwell & Davis 1991), but for Hilbert spaces, the estimators in Equation 7 are unbiased and $\sqrt{N}$-consistent (Bosq 2000). Mas (2002) showed that, for a prespecified lag H, the set of estimators $\{\widehat{\Xi}_h : h = 1, \ldots, H\}$ is jointly asymptotically normal; in particular, the eigencomponents of the simultaneous covariance operator are asymptotically normal. Kokoszka & Reimherr (2013a) studied these properties under the weaker condition of L^4 m-approximability (Hörmann & Kokoszka 2010). A natural estimator of the long-run covariance is a kernel-based estimator

$$\widehat{\Sigma}_{\mathrm{long}}(d, d') = \sum_{b=-(N-1)}^{N-1} K\left(\frac{h}{b_N}\right) \widehat{\Sigma}_b(d, d'),$$

where the weights are based on a suitable kernel function $K(\cdot)$ with bandwidth b_N, which varies with N; Hörmann & Kokoszka (2010) showed the consistency of this estimator, and Horváth et al. (2016) proposed an adaptive way to select the bandwidth in practice.

The situation of sparse functional design is less common for FTS. Nonetheless, different estimation and prediction methods are needed; Kowal et al. (2019) considered this problem and assumed *FAR(p)* using Bayesian hierarchical Gaussian modeling. Research in this direction has only recently taken off (Rubín & Panaretos 2020).

4.4. Change Point Detection

Research in FTS often requires detecting changes in the mean structure or the autoregressive behavior, or beyond. Here we focus on single change point identification.

4.4.1. Change point detection in the mean.
Berkes et al. (2009) first studied the testing problem $H_0 : E[X_1] = \cdots = E[X_N]$ versus $H_A : E[X_n] = \cdots = E[X_{k^*}] \neq E[X_{k^*+1}] = \cdots = E[X_N]$ for unknown $k^* \in \{1, \ldots, N\}$ when the functional observations X_n are independent. A common way to test for a change point in the mean is by using cumulative sum (CUMSUM) type statistics such as

$$S_{N,k}(d) = N^{-1/2} \left\{ \sum_{n=1}^{k} X_n(d) - \frac{k}{N} \sum_{n=1}^{N} X_n(d) \right\}$$

to define the test statistic $T_N = \max_{k'} \|S_{N,k'}\|^2$. Under the null hypothesis and assuming the residuals of the X_ns are weakly stationary, Aue et al. (2018) proved that $T_N \rightarrow_d \sup_{x \in [0,1]} \sum_{l \geq 1} \lambda_l B_l(x)^2$ as $N \rightarrow \infty$, where λ_l are the eigenvalues of the long-run covariance kernel of X_ns and $B_l, l \geq 1$, are i.i.d. standard Brownian bridges defined on [0, 1]. In practice, the asymptotic null distribution is approximated with Monte Carlo simulation and using the leading eigenvalues of the estimated long-run covariance. For this purpose, the change point is estimated as $\widehat{k}_N = \min\{k : \|S_{N,k}\| = \max_{k'} \|S_{N,k'}\|\}$. The lag-$h$ covariances, required by the estimation of the long-run covariance, are estimated as in Equation 7, except the pairwise products $\{X_{n+h}(d) - \widehat{\mu}(d)\}\{X_n(d') - \widehat{\mu}(d')\}$ are adjusted to reflect a potential change in the mean; $\widehat{\mu}$ is calculated by $\widehat{\mu}_n^* = (\widehat{k}_N)^{-1} \sum_{n=1}^{\widehat{k}_n} X_n$ if $n \leq \widehat{k}_N$ and $\widehat{\mu}_n^* = (N - \widehat{k}_N)^{-1} \sum_{n=\widehat{k}_N+1}^{N} X_n$ if $n > \widehat{k}_N$. Asymptotic properties of the estimated change point are studied along with methods to construct confidence intervals. For detection of multiple changes in the mean function (epidemic change), Aston & Kirch (2012) considered low-rank projections of the functional observations using the long-run covariance and used CUMSUM tests for multivariate data.

4.4.2. Change point detection in the autoregressive behavior.
We turn now to testing that the lag covariance is constant over time; without loss of generality, the process is assumed zero mean. Horváth et al. (2010) investigated this problem for *FAR*(1) models, $X_n = \Psi_n(X_{n-1}) + \varepsilon_n$ for Ψ_n bounded and linear operators, where it simplifies to testing $H_0 : \Psi_1 = \cdots = \Psi_N$ and $H_A : \Psi_1 = \cdots = \Psi_{k^*} \neq \Psi_{k^*+1} = \cdots = \Psi_N$, for some k^*.

The idea is to use the leading directions of variability of the FTS and assess whether the action of the assumed common operator Ψ on the span of these directions changes over time. Testing the stability of the lag one covariance is now reduced to testing the stability of a long vector of appropriate means, and one can use a CUMSUM testing procedure similar to the one discussed in Section 4.4.1. Horváth et al. (2010) proposed a quadratic form based on the CUMSUM test and studied the null and alternative asymptotic distributions by using the long-run covariance estimator. Zhang et al. (2011) considered a self-normalized test that allows the investigation of the constancy of the two or higher-lag covariances.

4.4.3. Testing for stationarity.
Horváth et al. (2014) considered the general problem for testing the null hypothesis that the process is stationary versus general alternatives that include the

ones listed above. They proposed a CUMSUM test procedure $T_N = \int_0^1 \int_{\mathcal{D}} Z_N^2(x, u) \mathrm{d}u \mathrm{d}x$, where $Z_N(x, u) = S_N(x, u) - x S_N(1, u)$ and $S_N(x, u) = N^{-1/2} \sum_{n=1}^{\lfloor Nx \rfloor} X_n(u)$; its asymptotic null distribution is based on the eigenvalues of the long-run covariance operator. The asymptotic properties are studied under specific alternatives. More recently, Aue & Van Delft (2020) considered a similar testing problem, but from a frequency domain perspective. The key idea is that an FTS is weakly stationary if and only if the corresponding elements in the frequency domain are uncorrelated. The testing procedure combines methods from FPCA and spectral density operators and is a quadratic form of the estimated spectral covariances at various lags in the spectral domain. The asymptotic distribution of the test is studied both when the null is true and under local alternatives.

4.5. Available Statistical Software

Analysis of FTS is implemented in R via several packages: `far` (Julien & Serge 2010) implements the *FAR*(1) framework and allows incorporation of exogenous covariates; `ftsa` (Hyndman & Shang 2021) performs model fitting, prediction using FPCs, change point detection, and much more; and the combination of `fda` (Ramsay et al. 2021) and `marss` (Holmes et al. 2012) is used by Aue et al. (2015).

5. SPATIAL FUNCTIONAL DATA ANALYSIS

5.1. Notation and Set-Up

Some applications involve functional observations recorded across many spatial locations, such as yearly nitrate profiles at many monitoring stations across the United States (King et al. 2018). For many spatial locations, indexed by i and identified by $s_i \in \mathcal{S}$ for domain $\mathcal{S} \subset \mathbb{R}^2$, the observed data are $\{(Y_{ij}, d_{ij})_{j=1}^{n_i}\}$, with n_i large and $d_{ij} \in \mathcal{D}$ for compact interval $\mathcal{D}$. Such data structures are becoming increasingly common in environmental science, climatology, agronomy, oceanography, economics, remote sensing, demographics material science, biostatistics, and genetics, to name a few. Typical goals include understanding and describing the mean and dependence structure and prediction at unobserved spatial locations.

Spatiotemporal models are often employed for these problems (Chen et al. 2021); however, when, for each spatial location, the measurements in time are very dense, flexible spatiotemporal methods become computationally intractable. Modeling such data requires an alternative perspective. Analysis of spatial functional data combines techniques from functional data and spatial statistics to develop flexible approaches to model dependence, while ensuring flexibility and reasonable computations. Due to the explosion of big data with such characteristics, spatial functional statistics has attracted huge interest in the past 20 years, with important challenges and contributions to the field described by Bosq (2000) and Horváth & Kokoszka (2012) and reviewed by Mateu & Romano (2017).

Assume the observed spatial functional data are noisy realizations of the form $Y_{ij} = X(s_i; d_{ij}) + \epsilon_{ij}$, where ϵ_{ij} is an i.i.d. measurement error with zero mean and variance τ^2, and for each $s \in \mathcal{S}, X(s, \cdot) \in L^2(\mathcal{D})$. The object $\{X(s, d), s \in \mathcal{S}, d \in \mathcal{D}\}$ is a functional random field, also known as a spatial functional process (SFP) (Ferraty & Vieu 2006, Delicado et al. 2010). While spatial functional models have also been studied in describing multilevel functional data (briefly discussed in the Section 6), where the SFP is observed through multiple repetitions (Morris & Carroll 2006, Staicu et al. 2010), here we focus on a single realization of the process.

Denote as $\mu(s, \cdot)$ the mean of $X(s, \cdot)$, $\mu(s, d) = \mathrm{E}[X(s, d)]$, and by $\Xi_{crss}(s, s')$ the cross-covariance operator with the associated cross-covariance kernel, $\Sigma_{crss}(s, s'; d, d') = \mathrm{E}[\{X(s, d) - \mu(s, d)\} \{X(s', d') - \mu(s', d')\}]$. In spatial statistics, a popular way to describe the process dependence is

via semivariogram. Here, we use its functional analogue: the semivariogram operator, $\Gamma(s, s')$, and the associated semivariogram kernel $\gamma(s, s'; d, d') = (1/2)E[\{X(s, d) - X(s', d)\}\{X(s, d') - X(s', d')\}]$. Using the trace of the cross-covariance and semivariogram, summary measures can be constructed to quantify the global spatial dependence of X:

$$\Sigma_{tr}(s, s') = \int_{\mathcal{D}} \Sigma_{crss}(s, s'; u, u) du \qquad \text{and} \qquad \gamma_{tr}(s, s') = \int_{\mathcal{D}} \gamma(s, s'; u, u) du.$$

The trace covariogram $\Sigma_{tr}(s, s')$ and trace semivariogram $\gamma_{tr}(s, s')$ are a valid covariogram and semivariogram, respectively (Bosq 2000, Menafoglio et al. 2013).

As the complexity can easily increase, simplifying assumptions are generally made on the dependence structure of $X(\cdot, \cdot)$. Two common assumptions are that (a) $X(s, \cdot)$ is weakly stationary in the sense that its mean and auto-covariance functions are invariant to the spatial location and (b) $X(s, \cdot)$ has isotropic cross-covariance, meaning that the cross-covariance and semivariogram functions depend on the lag between the spatial locations and not on the locations per se. By an abuse of notation, in the case of a weakly stationary and isotropic residual process, we use the notation $\Sigma(d, d') = \text{cov}\{X(s, d), X(s, d')\}$ to denote the auto-covariance function, $\Sigma_{crss}(h; d, d')$ for the cross-covariance function at spatial lag h, and $\gamma(h; d, d')$ to denote the semivariogram function corresponding to spatial lag $h = \|s - s'\|$.

5.2. Modeling Framework: Semivariogram and Covariogram Estimation

We consider possible frameworks for modeling spatial functional data and discuss estimation of the corresponding model components. When estimation accounts for the spatial dependence, this is typically done via process semivariogram or covariogram.

5.2.1. Preset orthogonal basis expansion.
One direct approach to modeling weakly stationary and isotropic SFPs that reduces the dimensionality of the problem is to use a preset truncated orthogonal basis. Specifically, let $\{B_l(\cdot)\}_{l=1}^L$ be an orthonormal basis of functions in $L^2(\mathcal{D})$; then $X(s, d) = \sum_{l=1}^L B_l(d)\zeta_l(s)$, where $\zeta_l(s) = \int_{\mathcal{D}} X(s, u)B_l(u) du$ are scalar weakly stationary random fields for all $l = 1, \ldots, L$ with isotropic covariance. For a fixed truncation L, fitting the SFP reduces to estimation and prediction of multivariate random fields (Cressie 2015, Schabenberger & Gotway 2017). Methods like generalized least squares or maximum likelihood estimation based on a Gaussian assumption, which rely on positing parametric dependence structures, can be used. This approach works well for fine temporal grids that ensure accurate integral approximation. More importantly, it heavily depends on the basis truncation L, which is a tuning parameter; in practice, L is selected via CV.

5.2.2. Mean model for weakly (non)stationary processes.
Another approach, also inspired by the modeling in classical spatial statistics, is to decompose the latent process into the mean and the random deviation, $X(s, d) = \mu(s, d) + \widetilde{X}(s, d)$, where the residual $\widetilde{X}(s, d)$ has zero mean and is assumed to be a weakly stationary and isotropic process.

To accommodate a possible variation of the mean across space, when the process is observed a single time, we model $\mu(s, d)$ as a linear combination of known spatial-based predictors. Specifically, let $\{f_l(s)\}_{l=1}^L$ be known spatial predictors and suppose $\mu(s, d) = \sum_{l=1}^L f_l(s)\beta_l(d)$, where $\{\beta_l(\cdot)\}_l$ are unknown functional regression coefficients defined on $\mathcal{D}$. For parameter estimation, Menafoglio et al. (2013) propose an iterative procedure using generalized least squares; the algorithm is based on positing a parametric variogram model to the empirical trace semivariogram of the residual process.

If the mean function is space invariant—for simplicity, use the notation $\mu(d)$ instead—the sample mean is an obvious estimator. Hörmann & Kokoszka (2011) proved that the consistency of this estimator depends on a combination of the spatial dependence decay of the functions, captured by $\Sigma_{\text{tr}}(s, s')$, and the sampling design for the spatial locations. In particular, the sample mean is shown to be inconsistent under an infill domain sampling scheme. The sample mean assigns equal weight to all curves in the data, irrespective of their spatial locations, which is not ideal when the locations are not uniformly distributed in the spatial domain. Gromenko et al. (2012) and Martínez-Hernández & Genton (2020) considered mean estimation using a weighted mean $\widehat{\mu}(d) = \sum_{i=1}^{n} w_i X(s_i, d)$, for $\sum_{i=1}^{n} w_i = 1$, where the weights are selected to minimize the expected squared norm of the error, $E\|\widehat{\mu} - \mu\|^2$. This approach requires specifying a parametric model for the trace covariogram; the optimal weights are determined analytically in terms of the trace covariogram.

5.2.3. Karhunen-Loéve representation for spatial functional data.
Assume the SFP is stationary and isotropic and that the covariance $\Sigma(d, d')$ is positive semidefinite and symmetric and, furthermore, is assumed continuous on $\mathcal{D} \times \mathcal{D}$. Using Mercer's theorem, $\Sigma(d, d')$ can be decomposed into eigenfunctions and eigenvalues; let $\{\phi_k(\cdot), \lambda_k\}_{k \geq 1}$ be the eigenelements ordered corresponding to $\lambda_1 \geq \lambda_2 \geq \cdots \geq 0$. KL representation of latent process $X(s, d)$ is

$$X(s, d) = \mu(d) + \sum_{k \geq 1} \xi_k(s)\phi_k(d), \qquad\qquad 8.$$

where $\xi_k(s) = \int_{\mathcal{D}} \{X(s, u) - \mu(u)\}\phi_r(u)\mathrm{d}u$ are scalar, zero mean, weakly stationary random fields defined on $\mathcal{S}$. This representation allows $X(s, d)$ to be approximated based on the leading eigenfunctions (Horváth & Kokoszka 2012, Bohorquez et al. 2016).

The leading eigenfunctions are estimated from the spectral decomposition of an appropriate estimator of the covariance function, $\widehat{\Sigma}(d, d')$. As they are estimated from the data, their number is generally very small, unlike the truncation L of Section 5.2.1. Nonetheless, the method requires estimation of the covariance function $\Sigma(d, d')$. Three common approaches exist in the literature. Let $\widetilde{X}(s, \cdot) = X(s, \cdot) - \mu(\cdot)$ denote the residual latent process. The first approach is to estimate the covariance function by the sample covariance or, more generally, by a weighted mean of the raw covariances (Hörmann & Kokoszka 2011, Gromenko et al. 2012), defined as $\widehat{\Sigma}(d, d') = \sum_{i=1}^{n} w_i \widetilde{X}(s_i, d)\widetilde{X}(s_i, d')$; when $w_i = 1/n$, this leads to the sample covariance estimator. When not specified, the weights w_i are determined to minimize the expected Hilbert–Schmidt norm $\mathrm{E}[\int_{\mathcal{D}} \int_{\mathcal{D}} \{\widehat{\Sigma}(u, v) - \Sigma(u, v)\}^2 \mathrm{d}u\mathrm{d}v]$, which can be solved by positing a parametric model for the empirical semivariogram defined in Section 5.2.5; the resulting covariance is shown to be consistent, under an increasing spatial domain assumption.

A different direction is to represent the latent process via a known set of basis functions, $\{B_l(\cdot)\}_l$, similar to the approach shown in Section 5.2.1. Represent $\widetilde{X}(s, d) = \sum_{l=1}^{L} B_l(d)\widetilde{\zeta}_l(s)$, with zero mean random field coefficients $\widetilde{\zeta}_l(s)$; it follows that $\Sigma(d, d') = \boldsymbol{B}^{\top}(d)\mathrm{E}[\widetilde{\boldsymbol{\zeta}}(s)\widetilde{\boldsymbol{\zeta}}(s)^{\top}]\boldsymbol{B}(d')$, where $\boldsymbol{B}(d)$ is the L-dimensional column vector of $B_l(d)$s and $\widetilde{\boldsymbol{\zeta}}(s)$ is the L dimensional vector of $\widetilde{\zeta}_l(s)$s. This approach is quite complex and involves fitting parametric covariance models to each pair product of random field coefficients $\widetilde{\zeta}_l(s)\widetilde{\zeta}_{l'}(s)$ (see Gromenko et al. 2012).

The last direction first fits a tensor product spline estimator $\widehat{\Sigma}_{crss}(h; d, d')$ to the cross-covariance function at spatial lag h, $\Sigma_{crss}(h; d, d')$, by using pair products of residual processes—specifically, $\{\widetilde{X}(s_i, d)\widetilde{X}(s_{i'}, d') : (i, i') \text{ such that } \|s_i - s_{i'}\| \leq h\}$ (Zhang & Li 2021). Assuming the model in Equation 8 is valid, the covariance $\Sigma(d, d')$ can be then estimated by $\widehat{\Sigma}(d, d') = \int_0^{\Delta} \widehat{\Sigma}_{crss}(h; d, d')W(h)\mathrm{d}h$, where $W(\cdot)$ is a weight function, taken for simplicity to be $W(h) = 1$. This direction requires the specification of the threshold Δ, a tuning parameter.

5.2.4. Penalized regression splines.

A completely different direction is to model the latent SFP using basis representation. Let $\{A_r(s)\}_{r=1}^{R}$ and $\{B_l(d)\}_{l=1}^{L}$ be finite-dimensional bases over domains $\mathcal{S}$ and $\mathcal{D}$, respectively, and assume the representation $X(s, d) = \sum_{r=1}^{R} \sum_{l=1}^{L} A_r(s) B_l(d) \beta_{rl}$ holds, where the β_{rl}s are unknown regression coefficients. Bernardi et al. (2017) propose estimating the coefficients using a penalized criterion that reflects the spatial and temporal dependence. The modeling framework has the advantage that it can directly accommodate additional covariates or effects of interest (Arnone et al. 2019). Additionally, once the parameters are estimated, the entire process is recovered, and thus prediction at new sites is readily available. One challenge is the selection of smoothing parameters involved in the penalized criterion and of the appropriate basis truncation; existing methods use CV and AIC.

5.2.5. Semivariogram and covariogram.

Most of the models presented in this section require a valid trace semivariogram or covariogram model. One advantage of using variogram-based methods is that the variogram is robust to mean misspecification, as long as the process is weakly spatial stationary. The empirical semivariogram can be calculated by

$$\widehat{\gamma}(h; d, d') = \frac{1}{2|N(h)|} \sum_{(i,i') \in N(h)} \{X(s_i, d) - X(s_{i'}, d)\} \{X(s_i, d') - X(s_{i'}, d')\},$$

where $N(h) = \{(i, i') : \|s_i - s_{i'}\| = h\}$ and $|N(h)|$ denotes the cardinality of the set $N(h)$. The trace semivariogram is then estimated as $\widehat{\gamma}_{\mathrm{tr}}(h) = \frac{1}{2|N(h)|} \sum_{(i,i') \in N(h)} \int_{\mathcal{D}} \{X(s_i, u) - X(s_{i'}, u)\}^2 \mathrm{d}u$. For irregular data, the neighboring set is modified to $N(h) = \{(i, i') : \|s_i - s_{i'}\| \in (h - \epsilon, h + \epsilon)\}$.

Denote the trace covariogram by $\Sigma_{\mathrm{tr}}(h) = \int_{\mathcal{D}} \Sigma_{crss}(h; u, u) \mathrm{d}u$; the notation emphasizes the spatial dependence on the spatial lag solely. The empirical trace covariogram is

$$\widehat{\Sigma}_{\mathrm{tr}}(h) = \frac{1}{2|N(h)|} \sum_{(i,i') \in N(h)} \int_{\mathcal{D}} \{X(s_i, u) - \mu(u)\} \{X(s_{i'}, u) - \mu(u)\} \mathrm{d}u.$$

These estimates need adjustment to ensure they are proper variogram (conditionally negative definite) and covariogram (symmetric and positive semidefinite) models. Alternatively, valid parametric variogram (such as spherical, Gaussian, exponential, or Matérn) or covariogram models can be fitted to these models using least squares (Cressie 2015).

5.3. Curve Prediction at Unobserved Spatial Locations

A huge advantage for modeling the spatial dependence is that one can recover an entire trajectory at a new, unobserved spatial location of the spatial domain. If the SFP is modeled as in Section 5.2.4, then estimation suffices to obtain the prediction of $X(s_0, \cdot)$ at a new spatial location $s_0 \in \mathcal{S}$. In any other situation, prediction of a new curve requires prediction of the random fields. Unlike other areas, spatial statistics, through kriging, allows us to perform prediction without having to estimate the model components.

5.3.1. Kriging.

Some of the oldest work in spatial functional data is on kriging (Goulard & Voltz 1993). Initial approaches assumed weak stationarity and included a two-step procedure: (*a*) to get a prediction at s_0, for each time point d, perform kriging using all the available spatial data $\{(Y_{ij}, s_i) : i, j, \text{ such that } d_{ij} = d\}$, and (*b*) smooth the resulting predictions to ensure smoothness across $d \in \mathcal{D}$. This method is suitable for the case when the grid for the functional observations is regular and not too fine. For dense functional observations, different versions of kriging have been developed.

Giraldo et al. (2011) coined the term ordinal kriging, a method for predicting $\widehat{X}(s_0, \cdot)$ by $\widehat{X}(s_0, d) = \sum_{i=1}^{n} \lambda_i X(s_i, d)$ for all $d \in \mathcal{D}$ where the coefficients λ_i are estimated based on the constrained criterion

$$\min_{\lambda_1, \ldots, \lambda_n} \int_{\mathcal{D}} \mathrm{Var}\left\{\widehat{X}(s_0, u) - X(s_0, u)\right\} \mathrm{d}u, \qquad \mathrm{E}\left[\widehat{X}(s_0, d) - X(s_0, d)\right] = 0, \; \forall d \in \mathcal{D},$$

that ensures the prediction $\widehat{X}(s_0, \cdot)$ has the smallest variance and is unbiased. Kriging has been extended to accommodate increased flexibility in the λ weights, such as time-varying weights of the form $\lambda_i(d)$ (Giraldo et al. 2010) or weighting like $\widehat{X}(s_0, u) = \sum_{i=1}^{n} \int_{\mathcal{D}} \lambda_i(u, v) X(s_i, v) \mathrm{d}v$ (Monestiez & Nerini 2008). The minimization criterion is over the univariate and bivariate λ-functions, respectively. In some cases it might be relevant to look for predictions of the form $\widehat{X}(s_0, u) = \sum_{i=1}^{n} \int_0^u \lambda_i(v) X(s_i, v) \mathrm{d}v$, thus using only the curve's behavior observed prior to the current time point u, $\{X(s_i, d) : d \leq u\}_i$ in predicting $\widehat{X}(s_0, u)$. Kriging amounts to representing the curves and the weights, λ-functions, using a finite basis system, and solving a constrained optimization algorithm that requires a valid variogram model. The basis dimension is typically selected via CV. Caballero et al. (2013) and Menafoglio et al. (2013) introduced universal kriging, which provides a framework for curve prediction for nonstationary processes with spatially varying mean function.

5.3.2. Model-based prediction.

When the SFP is modeled as described in Section 5.2, it makes sense to perform prediction by predicting the random terms. We summarize the ideas for the case of the KL model representation. Suppose the mean and the leading eigenfunctions are estimated, and denote them by $\widehat{\mu}(d)$ and $\widehat{\phi}_k(d)$, respectively. Projecting on the leading directions, we obtain the pseudodata $\widetilde{\xi}_k(s_i) = \{\int X(s_i, u) - \widehat{\mu}(u)\}\widehat{\phi}_k(u)\mathrm{d}u$, for $i = 1, \ldots, n$. Prediction at new location $\widehat{\xi}_k(s_0)$ can be obtained from classical kriging or cokriging of these pseudodata (Liu et al. 2017). The entire curve is then predicted by $\widehat{X}(s_0, \cdot) = \widehat{\mu}(\cdot) + \sum_{k=1}^{K} \widehat{\xi}_k(s_0)\widehat{\phi}_k(\cdot)$, for some finite truncation K determined using PVE.

Recently, Zhang & Li (2021) reported good numerical performance by obtaining $\widehat{\xi}_k(s_0)$ as best linear unbiased predictions in the random effects model $Y_{ij} = \mu(d_{ij}) + \sum_{k=1}^{K} \xi_k(s_i)\phi_k(d_{ij}) + \epsilon_{ij}$, for $[\{(Y_{ij}, d_{ij})_j, s_i\} : |s_i - s_0| \leq \Delta$ for some threshold Δ and using consistent estimators of the mean, leading eigenfunctions and covariances of the scalar random fields.

5.4. Available Statistical Software

Kriging is implemented in the R package `geofd` (Giraldo et al. 2012), or by combining the functions `krigeST` and `fit.StVariogram` in the `gstat` R package (Pebesma 2004), the `RandomFields` R package (Schlather et al. 2015), and functions from the `fda` (Ramsay et al. 2021) R package. Prediction via regression is implemented in `fdaPDE` (Arnone et al. 2022).

6. FUTURE DIRECTIONS

In this review we attempt to cover a wide class of second-generation functional data that are collected using several classical schemes: multivariate, longitudinal, time series, and spatial. For each setting we discuss only a limited range of problems. In particular, we focus on Hilbert-valued latent processes observed with measurement error; we do not discus methods for binary-valued (Serban et al. 2013), nor mixed binary-continuous (Tidemann-Miller et al. 2016), nor manifold-valued functional data, nor do we carefully consider how to accommodate additional covariate information; for the latter, Morris (2015) discusses functional regression. Additionally, we do not review methods for significance testing for the second order of functional data (Crainiceanu et al. 2012).

The research in this area is motivated and often outpaced by data applications that increase not only in dimension but also in complexity. Hierarchical complexity enhances for all the discussed second-generation functional data, when multiple (nested) functions are observed within one subject or cluster, leading to multilevel second-generation functional data. Key modeling approaches for multilevel second-generation functional data (see Scott et al. 2013, chapter 13, for a review and data examples) can be adopted for more complex structured data. For example, multiple trajectories can be measured at each spatial location, leading to multilevel spatial functional data (Staicu et al. 2015, Scheffler et al. 2020); multilevel MVFD occurs if multidimensional trajectories are hierarchically observed (Wang & Tsung 2021). Some other pressing open problems are modeling data with non-Euclidean-valued functional data, modeling FTS when the functional observations are observed partially or are generalized-valued, and developing significance tests for MVFD where the functional observations are recorded on sparse grids.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

LITERATURE CITED

Antoniadis A, Paparoditis E, Sapatinas T. 2006. A functional wavelet–kernel approach for time series prediction. *J. R. Stat. Soc. Ser. B* 68(5):837–57

Arnone E, Azzimonti L, Nobile F, Sangalli LM. 2019. Modeling spatially dependent functional data via regression with differential regularization. *J. Multivar. Anal.* 170:275–95

Arnone E, Sangalli LM, Lila E, Ramsay J, Formaggia L. 2022. `fdapde`: Functional data analysis and partial differential equations (PDE); statistical analysis of functional and spatial data, based on regression with PDE regularization. *R Package* version 1.1-8. **https://CRAN.R-project.org/package=fdaPDE**

Aston JA, Kirch C. 2012. Evaluating stationarity via change-point alternatives with applications to fMRI data. *Ann. Appl. Stat.* 6(4):1906–48

Aue A, Horváth L, Pellatt DF. 2017. Functional generalized autoregressive conditional heteroskedasticity. *J. Time Ser. Anal.* 38(1):3–21

Aue A, Norinho DD, Hörmann S. 2015. On the prediction of stationary functional time series. *J. Am. Stat. Assoc.* 110(509):378–92

Aue A, Rice G, Sönmez O. 2018. Detecting and dating structural breaks in functional data without dimension reduction. *J. R. Stat. Soc. Ser. B* 80(3):509–29

Aue A, Van Delft A. 2020. Testing for stationarity of functional time series in the frequency domain. *Ann. Stat.* 48(5):2505–47

Berkes I, Gabrys R, Horváth L, Kokoszka P. 2009. Detecting changes in the mean of functional observations. *J. R. Stat. Soc. Ser. B* 71(5):927–46

Bernardi MS, Sangalli LM, Mazza G, Ramsay JO. 2017. A penalized regression model for spatial functional data with application to the analysis of the production of waste in Venice province. *Stochastic Environ. Res. Risk Assess.* 31(1):23–38

Berrendero JR, Justel A, Svarc M. 2011. Principal components for multivariate functional data. *Comput. Stat. Data Anal.* 55(9):2619–34

Besse PC, Cardot H. 1996. Approximation spline de la prévision d'un processus fonctionnel autorégressif d'ordre 1. *Can. J. Stat.* 24(4):467–87

Besse PC, Cardot H, Stephenson DB. 2000. Autoregressive forecasting of some functional climatic variations. *Scand. J. Stat.* 27(4):673–87

Biernacki C, Celeux G, Govaert G. 2000. Assessing a mixture model for clustering with the integrated completed likelihood. *IEEE Trans. Pattern Anal. Mach. Intel.* 22(7):719–25

Bohorquez M, Giraldo R, Mateu J. 2016. Optimal sampling for spatial prediction of functional data. *Stat. Methods Appl.* 25(1):39–54

Bosq D. 1991. Modelization, nonparametric estimation and prediction for continuous time processes. In *Nonparametric Functional Estimation and Related Topics*, ed. G Roussas, pp. 509–29. New York: Springer

Bosq D. 2000. *Linear Processes in Function Spaces: Theory and Applications*. New York: Springer

Bosq D. 2014. Computing the best linear predictor in a Hilbert space. Applications to general ARMAH processes. *J. Multivar. Anal.* 124:436–50

Brockwell PJ, Davis RA. 1991. *Time Series: Theory and Methods*. New York: Springer. 2nd ed.

Caballero W, Giraldo R, Mateu J. 2013. A universal kriging approach for spatial functional data. *Stochastic Environ. Res. Risk Assess.* 27(7):1553–63

Cerovecki C, Francq C, Hörmann S, Zakoian JM. 2019. Functional garch models: the quasi-likelihood approach and its applications. *J. Econom.* 209(2):353–75

Chen K, Delicado P, Müller HG. 2017. Modelling function-valued stochastic processes, with applications to fertility dynamics. *J. R. Stat. Soc. Ser. B* 79(1):177–96

Chen K, Müller HG. 2012. Modeling repeated functional observations. *J. Am. Stat. Assoc.* 107(500):1599–609

Chen W, Genton MG, Sun Y. 2021. Space-time covariance structures and models. *Annu. Rev. Stat. Appl.* 8:191–215

Chiou JM. 2012. Dynamical functional prediction and classification, with application to traffic flow prediction. *Ann. Appl. Stat.* 6(4):1588–614

Chiou JM, Chen YT, Yang YF. 2014. Multivariate functional principal component analysis: a normalization approach. *Stat. Sin.* 24:1571–96

Chiou JM, Li PL. 2007. Functional clustering and identifying substructures of longitudinal data. *J. R. Stat. Soc. Ser. B* 69(4):679–99

Chiou JM, Müller HG. 2014. Linear manifold modelling of multivariate functional data. *J. R. Stat. Soc. Ser. B* 76(3):605–26

Chiou JM, Müller HG. 2016. A pairwise interaction model for multivariate functional and longitudinal data. *Biometrika* 103(2):377–96

Claeskens G, Hubert M, Slaets L, Vakili K. 2014. Multivariate functional halfspace depth. *J. Am. Stat. Assoc.* 109(505):411–23

Crainiceanu CM, Staicu AM, Ray S, Punjabi N. 2012. Bootstrap-based inference on the difference in the means of two correlated functional processes. *Stat. Med.* 31(26):3223–40

Cressie N. 2015. *Statistics for Spatial Data*. New York: Wiley

Dai W, Genton MG. 2018a. An outlyingness matrix for multivariate functional data classification. *Stat. Sin.* 28(4):2435–54

Dai W, Genton MG. 2018b. Functional boxplots for multivariate curves. *Stat.* 7(1):e190

Dai W, Genton MG. 2019. Directional outlyingness for multivariate functional data. *Comput. Stat. Data Anal.* 131:50–65

Damon J, Guillas S. 2002. The inclusion of exogenous variables in functional autoregressive ozone forecasting. *Environmetrics* 13(7):759–74

Delicado P, Giraldo R, Comas C, Mateu J. 2010. Statistics for spatial functional data: some recent contributions. *Environmetrics* 21(3-4):224–39

Didericksen D, Kokoszka P, Zhang X. 2012. Empirical properties of forecasts with the functional autoregressive model. *Comput. Stat.* 27(2):285–98

Ferraty F, Vieu P. 2006. *Nonparametric Functional Data Analysis: Theory and Practice*. New York: Springer

Giraldo R, Delicado P, Mateu J. 2010. Continuous time-varying kriging for spatial prediction of functional data: an environmental application. *J. Agric. Biol. Environ. Stat.* 15(1):66–82

Giraldo R, Delicado P, Mateu J. 2011. Ordinary kriging for function-valued spatial data. *Environ. Ecol. Stat.* 18(3):411–26

Giraldo R, Mateu J, Delicado P. 2012. geofd: An R package for function-valued geostatistical prediction. *Rev. Colomb. Estad.* 35(3):385–407

Goldsmith J, Scheipl F, Huang L, Wrobel J, Di C, et al. 2021. `refund`: Regression with functional data. *R Package*, version 0.1-24. **https://CRAN.R-project.org/package=refund**

Goulard M, Voltz M. 1993. Geostatistical interpolation of curves: a case study in soil science. In *Geostatistics Tróia '92*, ed. A Soares, pp. 805–16. New York: Springer

Greven S, Crainiceanu C, Caffo B, Reich D. 2011. Longitudinal functional principal component analysis. In *Recent Advances in Functional Data Analysis and Related Topics*, ed. F Ferraty, pp. 149–54. New York: Springer

Gromenko O, Kokoszka P, Zhu L, Sojka J. 2012. Estimation and testing for spatially indexed curves with application to ionospheric and magnetic field trends. *Ann. Appl. Stat.* 6:669–96

Guo W. 2002. Functional mixed effects models. *Biometrics* 58(1):121–28

Guo W. 2004. Functional data analysis in longitudinal settings using smoothing splines. *Stat. Methods Med. Res.* 13(1):49–62

Han K, Hadjipantelis PZ, Wang JL, Kramer MS, Yang S, et al. 2018. Functional principal component analysis for identifying multivariate patterns and archetypes of growth, and their association with long-term cognitive development. *PLOS ONE* 13(11):e0207073

Happ C, Greven S. 2018. Multivariate functional principal component analysis for data observed on different (dimensional) domains. *J. Am. Stat. Assoc.* 113:649–59

Happ-Kurz C. 2021. `mfpca`: Multivariate functional principal component analysis for data observed on different dimensional domains. *R Package*, version 1.3-9. **https://CRAN.R-project.org/package=MFPCA**

Holmes EE, Ward EJ, Wills K. 2012. MARSS: multivariate autoregressive state-space models for analyzing time-series data. *R J.* 4(1):11–19

Hörmann S, Horváth L, Reeder R. 2013. A functional version of the arch model. *Econom. Theory* 29(2):267–88

Hörmann S, Kidziński Ł, Hallin M. 2015. Dynamic functional principal components. *J. R. Stat. Soc. Ser. B* 77(2):319–48

Hörmann S, Kokoszka P. 2010. Weakly dependent functional data. *Ann. Stat.* 38(3):1845–84

Hörmann S, Kokoszka P. 2011. Consistency of the mean and the principal components of spatially distributed functional data. In *Recent Advances in Functional Data Analysis and Related Topics*, ed. F Ferraty, pp. 169–75. New York: Springer

Horváth L, Hušková M, Kokoszka P. 2010. Testing the stability of the functional autoregressive process. *J. Multivar. Anal.* 101(2):352–67

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

Horváth L, Kokoszka P, Rice G. 2014. Testing stationarity of functional time series. *J. Econom.* 179(1):66–82

Horváth L, Rice G, Whipple S. 2016. Adaptive bandwidth selection in the long run covariance estimator of functional time series. *Comput. Stat. Data Anal.* 100:676–93

Huang L, Reiss PT, Xiao L, Zipunnikov V, Lindquist MA, Crainiceanu CM. 2017. Two-way principal component analysis for matrix-variate data, with an application to functional magnetic resonance imaging data. *Biostatistics* 18(2):214–29

Hubert M, Rousseeuw PJ, Segaert P. 2015. Multivariate functional outlier detection. *Stat. Methods Appl.* 24(2):177–202

Hyndman RJ, Shang HL. 2021. `ftsa`: Functional time series analysis. *R Package*, version 6.1. **https://CRAN.R-project.org/package=ftsa**

Hyndman RJ, Ullah MS. 2007. Robust forecasting of mortality and fertility rates: a functional data approach. *Comput. Stat. Data Anal.* 51(10):4942–56

Ieva F, Paganoni AM. 2013. Depth measures for multivariate functional data. *Commun. Stat. Theory Methods* 42(7):1265–76

Ieva F, Paganoni AM, Pigoli D, Vitelli V. 2013. Multivariate functional clustering for the morphological analysis of electrocardiograph curves. *J. R. Stat. Soc. Ser. C* 62(3):401–18

Jacques J, Preda C. 2014. Model-based clustering for multivariate functional data. *Comput. Stat. Data Anal.* 71:92–106

Julien D, Serge G. 2010. `far`: Modelization for functional autoregressive processes. *R Package*, version 0.6-3. **https://CRAN.R-project.org/package=far**

Kayano M, Dozono K, Konishi S. 2010. Functional cluster analysis via orthonormalized Gaussian basis expansions and its application. *J. Classif.* 27(2):211–30

King MC, Staicu AM, Davis JM, Reich BJ, Eder B. 2018. A functional data analysis of spatiotemporal trends and variation in fine particulate matter. *Atmos. Environ.* 184:233–43

Klepsch J, Klüppelberg C, Wei T. 2017. Prediction of functional ARMA processes with an application to traffic data. *Econom. Stat.* 1:128–49

Kokoszka P, Reimherr M. 2013a. Asymptotic normality of the principal components of functional time series. *Stochastic Proc. Appl.* 123(5):1546–62

Kokoszka P, Reimherr M. 2013b. Determining the order of the functional autoregressive model. *J. Time Ser. Anal.* 34(1):116–29

Koner S, Park SY, Staicu AM. 2021. PROFIT: projection-based test in longitudinal functional data. arXiv:2104.11355 [stat.ME]

Koner S, Staicu AM, Maity A. 2022. PROLIFIC: projection-based test for lack of importance of smooth functional effect in crossover design. arXiv:2205.08577 [stat.ME]

Kowal DR, Matteson DS, Ruppert D. 2019. Functional autoregression for sparsely sampled data. *J. Bus. Econ. Stat.* 37(1):97–109

Li C, Xiao L. 2021. `mfaces`: Fast covariance estimation for multivariate sparse functional data. *R Package*, version 0.1-3. **https://CRAN.R-project.org/package=mfaces**

Li C, Xiao L, Luo S. 2020. Fast covariance estimation for multivariate sparse functional data. *Stat* 9(1):e245

Liu C, Ray S, Hooker G. 2017. Functional principal component analysis of spatially correlated data. *Stat. Comput.* 27(6):1639–54

Liu X, Ma S, Chen K. 2022. Multivariate functional regression via nested reduced-rank regularization. *J. Comput. Graph. Stat.* 31(1):231–40

Martínez-Hernández I, Genton MG. 2020. Recent developments in complex and spatially correlated functional data. *Braz. J. Probab. Stat.* 34(2):204–29

Martínez-Hernández I, Genton MG, González-Farías G. 2019. Robust depth-based estimation of the functional autoregressive model. *Comput. Stat. Data Anal.* 131:66–79

Mas A. 2002. Weak convergence for the covariance operators of a Hilbertian linear process. *Stochastic Proc. Appl.* 99(1):117–35

Mateu J, Romano E. 2017. Advances in spatial functional statistics. *Stochastic Environ. Res. Risk Assess.* 31:1–6

Menafoglio A, Secchi P, Dalla Rosa M. 2013. A universal kriging predictor for spatially dependent functional data of a Hilbert space. *Electron. J. Stat.* 7:2209–40

Monestiez P, Nerini D. 2008. A cokriging method for spatial functional data with applications in oceanology. In *Functional and Operatorial Statistics*, ed. S Dabo-Niang, F Ferraty, pp. 237–42. New York: Springer

Morris JS. 2015. Functional regression. *Annu. Rev. Stat. Appl.* 2:321–59

Morris JS, Baladandayuthapani V, Herrick RC, Sanna P, Gutstein H. 2011. Automated analysis of quantitative image data using isomorphic functional mixed models, with application to proteomics data. *Ann. Appl. Stat.* 5(2A):894

Morris JS, Carroll RJ. 2006. Wavelet-based functional mixed models. *J. R. Stat. Soc. Ser. B* 68(2):179–99

Panaretos VM, Tavakoli S. 2013. Fourier analysis of stationary time series in function space. *Ann. Stat.* 41(2):568–603

Park J, Ahn J. 2017. Clustering multivariate functional data with phase variation. *Biometrics* 73(1):324–33

Park SY, Staicu AM. 2015. Longitudinal functional data analysis. *Stat* 4(1):212–26

Park SY, Staicu AM, Xiao L, Crainiceanu CM. 2017. Simple fixed-effects inference for complex functional models. *Biostatistics* 19(2):137–52

Pebesma EJ. 2004. Multivariable geostatistics in S: the gstat package. *Comput. Geosci.* 30(7):683–91

Peng J, Müller HG. 2008. Distance-based clustering of sparsely observed stochastic processes, with applications to online auctions. *Ann. Appl. Stat.* 2(3):1056–77

Qiao X, Guo S, James GM. 2019. Functional graphical models. *J. Am. Stat. Assoc.* 114(525):211–22

Ramsay J, Silverman B. 2005. Principal components analysis for functional data. In *Functional Data Analysis*, ed. JO Ramsay, BW Silverman, pp. 147–72. New York: Springer. 2nd ed.

Ramsay JO, Graves S, Hooker G. 2021. fda: Functional data analysis. R Package version 5.5.1. **https://CRAN.R-project.org/package=fda**

Rubín T, Panaretos VM. 2020. Sparsely observed functional time series: estimation and prediction. *Electron. J. Stat.* 14(1):1137–210

Schabenberger O, Gotway CA. 2017. *Statistical Methods for Spatial Data Analysis*. Boca Raton, FL: Chapman and Hall/CRC

Scheffler A, Telesca D, Li Q, Sugar CA, Distefano C, et al. 2020. Hybrid principal components analysis for region-referenced longitudinal functional EEG data. *Biostatistics* 21(1):139–57

Schlather M, Malinowski A, Menck PJ, Oesting M, Strokorb K. 2015. Analysis, simulation and prediction of multivariate random fields with package RandomFields. *J. Stat. Softw.* 63(8):1–25

Scott MA, Simonoff JS, Marx BD. 2013. The SAGE handbook of multilevel modeling. London: SAGE

Serban N, Staicu AM, Carroll RJ. 2013. Multilevel cross-dependent binary longitudinal data. *Biometrics* 69(4):903–13

Shamshoian J, Şentürk D, Jeste S, Telesca D. 2022. Bayesian analysis of longitudinal and multidimensional functional data. *Biostatistics* 23(2):558–73

Staicu AM, Crainiceanu CM, Carroll RJ. 2010. Fast methods for spatially correlated multilevel functional data. *Biostatistics* 11(2):177–94

Staicu AM, Lahiri SN, Carroll RJ. 2015. Significance tests for functional data with complex dependence structure. *J. Stat. Plan. Inference* 156:1–13

Tidemann-Miller B, Reich B, Staicu AM. 2016. Modeling multivariate mixed-response functional data. arXiv:1601.02461 [stat.ME]

Tokushige S, Yadohisa H, Inada K. 2007. Crisp and fuzzy k-means clustering algorithms for multivariate functional data. *Comput. Stat.* 22(1):1–16

Volkmann A. 2021. `multifamm`: Multivariate functional additive mixed models. *R Package*, version 0.1.1. **https://CRAN.R-project.org/package=multifamm**

Volkmann A, Stöcker A, Scheipl F, Greven S. 2021. Multivariate functional additive mixed models. arXiv:2103.06606 [stat.ME]

Wang JL, Chiou JM, Müller HG. 2016. Functional data analysis. *Annu. Rev. Stat. Appl.* 3:257–95

Wang K, Tsung F. 2021. Hierarchical sparse functional principal component analysis for multistage multivariate profile data. *IISE Trans.* 53(1):58–73

Wood SN. 2006. *Generalized Additive Models: An Introduction with R*. Boca Raton, FL: Chapman and Hall/CRC

Wrobel J, Park SY, Staicu AM, Goldsmith J. 2016. Interactive graphics for functional data analyses. *Stat* 5(1):108–18

Xiao L. 2020. Asymptotic properties of penalized splines for functional data. *Bernoulli* 26(4):2847–75

Yuan Y, Gilmore JH, Geng X, Martin S, Chen K, et al. 2014. FMEM: Functional mixed effects modeling for the analysis of longitudinal white matter Tract data. *NeuroImage* 84:753–64

Zhang H, Li Y. 2021. Unified principal component analysis for sparse and dense functional data under spatial dependency. *J. Bus. Econ. Stat.* **https://doi.org/10.1080/07350015.2021.1938085**

Zhang JT, Chen J. 2007. Statistical inferences for functional data. *Ann. Stat.* 35(3):1052–79

Zhang X, Shao X, Hayhoe K, Wuebbles DJ. 2011. Testing the structural stability of temporally dependent functional observations and application to climate projections. *Electron. J. Stat.* 5:1765–96

Zhu H, Chen K, Luo X, Yuan Y, Wang JL. 2019. FMEM: functional mixed effects models for longitudinal functional responses. *Stat. Sin.* 29(4):2007

Zhu H, Li R, Kong L. 2012. Multivariate varying coefficient model for functional responses. *Ann. Stat.* 40(5):2634

Zhu H, Morris JS, Wei F, Cox DD. 2017. Multivariate functional response regression, with application to fluorescence spectroscopy in a cervical pre-cancer study. *Comput. Stat. Data Anal.* 111:88–101

Zhu H, Strawn N, Dunson DB. 2016. Bayesian graphical models for multivariate functional data. *J. Mach. Learn. Res.* 17(204):1–27

Zipunnikov V, Greven S, Shou H, Caffo B, Reich DS, Crainiceanu C. 2014. Longitudinal high-dimensional principal components analysis with application to diffusion tensor imaging of multiple sclerosis. *Ann. Appl. Stat.* 8(4):2175

Zuo Y, Serfling R. 2000. General notions of statistical depth function. *Ann. Stat.* 28(2):461–82