

REVIEW ARTICLE OPEN ACCESS 

Functional Data Analysis: An Introduction and Recent Developments

Jan Gertheiss¹  | David Rügamer^{2,3} | Bernard X. W. Liew⁴ | Sonja Greven⁵

¹Department of Mathematics and Statistics, School of Economics and Social Sciences, Helmut Schmidt University, Hamburg, Germany | ²Department of Statistics, LMU Munich, Munich, Germany | ³Munich Center for Machine Learning, Munich, Germany | ⁴School of Sport, Rehabilitation and Exercise Sciences, University of Essex, Essex, UK | ⁵Chair of Statistics, School of Business and Economics, Humboldt-Universität zu Berlin, Berlin, Germany

Correspondence: Jan Gertheiss (jan.gertheiss@hsu-hh.de)**Received:** 1 December 2023 | **Revised:** 17 May 2024 | **Accepted:** 27 May 2024**Funding:** Sonja Greven received funding from grants GR 3793/2-2, GR 3793/3-1, and GR 3793/8-1 from the German Research Foundation (DFG). Bernard Liew is supported by The Academy of Medical Sciences, UK, Springboard Award SBF006/1019.**Keywords:** curve data | functional regression | image data | longitudinal data analysis | object-oriented data analysis

ABSTRACT

Functional data analysis (FDA) is a statistical framework that allows for the analysis of curves, images, or functions on higher dimensional domains. The goals of FDA, such as descriptive analyses, classification, and regression, are generally the same as for statistical analyses of scalar-valued or multivariate data, but FDA brings additional challenges due to the high- and infinite dimensionality of observations and parameters, respectively. This paper provides an introduction to FDA, including a description of the most common statistical analysis techniques, their respective software implementations, and some recent developments in the field. The paper covers fundamental concepts such as descriptives and outliers, smoothing, amplitude and phase variation, and functional principal component analysis. It also discusses functional regression, statistical inference with functional data, functional classification and clustering, and machine learning approaches for functional data analysis. The methods discussed in this paper are widely applicable in fields such as medicine, biophysics, neuroscience, and chemistry and are increasingly relevant due to the widespread use of technologies that allow for the collection of functional data. Sparse functional data methods are also relevant for longitudinal data analysis. All presented methods are demonstrated using available software in R by analyzing a dataset on human motion and motor control. To facilitate the understanding of the methods, their implementation, and hands-on application, the code for these practical examples is made available through a code and data supplement and on [GitHub](#).

1 | Introduction

In functional data analysis (FDA), curves, images, or functions on higher dimensional domains constitute the observations and the objects of interest in the analysis (Ramsay and Silverman 2005). The goals of FDA, such as descriptive analyses, classification, regression, etc., are often the same as for statistical analyses of scalar-valued or multivariate data, and many corresponding methods have been transferred to function-valued data, while FDA also brings additional challenges including infinite/high

dimensionality of observations and parameters as well as misalignment. In contrast to simpler methods that reduce the functional observations to scalar summary values, FDA retains all important information by directly using the functional observations in the analysis. Functional data occur in many different fields (Ullah and Finch 2013), including medicine (Sørensen, Goldsmith, and Sangalli 2013), biophysics (Liew et al. 2020), neuroscience (Rügamer et al. 2018), or chemistry (Brockhaus et al. 2015). Nowadays, technologies such as imaging techniques, electroencephalographs and electrocardiographs, gyroscopes, or

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](#) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2024 The Author(s). *Biometrical Journal* published by Wiley-VCH GmbH.

accelerometers, standard in biomedical research or even built into everyday devices such as smartphones, controllers, or wristwatches, allow for seamlessly collecting functional data. Consequently, many research fields observe an increased interest in methods to analyze this type of data. In this paper, we describe a range of the most common statistical analysis techniques in FDA, focusing primarily on their practical application, their respective software implementations, and some more recent developments in the field. We aim to provide an introduction and tutorial for some of the most important topics in FDA, while providing a partial review with further reading for some additional, more advanced topics. For readers interested in theoretical details, we refer to Cuevas (2014), Hsing and Eubank (2015), Kokoszka and Reimherr (2017), and Zhang (2013).

1.1 | Functional Data, Longitudinal Data, and Time Series

Datasets collected over time occur in different fields and under different names, and we thus briefly discuss differences between FDA and longitudinal as well as time-series data for readers new to the field.

In health and social sciences, repeated observations over time on the same subjects or observational units are often termed longitudinal data (LD) or panel data. LD usually involves only a few repeated measurements per observation unit (e.g., subject observed before medication and at two follow-up appointments) with potentially different (numbers of) time points for different observation units. While functional data (FD) with different measurement points across observations exists (so-called sparse FD), the data is often recorded on the same (time) interval, with the same frequency, and exhibits a large number of measurements per observation unit. Methods to analyze FD and LD thus traditionally differ in how data is understood and treated. While the repeated measurements of observation units in the analysis of LD are often accounted for by using random effects in a parametric model, FD usually treats the observed sequences per measurement unit as observations of a smooth process observed at discrete time points. It thus models this data using non- or semiparametric methods, aided by the typically larger number of available measurement points per curve. Nevertheless, there is an overlap between the two areas, and ideas and methods from (sparse) FDA have more recently been used to make LD models less parametric and more flexible (e.g., Goldsmith, Greven, and Crainiceanu 2013; Köhler et al. 2017; Yao, Müller, and Wang 2005). A recent overview of developments and emerging areas can be found in Li, Qiu, and Xu (2022).

Time-series analysis, on the other hand, typically considers a single observation of a stochastic process on an equidistant time grid from a single observation unit (e.g., values of a share at one stock market observed over time). It is thus different from LD and FD where repeated observations on each of a sample of observational units are available. Due to the limited amount of data, methods usually focus on parametric estimation and often aim for good forecasting performance for unseen future time points. Goals and data structure thus differ from LD and FDA. If a sample of different time series is available, however, as, for example, considered in the field of time series classification, and

the focus is less on parametric methods, the setting can also be considered from a functional data viewpoint.

1.2 | Notation

In this paper, we will usually focus on a sample of functional data x_i considered to be realizations of random functions X_i , $i = 1, \dots, n$, with values $X_i(t) \in \mathbb{R}$, $t \in \mathcal{T}$, over an interval $\mathcal{T} \subset \mathbb{R}$. Note that both t and $X_i(t)$ can be generalized to take values in some other domain than \mathbb{R} , but this will not be the focus here. The mean function will be denoted by $\mathbb{E}(X(t)) = \mu(t)$ and the covariance function by $\text{Cov}(X(s), X(t)) = \Sigma(s, t)$, $s, t \in \mathcal{T}$. In practice, the functional data are only observed on a discrete grid of measurement points $t_{il} \in \mathcal{T}$, giving observations $x_i(t_{il})$, $l = 1, \dots, L_i$, $i = 1, \dots, n$. For so-called dense functional data, the grid is equal across functions and (relatively) dense in \mathcal{T} . For so-called sparse functional data, the grids are curve-specific and can be relatively sparse, but should together still cover \mathcal{T} well. If the grids differ but are relatively dense in \mathcal{T} , one also speaks of irregular functional data. While a common assumption in FDA is the smoothness of the functions, they are often observed with additional measurement errors. Then, we observe $X_i(t_{il}) = \tilde{X}_i(t_{il}) + \varepsilon_{il}$ at $t_{il} \in \mathcal{T}$, $l = 1, \dots, L_i$, $i = 1, \dots, n$, where $\tilde{X}_i(t)$ is the underlying (smooth) function and ε_{il} are independent and identically distributed (i.i.d.) noise terms. Functions are commonly considered as elements of the \mathcal{L}^2 space of functions

$$\mathcal{L}^2(\mathcal{T}) = \left\{ f : \mathcal{T} \rightarrow \mathbb{R} \mid \int_{\mathcal{T}} f^2(t) dt < \infty \right\}$$

with scalar product $\langle f, g \rangle = \int_{\mathcal{T}} f(t)g(t) dt$

for functions $f, g \in \mathcal{L}^2(\mathcal{T})$, which induces the norm $\|f\|_{\mathcal{L}^2} = \langle f, f \rangle^{1/2}$ and a Hilbert space structure.

1.3 | Running Data Example(s): Human Motion and Motor Control

Data from the biomechanical analysis of human motion and motor control almost always exhibit temporal, spatial, or both spatiotemporal characteristics within regular discrete bounds (e.g., gait events). For example, the flexion-extension knee joint angle during running over the stance phase of running reflects a one-dimensional (1D) time-series variable (Pataky 2010); the plantar pressure distribution of the foot at an instant during standing reflects a two-dimensional (2D) spatial variable (Montagnani et al. 2021), and the three-dimensional (3D) bone strain patterns of the femur at an instant during an exercise reflects a spatial variable (Martelli et al. 2014). The typical approach to analyzing biomechanical data is to first reduce multidimensional data into a single scalar variable (e.g., maximal value), followed by statistical inference. However, statistically analyzing discretized biomechanical data could, in some instances, lead to elevated Type 1 and 2 errors (Pataky, Vanrenterghem, and Robinson 2015, 2016; Robinson, Vanrenterghem, and Pataky 2015). In addition, it is commonly the interest of researchers to not only know “what” is statistically different but, in human movement, “where” and “when” the differences occur. One approach used to answer these

TABLE 1 | Variables available in the human motion and motor control data example, obtained as the combinations of “body part,” “kinematics/kinetics,” and “axis”; the latter also corresponds to specific types of motion, which may differ between body parts.

Body part	Kinematics/kinetics	Axis	Type of motion
Ankle hip knee	Acceleration $\left(\frac{\text{degree}}{\text{s}^2}\right)$ Angle (degree) Moment $\left(\frac{\text{Nm}}{\text{kg}}\right)$ Velocity $\left(\frac{\text{degree}}{\text{s}}\right)$	Anterior-posterior Medial-lateral Vertical	Inversion-eversion (ankle)/ abduction-adduction (hip, knee) Plantarflexion-dorsiflexion (ankle)/ flexion-extension (hip, knee) Abduction-adduction (ankle)/ axial rotation (hip, knee)

questions is FDA, allowing, for example, in functional regression to model repeated measurements and to include functional biomechanical variables as the outcome (Warmenhoven et al. 2018), the covariates (Liew et al. 2020), or both (Liew et al. 2021).

The data we use in this paper are derived from three primary sources—a publicly available running dataset (Fukuchi, Fukuchi, and Duarte 2017), and two datasets from the third author’s (BL) research on load carriage running (Liew, Morris, Keogh, et al. 2016; Liew, Morris, and Netto 2016). Details of the experimental methods will be briefly mentioned here. In the first publicly available dataset on running (with sample size $n = 28$) in healthy adults (Fukuchi, Fukuchi, and Duarte 2017), running was performed on a dual-belt, force-instrumented treadmill (300 Hz; Bertec, USA), while lower limb kinematic trajectories were captured with 12 optoelectronic cameras (150 Hz; Motion Analysis Corporation, USA). Participants performed running with shoes at 2.5, 3.5, and 4.5 m/s. The second dataset comes from a previously published work investigating the effects of load carriage on running biomechanics ($n = 31$) (Liew, Morris, and Netto 2016), which involved participants performing overground running across in-ground embedded force platforms (2000 Hz; AMTI, Watertown, MA, USA) while carrying three load conditions (0%, 10%, 20% body weight (BW)) across three velocities (3.0, 4.0, 5.0 m/s). Lower-limb kinematic trajectories were captured using an 18 camera motion capture system (Vicon T-series, Oxford Metrics, UK; 250 Hz). The last dataset comes from a project investigating the influence of strength training on load carriage running biomechanics ($n = 31$) (Liew, Morris, Keogh, et al. 2016). Participants performed overground running at a fixed velocity of 3.5 m/s ($\pm 10\%$) while carrying two load conditions (0%, 20% BW). Motion capture equipment was identical to Liew, Morris, Keogh, et al. (2016).

For all datasets, 3D bilateral lower limb kinematics (joint angle, velocity, acceleration) and kinetics (moments) of the ankle, knee, and hip joints were extracted as variables, which yields $3 \times 4 \times 3 = 36$ variables as the combinations of the entries in the first three columns of Table 1. All kinematic and moment variables were time normalized to 101 data points within the stance phase of each lower limb. The joint moment was normalized to body mass (N/kg).

For illustration, Figure 1 (left) shows the knee flexion-extension acceleration (divided by 10,000) across the stance phase of running for all participants from the second study (Liew, Morris, and Netto 2016), when running at 3 m/s with no backpack. Colors correspond to the maximum moment. One goal of the data analysis could be to investigate how acceleration profiles relate

to the often-used scalar quantity “maximum moment.” In the right panel of Figure 1, the hip abduction-adduction acceleration is shown for females (solid blue) and males (dashed black). Here, it could be interesting to study if and where males and females differ in their profiles.

The remainder of this paper is structured as follows. We first introduce fundamental aspects of functional data in Section 2 including descriptive analysis methods, smoothing approaches, amplitude and phase variation, and functional principal component analysis (FPCA). In Section 3, we then summarize different parametric and nonparametric regression approaches before discussing methods for statistical inference with functional data in Section 4. Methods beyond regression such as classification and clustering are considered in Section 5 followed by an outline of machine and deep learning approaches in Section 6. In a final outlook section (Section 7), we briefly discuss relationships of FDA to other fields as well as FDA beyond 1D functional observations. All sections are complemented with a selection of existing software packages that implement the respective approaches.

2 | Fundamentals

2.1 | Descriptives and Outliers

The first step in FDA usually is a descriptive analysis of the data, including visualizations and potentially the identification of outlying observations that may be influential in the analysis. For scalar data, this is relatively straightforward using tools such as a boxplot. For functional data, the task is more complex. In particular, outlying curves can differ from the majority of the sample by the range of their function values (“magnitude outliers”) and/or by their shape (“shape outliers”) (see Hyndman and Shang 2010).

It is thus first necessary to define a notion of outlyingness or centrality of such data, which can be done using so-called functional data depths. Different versions have been developed; see Gijbels and Nagy (2017) for an in-depth discussion of the desirable properties of statistical depths for functional data and their (non-)fulfillment for several proposals, and Chakraborty and Chaudhuri (2014) for results showing that some of the developed depth functions actually have degenerate behavior in infinite dimensions (leading to population version depth measures of identically zero for a broad class of models).

To define a functional boxplot, Sun and Genton (2011) use the band-depth or modified band-depth of López-Pintado and Romo

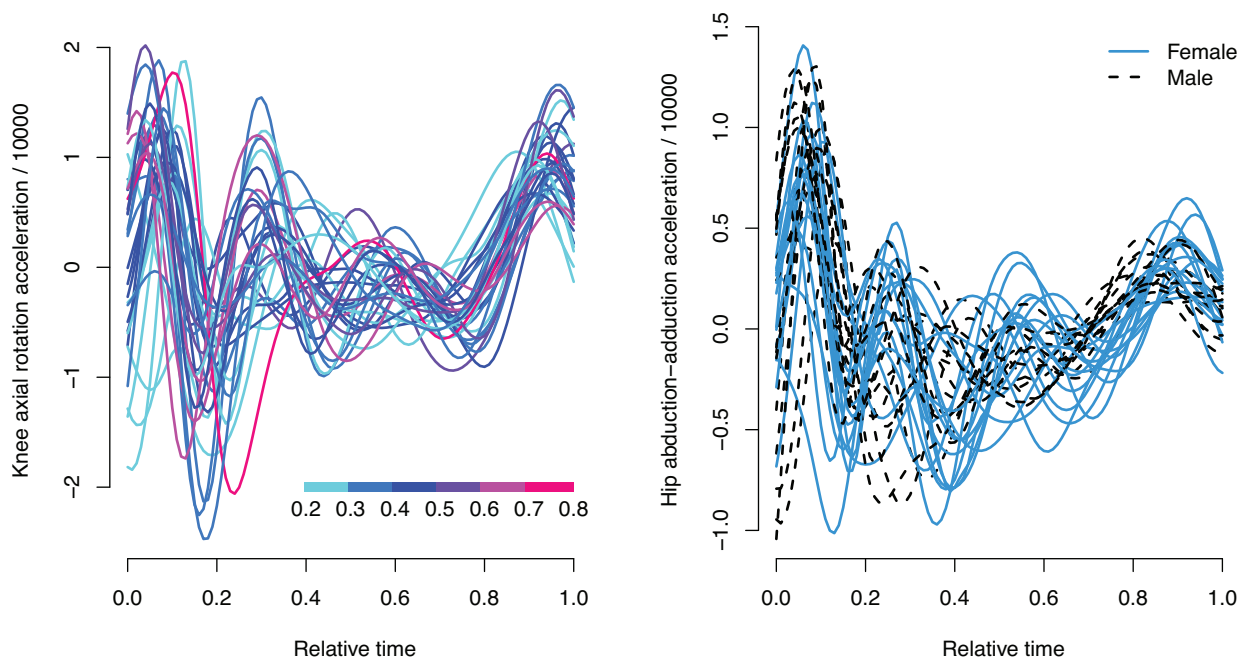


FIGURE 1 | Knee axial rotation acceleration measurements (left) with colors corresponding to the maximum moment, and hip abduction-adduction acceleration profiles (right) with colors corresponding to subjects' sex.

(2009), the latter showing nondegenerate behavior in infinite dimensions as shown in Chakraborty and Chaudhuri (2014), which looks at the fraction of bands between pairs of other curves in the sample that contain (part of) a given curve. The sample median then is the curve with the highest functional depth, whereas the 50% central region (corresponding to the interquartile range of a standard boxplot) is defined to envelop the 50% most central curves in the sample with the highest depth. This region is inflated by 1.5 to generate the analog of whiskers in the boxplot, and any curve lying outside this region is flagged as an outlier. Note that while here the functional median is defined via the highest value of the chosen functional depth measure, other proposals for a (geometric) median extend the expected absolute difference minimization property of the scalar median to general Hilbert spaces (Cardot, Cénac, and Zitt 2013).

Alternative visualizations include the rainbowplot, functional bagplot, and functional highest density region boxplot of Hyndman and Shang (2010), which are built on corresponding versions for multivariate data for the first two scores of an FPCA (see Section 2.4), thus capturing only the two main modes of variation in the data. A simple visualization method can also be to color functions according to another variable of interest, for example, to inform a regression analysis (compare Figure 1, left), or according to a well-chosen functional depth. The outliergram of Arribas-Gil and Romo (2014) specifically aims to detect shape outliers. Dai and Genton (2018) discuss visualization and outlier detection for multivariate functional data.

For our running example, Figure 2 shows an example rainbow plot and functional bagplot for the knee axial rotation acceleration profiles of Figure 1 (left) obtained using the `rainbow` R package. Note that the depth measure in both plots is not a proper functional depth, but a multivariate depth using only the first two functional principal components (FPCs; compare

Section 2.4) and thus only captures about 62% of the variation in the functional data.

2.2 | Smoothing for Functional Data

Functional data are in practice observed on finite grids—which may be curve-specific—and commonly with measurement error. A typical model for the observations thus is $X_i(t_{il}) = \tilde{X}_i(t_{il}) + \varepsilon_{il}$, where $X_i(t_{il})$ constitutes the observable value at $t_{il} \in \mathcal{T}$, $l = 1, \dots, L_i$, $i = 1, \dots, n$, $\tilde{X}_i(t)$ is the underlying (smooth) function and ε_{il} are i.i.d. noise terms (compare Section 1.2). As part of an analysis, it is often of interest to reconstruct the underlying “true” curves $\tilde{X}_i(t)$.

The earliest approaches in FDA usually do this reconstruction as an initial step using presmoothing of the curves (Ramsay and Silverman 2005). For a given curve i , we can view $X_i(t_{il}) = \tilde{X}_i(t_{il}) + \varepsilon_{il}$ with i.i.d. noise ε_{il} as a nonparametric regression problem and use corresponding approaches for scatterplot smoothing. Typical methods used include kernel smoothing, local polynomial smoothing, and spline or penalized spline approaches; see Ramsay and Silverman (2005) and Zhang (2013) for fuller discussions.

We here focus on a (spline) basis expansion approach for $\tilde{X}_i = \tilde{X}_i(\cdot)$, that is, we assume that it can be written or approximated (asymptotically) as a linear combination of spline basis functions ϕ_k , $k = 1, \dots, K$, and write

$$X_i(t_{il}) = \tilde{X}_i(t_{il}) + \varepsilon_{il} = \sum_{k=1}^K \phi_k(t_{il})\theta_{ik} + \varepsilon_{il}. \quad (1)$$

Unknown coefficients θ_{ik} are estimated using a least squares or penalized least squares criterion, penalizing the integrated

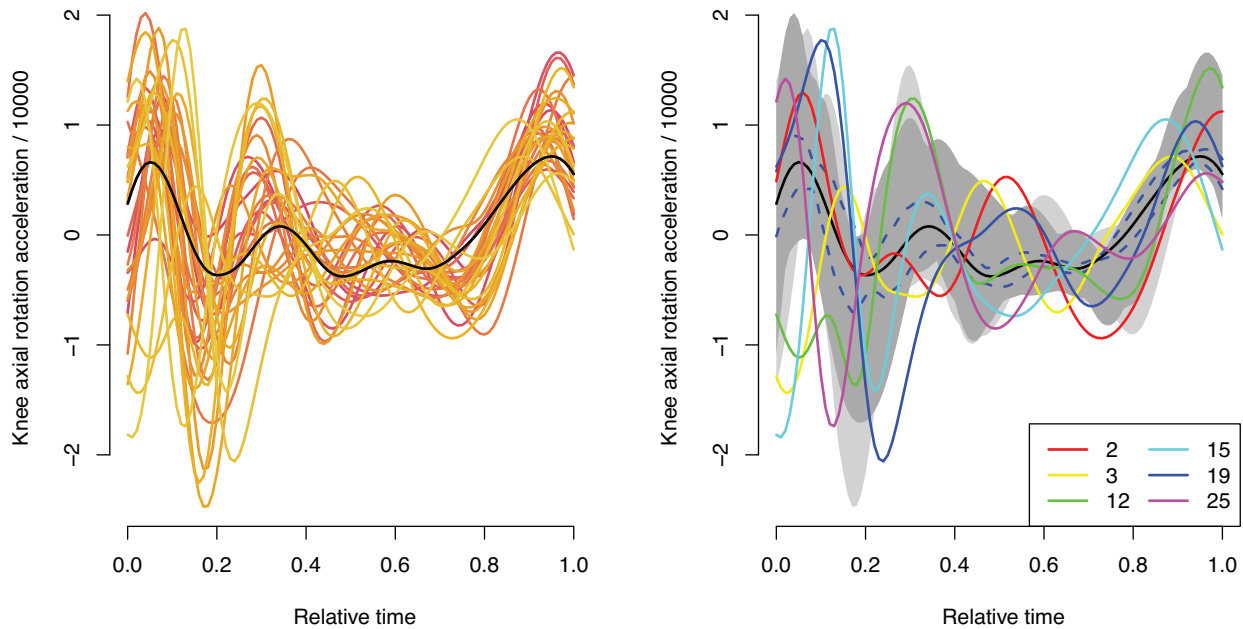


FIGURE 2 | Rainbow plot for the knee axial rotation acceleration profiles from Figure 1 (left) with yellow colors indicating low and red high depth (of the first two FPC scores); sample median given in black in both plots, with 95% pointwise confidence intervals (right); functional bagplot with 50% inner region and 99% fence region shaded in dark and light gray, respectively, and flagged outlying curves highlighted in color (right).

squared (e.g., second) derivative of the function to encourage smoothness or an approximation such as the finite differences of neighboring B-spline coefficients (Eilers and Marx 2021).

After presmoothing, the resulting curves are treated as if they were truly functional observations, that is, fully observed curves. The advantage of this presmoothing approach is that theory and methods developed for observations that are truly functional (e.g., Bosq 2000; Dauxois, Pousse, and Romain 1982) can be directly applied. The clear disadvantage is that the uncertainty in estimating $\tilde{X}_i(t)$ from noisy and discrete data is ignored in subsequent analysis steps. In particular, for sparse functional data, this source of uncertainty can be substantial and vary across curves depending on the number of available observation points. Also, smoothing can fail or not work well if only a few points for a curve are observed (Yao, Müller, and Wang 2005).

Thus, more recent approaches have often worked with the observed data and incorporated one of the smoothing methods into the analysis itself, in particular for sparse functional data (Yao, Müller, and Wang 2005). This has the clear advantage that the true information content of the data is acknowledged both in the analysis and the uncertainty quantification. Also, and importantly, such an approach allows “borrowing of strength” across functions, which is particularly useful in the case of sparse functional data. This approach was spearheaded by Yao, Müller, and Wang (2005) for the case of FPCA discussed in Section 2.4 below, where smoothing is incorporated in the estimation of the covariance function. Similar ideas were then included in other methods for functional data, such as in regression (e.g., Greven and Scheipl 2017; Goldsmith et al. 2020; Scheipl, Staicu, and Greven 2015). We will introduce these approaches in the corresponding sections.

A further approach that is similar to presmoothing is to expand the functions in a basis prior to analysis and subsequently work with the multivariate vector of basis coefficients instead of the presmoothed functions. If the basis transformation is lossless, as can be the case for, for example, wavelets and functions on regular grids (Morris and Carroll 2006), this approach can be equivalent to working with the functions (up to potential simplifying assumptions in the subsequent model such as independence between coefficients). For splines (Ramsay and Silverman 2005), this approach is very similar to presmoothing, and after the denoising usually the resulting uncertainty in curve reconstruction is also ignored. Other approaches besides presmoothing that include some sort of denoising as preprocessing of functional data include FPCA (e.g., Goldsmith et al. 2011) or factor models (Hörmann and Jammoul 2022).

As data from our running example are already smooth, they can be nearly perfectly reconstructed using a B-spline basis expansion with a sufficient number of knots, for example, $K = 30$. Figure 3 illustrates this for the first knee axial rotation acceleration profile using the `fda` package. On the left, a basis with only 10 basis functions is shown for better visibility. The reconstruction of the functional observation on the right with 30 basis functions is extremely close to the observed data points.

2.3 | Amplitude and Phase Variation

In functional data, we often observe two kinds of variation: *amplitude* and *phase* variation. When displaying functions by plotting $x_i(t)$ versus $t \in \mathcal{T}$, these become visible as variations “along the y-axis” and “along the x-axis,” respectively. Typical examples are functions over time such as, for example, growth curves (Ramsay and Silverman 2005), where there is variation both in the height

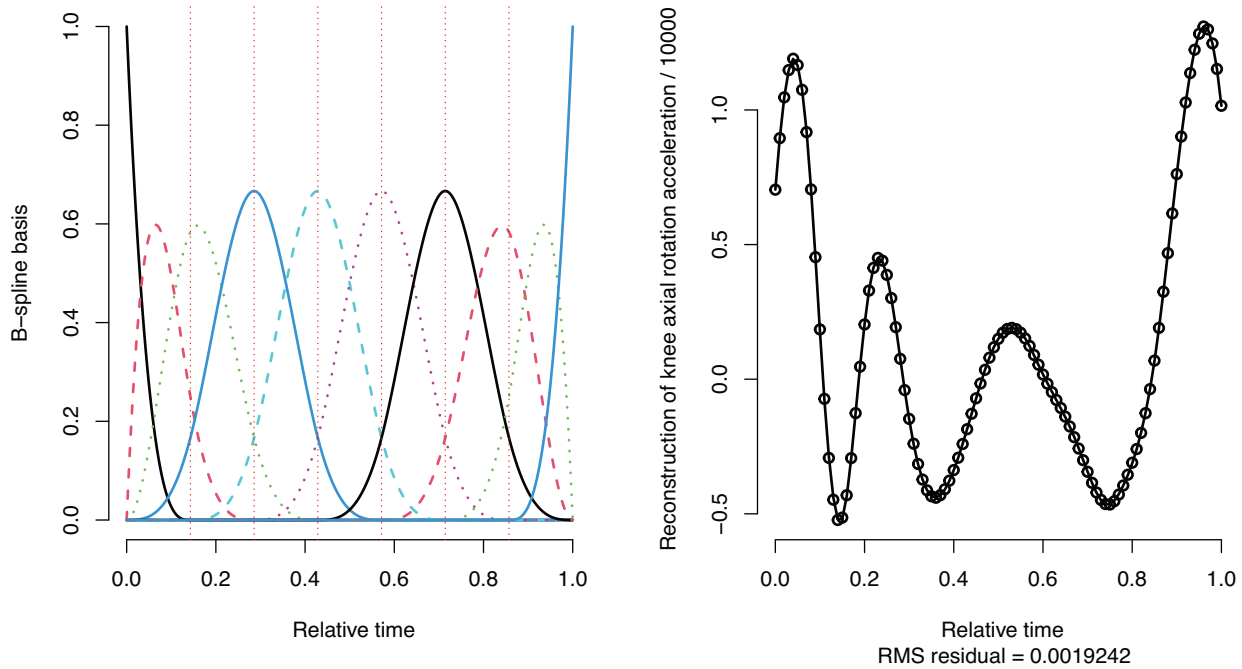


FIGURE 3 | B-spline basis with 10 knots (left); reconstruction (solid curve) of the first knee axial rotation acceleration profile (observed points as circles) with 30 B-spline basis functions (right).

(amplitude) of children as well as in the timing of growth spurts (phase). Further examples are functions over space such as, for example, diffusion tensor imaging measurements along different neuronal tracks in the brain (Greven et al. 2010), where there is variation both in the fractional anisotropy measurement as a proxy for neuronal health (amplitude), as well as variation in the brain location corresponding to a given t (phase), due to different proportions of anatomies and due to imprecisions in the measurement process. In our running example, phase variability is clearly visible, as maxima and minima in vertical knee acceleration are not aligned in Figures 1 and 2. This means that there is not only variation in the magnitude of acceleration between observations but also in the timing of the acceleration pattern.

Ignoring phase variation in an analysis can lead to confounding effects between amplitude and phase and to less interpretable results. Intuitively speaking, in FDA, we usually assume that the same argument t corresponds across functions—for example, corresponds to the same event in time or to the same location in space—and comparisons across functions per t are thus sensible. If there is phase variation, this is no longer the case. Depending on the context, it is thus often of interest to either (i) align functions before the analysis (of amplitude only) if phase variation is not of interest, which is also known as *registration* or *warping*, or (ii) to separate amplitude and phase variation and analyze them separately or jointly.

Most existing approaches focus on (i), the registration or warping problem. If we want to align the functions x_i , for example, to some overall mean μ , this can be formulated as finding suitable so-called *warping functions* γ_i that monotonically map \mathcal{T} onto \mathcal{T} , such that $x_i(\gamma_i(t))$ is optimally aligned to $\mu(t)$ for all $t \in \mathcal{T}$. The γ_i then capture different “speeds” (phase) of the different

functions x_i along the interval \mathcal{T} . Traditionally (e.g., Ramsay and Silverman 2005, Ch. 7), registration has often been performed by matching certain landmarks—for example, maxima, minima, zero crossings, or corresponding values for derivatives—across functions and linearly interpolating γ_i in between. This approach is simple and works reasonably well in the case of clearly defined and equal numbers of such landmarks. More sophisticated approaches take the whole function into account and often work with minimizing the \mathcal{L}^2 distance between one function x_j and a second aligned function x_i over γ , that is,

$$\inf_{\gamma} \|x_i \circ \gamma - x_j\|_{\mathcal{L}^2}^2 = \inf_{\gamma} \int_{\mathcal{T}} (x_i(\gamma(t)) - x_j(t))^2 dt. \quad (2)$$

In addition to not being symmetric, that is, warping x_i to x_j and warping x_j to x_i are not equivalent, this approach carries the so-called *pinching problem* (Marron et al. 2015). This means that the distance (2) can become small or zero even if x_i and x_j are not warped versions of each other, by using a warping function γ that compresses areas of \mathcal{T} , where x_i and x_j are dissimilar and expands those where they are close, resulting in spiky warping functions. Regularization or Bayesian approaches (e.g., Lu, Herbei, and Kurtek 2017; Matuk et al. 2022; Ramsay and Li 1998; Ramsay and Silverman 2005) restrict the amount of warping possible but can only partially solve this problem. Srivastava et al. (2010) thus proposed to work with the so-called *elastic distance* instead, the Fisher–Rao metric optimized over warping. This distance can be shown to simplify the \mathcal{L}^2 distance in (2) for the optimally aligned square-root-velocity transformed curves $Q(x_i)(t)$ and $Q(x_j)(t)$, which simplifies computations. Here, $Q(x_i)(t) = \dot{x}_i(t) / \sqrt{|\dot{x}_i(t)|}$ (if $\dot{x}_i(t) \neq 0$, else $Q(x_i)(t) = 0$) for the first derivative \dot{x} of x . The elastic distance constitutes a proper distance between functions modulo warping (and level due to the derivative) and avoids the pinching problem. It can thus

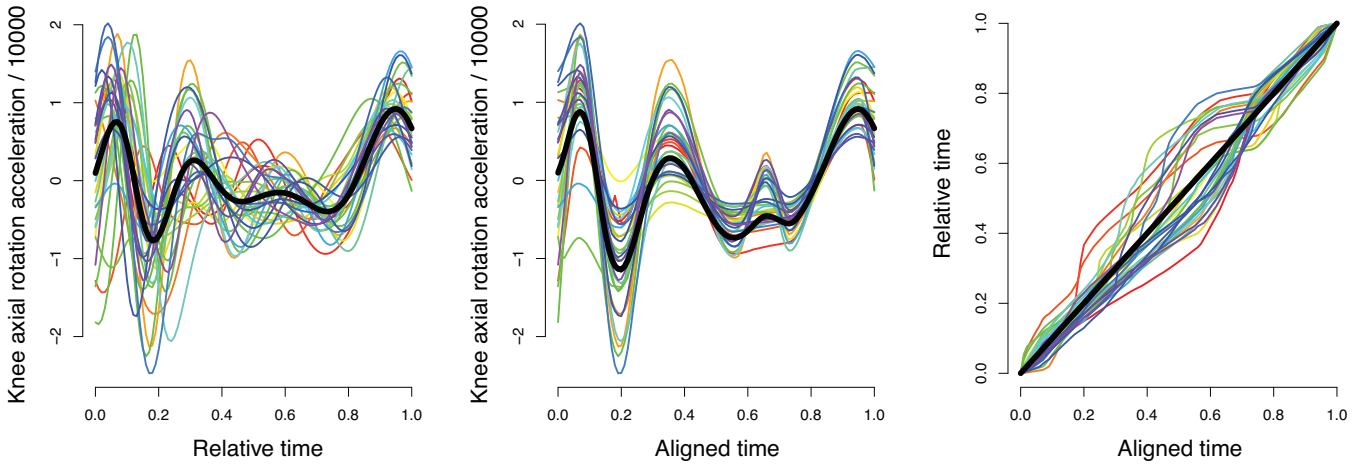


FIGURE 4 | Knee axial rotation acceleration profiles from Figure 1 (left) with their mean function in black (left); same functions iteratively aligned to their mean given at the end of the algorithm in black (middle); warping functions warping the original time to the aligned time (right).

be used to align functions or to obtain, for example, a mean estimate after alignment, by taking out all information about the phase. Approaches for computations with sparsely observed (multivariate) functional data have also been developed (Steyer, Stöcker, and Greven 2023a).

In some settings, phase is of interest as well. Some more recent approaches thus consider a joint analysis of both phase and amplitude, that is, point (ii) above, for example, in the context of FPCA (cf. Section 2.4) (Happ et al. 2019) or regression (cf. Section 3) (Hadjipantelis et al. 2014).

In our running example, there is a clear misalignment of maxima and minima between functions. As the number of maxima and minima can differ between functions, simple warping approaches based on landmarks do not work well here. Figure 4 shows the knee axial rotation acceleration profiles from Figure 1 (left) before and after elastic alignment using the `time_warping` function in R package `fdasrvf`, together with the warping functions transforming the time for alignment. Note that warping functions (up to numerics) are centered around the identity line. Plotted are also the means of the unaligned and aligned functions in black. After the alignment, the mean shows more pronounced peaks and valleys, as it is no longer averaging over functions with their maxima and minima at different time points, and thus better represents the functions in the data.

2.4 | Functional Principal Component Analysis

FPCA is an important tool in FDA as it allows dimension reduction of at least in theory infinite-dimensional functional data to manageable finite vectors of FPC scores. It is based on the Karhunen-Loève expansion (Karhunen 1947; Loève 1945), which allows us to expand a random function, that is, a stochastic process, X over \mathcal{T} as

$$X(t) = \mu(t) + \sum_{m=1}^{\infty} \xi_m \phi_m(t), \quad t \in \mathcal{T}, \quad (3)$$

where μ is the mean function of X and ϕ_m are orthonormal eigenfunctions of the covariance, with $\int_{\mathcal{T}} \phi_m(t) \phi_k(t) dt = 1$ iff $k = m$ and $= 0$ otherwise. In particular, according to Mercer's theorem (Mercer 1909),

$$\text{Cov}(X(s), X(t)) = \sum_{m=1}^{\infty} \nu_m \phi_m(s) \phi_m(t), \quad (4)$$

with decreasing eigenvalues $\nu_1 \geq \nu_2 \geq \dots \geq 0$. Furthermore, ξ_m are uncorrelated random scores with mean zero and variance ν_m , $m = 1, 2, \dots$, and are independent normal if X is a Gaussian process.

Truncating the sum in (3) at a finite upper limit M gives the best approximation to X with M basis functions in terms of mean-square prediction error (Rice and Silverman 1991). This allows for a parsimonious representation that approximates a sample X_i , $i = 1, \dots, n$, of independent replicates of X with finitely many functions ϕ_m common to all observations, and individual scores $\xi_{m,i}$ that capture the function-specific deviations from the overall mean. As the components have decreasing importance as measured by ν_m , the first few components often capture most of the relevant features in the data. Looking at the fraction $\sum_{m=1}^M \nu_m / \sum_{m=1}^{\infty} \nu_m$ allows a quantitative assessment of the (integrated) variance explained by the approximation, and a choice of M to explain, for example, 95% or 99% of the overall variance.

FPCA estimates μ , the $\xi_{m,i}$, ν_m and ϕ_m in practice from a sample of curves taken to be realizations x_i of $X_i(t) = \mu(t) + \sum_{m=1}^{\infty} \xi_{m,i} \phi_m(t)$, where each X_i is an independent copy of X given in (3). Usually, a smoothness assumption is incorporated for the estimation of the ϕ_m , which can be done in several ways. We here briefly describe an approach based on Yao, Müller, and Wang (2005) that accommodates sparsely observed functions with additional white noise errors ϵ_{it} and thus works relatively generally. As in this setting presmoothing the observed functions x_i or computing the scores using numerical integration based on $\xi_{m,i} = \int_{\mathcal{T}} (X_i(t) - \mu(t)) \phi_m(t) dt$ does not work well, the idea is to incorporate smoothing into the estimation of the covariance in (4). After estimation of $\mu(t)$ by $\hat{\mu}(t)$, and detrending $\tilde{x}_i(t) =$

$x_i(t) - \hat{\mu}(t)$, estimation is based on the cross-products of observed points within curves, $\bar{x}_i(t_{ij})\bar{x}_i(t_{ik})$, which are roughly unbiased estimates of $\text{Cov}(X(t_{ij}), X(t_{ik}))$. All cross-products are pooled and a smoothing method of choice is used for bivariate smoothing—local polynomial smoothing in Yao, Müller, and Wang (2005), penalized spline smoothing in the `refund` R package (Goldsmith et al. 2020), with further improvements for the particular case of covariance smoothing regarding speed and symmetry proposed for example, in Xiao, Li, and Ruppert (2013) and Cederbaum, Scheipl, and Greven (2018). If an additional error is assumed, the diagonal cross-products approximate $\text{Cov}(X(t_{ij}), X(t_{ij}))$ plus the error variance, the diagonal is thus left out for smoothing and the error variance estimated from the difference of the smooth diagonal and the average (possibly boundary-trimmed) diagonal cross-products.

Once the smooth covariance is available, the orthogonal decomposition (4) is typically done numerically on a fine grid using the usual matrix eigendecomposition. Eigenvectors are then rescaled to approximate orthonormality of the ϕ_m with respect to the functional \mathcal{L}^2 norm rather than the Euclidean vector norm. Finally, the scores $\xi_{m,i}$ are estimated. In the dense case, this could be based on $\xi_{m,i} = \int_{\mathcal{T}} (X_i(t) - \mu(t))\phi_m(t)dt$ with the estimated eigenfunctions and numerical integration. In the sparse case, where numerical integration based on only a few grid points does not give good results, it is preferable to work with conditional expectations (Yao, Müller, and Wang 2005) of the $\xi_{m,i}$ given the data. Another way of looking at this is viewing Equation (3) truncated to M eigenfunctions as a linear mixed model with random effects $\xi_{1,i}, \dots, \xi_{M,i}$, which can be assumed to be uncorrelated and thus have diagonal covariance matrix due to the Karhunen–Loève expansion (e.g., Greven et al. 2010; Goldsmith, Greven, and Crainiceanu 2013; Scheipl, Staicu, and Greven 2015). The eigenfunctions ϕ_m take the place of, for example, a constant function and a linear function for a random intercept random slope model, thus allowing for more modeling flexibility with data-driven basis functions. The conditional expectations then correspond to the usual best linear unbiased prediction of the random effects.

The estimated scores $\xi_i = (\xi_{1,i}, \dots, \xi_{M,i})^T$ give a multivariate summary vector of the individual functions that provides dimension reduction and enables many methods developed for multivariate data to be transferred to functional data. In addition, the estimated approximation $\hat{\mu}(t) + \sum_{m=1}^M \hat{\xi}_{m,i} \hat{\phi}_m(t)$ allows predicting the whole function even from sparse individual data by pooling of information across functions. Goldsmith, Greven, and Crainiceanu (2013) provide confidence bands for the individual functions based on this approach that additionally takes the uncertainty in the estimated eigendecomposition into account.

In our running example, the first four FPCs explain about 83% of the total variation of the first knee axial rotation acceleration profiles, as shown in Figure 5. The importance decreases quickly, with the fifth to seventh FPCs only explaining further 6%, 4%, and 2%, respectively. Positive scores for the first FPC mostly indicate more pronounced movements with higher peaks and lower valleys, and scores for this component tend to be slightly higher for females. The second FPC contrasts more pronounced movements at different times during the whole movement. Please note that FPCs are only unique up to sign, and FPCs and corresponding

scores could also be flipped compared to the output of function `fpca.face` in R package `refund` in case this allowed for simpler interpretation. We here rescaled the `fpca.face` output to ensure orthonormality of estimated eigenfunctions with respect to the \mathcal{L}^2 instead of the vector inner product, please see the code and data supplement or the accompanying GitHub repository for details. Also note that clearly, the eigenfunctions are picking up some of the phase variability in the data that we already saw in Figure 4, such that an alternative FPCA would first align the functions and then analyze either the main directions of variation in the aligned data or in the aligned data and warping functions jointly.

Given its centrality as a tool for dimension reduction and to obtain a parsimonious empirical basis, there are several extensions of FPCA to more complex settings. Multivariate FPCA extends FPCA to multivariate functional data. These could be several functions observed over \mathcal{T} such that the setting is that of vector-valued functional data (e.g., Chiou, Chen, and Yang 2014). Or it could be several functions observed potentially over different domains. Happ and Greven (2018) develop this setting, with an example of one function being a sparsely observed longitudinal trajectory over time and another being a 2D or 3D image, and provide a link between the univariate and multivariate FPCAs that simplifies estimation. Covariate-dependent FPCA is, for instance, considered by Ding et al. (2022). Di et al. (2009), Greven et al. (2010), and Shou et al. (2015) develop multilevel FPCA extensions for settings where functions are, for example, repeatedly observed within observational units such as subjects, and where an FPCA decomposition is desired of both the within-subject and the between-subject functional variation. In the case of correlated functional data, we cannot view the X_i as independent copies of (3). Extensions of FPCA have thus also been developed, for example, for functional time series and functions observed repeatedly over time (e.g., Chen and Müller 2012; Panaretos and Tavakoli 2013; Park and Staicu 2015), using a double decomposition in both the time dimension and the t dimension.

2.5 | Software

The Comprehensive R Archive Network (CRAN) (R Core Team 2023) contains a “Task View: Functional Data Analysis” listing most R packages related to FDA. We here name some of the packages implementing methods mentioned in the previous subsections.

For descriptives and outlier detection (Section 2.1), the functional boxplot of (Sun and Genton 2011) is implemented in the function `fbplot` in the R package `fda` (Ramsay, Graves, and Hooker 2024). The `rainbow` package (Shang and Hyndman 2024) can be used to obtain functional bagplots, boxplots, and rainbow plots (Hyndman and Shang 2010). `fdaoutlier` (Ojo, Lillo, and Fernandez Anta 2023) provides a collection of functions for functional data outlier detection.

For smoothing of curves (Section 2.2), many methods developed for scatterplot smoothing can be used to presmooth functions. The `fda` package provides functionality to expand observed functions in bases such as, for example, B-splines. A very flexible R package for smoothing in general is the `mgcv` package (Wood

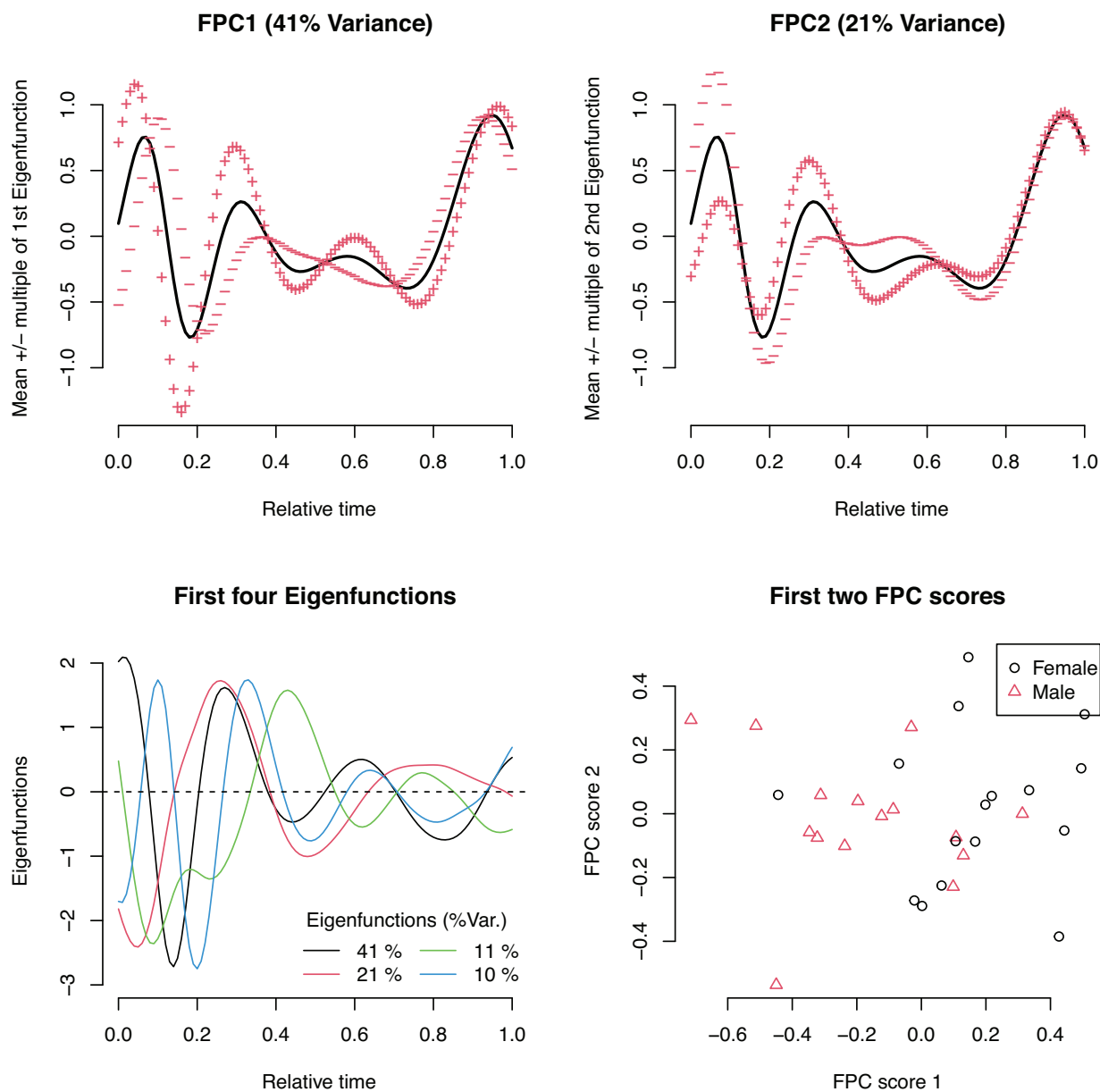


FIGURE 5 | First four FPCs in terms of the first four eigenfunctions (bottom left) with the percentage of explained variation in the knee axial rotation acceleration profiles from Figure 1; mean function plus/minus $\sqrt{v_1}$ times the first FPC/eigenfunction (top left), respectively, $\sqrt{v_2}$ times the second FPC/eigenfunction (top right); scatterplot of second against first FPC scores (bottom right). Note that $\sqrt{v_m}$ corresponds to the standard deviation of the m th score ξ_m , $m = 1, 2$.

2017). Further packages build smoothing into the analysis, such as, for example, into the estimation of covariances for FPCA as discussed below.

For warping/alignment of functions and separation of their amplitude and phase (Section 2.3), elastic options based on the square-root-velocity framework are implemented in R packages `fdasrvf` (Tucker 2024) for functions with the same number of sample points and in `elasdics` (Steyer 2024) for irregularly or sparsely observed (multivariate) functions. The `fda` package offers several alternative options based, for example, on landmarks.

For FPCA (Section 2.4) of univariate functional data, different packages offer options based on covariance smoothing and conditional expectations for the scores. Function `FPCA` in R package `fdapace` (Zhou et al. 2022) uses local polynomial smoothing, while several functions in R package `refund` (Goldsmith et al. 2020) use penalized spline smoothing, such as `fpca.sc` and `fpca.face` with the faster FACE method. `ccb.fpc` provides corrected confidence bands for the predicted functions incorporating uncertainty in estimated FPCs (Goldsmith, Greven, and Crainiceanu 2013). R package `MFPCA` (Happ-Kurz 2022) implements FPCA for multivariate functional data (Happ and Greven 2018).

3 | Functional Regression

3.1 | Introduction/Overview

We call a regression problem “functional regression” if at least one functional variable is found on the left- and/or right-hand side of the model equation. This means that functional variables may be the response, covariate(s), or both. To distinguish the different settings, we use the terms “scalar-on-function(s)” regression (SOFR), “function-on-scalar” regression (FOFR), and “function-on-function(s)” regression (FOFR), respectively. In what follows, we will first give an overview of those three settings. Then, we will describe two popular strategies for statistical modeling, inference, and prediction, namely semiparametric models using basis functions, and nonparametric approaches using kernels. We will focus on the former, semiparametric models and basis functions here because results are easier to interpret (from our point of view). In addition to the “statistical” approaches described in Section 3, also methods borrowed from and inspired by the machine learning (ML) community can be used, which will be discussed in more detail in Section 6.

3.1.1 | Scalar-on-Functions Regression

“Scalar-on-function(s)” regression corresponds to the case that the response is scalar, and among the covariates, there is at least one functional variable. In the simplest case, there is only a single covariate, which is functional. For instance, Figure 1 (left) showed knee axial rotation acceleration with colors corresponding to the maximum moment. For illustration, we may try to explain/predict the latter using the acceleration profiles. Then the simplest, but probably still the most popular model for SOFR is the so-called *functional linear model*

$$Y_i = \alpha + \int_{\mathcal{T}} x_i(t)\beta(t) dt + \epsilon_i, \quad i = 1, \dots, n, \quad (5)$$

where β is the so-called *coefficient function*, Y_i is the (scalar) response and x_i is the functional covariate observed for subject/unit i (such as the curves shown in Figure 1, left). As in the classical linear model with both scalar covariate(s) and response, it is typically assumed that x_i is set by the experimenter, or if the covariate is, in fact, a random variable, we take the “conditional view” in terms of modeling Y_i given $X_i = x_i$. With functional data, however, it is usually the latter. Assumptions for the error term ϵ_i are typically also the same as in the classical linear model. Thus, ϵ_i are assumed to be i.i.d. (scalar) random variables with mean zero and constant variance σ^2 , and sometimes the additional assumption of normality is made. The latter can, for instance, be relevant when using likelihood-based approaches for estimating unknown model parameters (compare Section 3.2.1 below). When interpreting the results after fitting model (5) to the data at hand, the coefficient function β is particularly important, analogously to the β -coefficient(s) in the standard linear model (see the examples below).

As pointed out above, the functional linear model (5) is the simplest model for SOFR. However, as in the scalar case, it can be seen as the starting point for numerous generalizations and extensions. The most obvious extension is including scalar

covariates Z_1, \dots, Z_q and multiple functional covariates X_1, \dots, X_p with domains $\mathcal{T}_1, \dots, \mathcal{T}_p$, respectively, leading to

$$Y_i = \alpha + \sum_{j=1}^p \int_{\mathcal{T}_j} x_{ij}(t)\beta_j(t) dt + \sum_{r=1}^q z_{ir}\gamma_r + \epsilon_i, \quad i = 1, \dots, n, \quad (6)$$

with further assumptions being analogous to (5). In the case of the data shown in Figure 1, for instance, we should at least take the person’s sex into account by including a corresponding factor as an additional scalar covariate. When fitting the corresponding model using function `pfr()` from R package `refund` (Goldsmith et al. 2020), the coefficient function for the functional covariate “knee axial rotation acceleration” as shown in Figure 6 (left) is obtained, together with approximate, pointwise 95% confidence intervals (shaded region). For modeling/fitting the coefficient function, we used a penalized, cubic spline with 15 (B-spline) basis functions and a penalty on the second-order differences of neighboring basis coefficients, a so-called *P-spline* (Marx and Eilers 1999). For details on fitting and software, see Sections 3.2.1 and 3.3, respectively. Further statistical inference, such as confidence intervals, is discussed in Section 4. Qualitatively speaking, the interpretation of $\beta(t)$ from Figure 6 (left) is as follows: (positive) values of knee acceleration at the beginning of the cycle have a positive effect on the maximum moment, whereas the association is negative at the end of the cycle, but uncertainty is very high in that area (as can be seen from the large confidence intervals). So, the main takeaway from this example is that people with large (positive) knee axial rotation acceleration at the beginning of the cycle tend to have a larger maximum moment than people with lower acceleration at the beginning of the cycle. This effect is also visible from the colors in Figure 1 (left). In general, however, such findings are not necessarily possible from descriptive plots only, particularly if more than one covariate is present.

If the error term ϵ_i in (6) is clearly non-Gaussian, for example, because the response is binary, model (6) can be generalized to a response variable with simple exponential-family distribution, analogously to the well-known generalized linear model (GLM). With functional covariates, we can define the linear predictor

$$\eta_i = \alpha + \sum_{j=1}^p \int_{\mathcal{T}_j} x_{ij}(t)\beta_j(t) dt + \sum_{r=1}^q z_{ir}\gamma_r, \quad i = 1, \dots, n. \quad (7)$$

Then, with $\mu_i = \mathbb{E}[Y_i | X_{i1} = x_{i1}, \dots, X_{ip} = x_{ip}, Z_{i1} = z_{i1}, \dots, Z_{iq} = z_{iq}]$ being the conditional mean of the response given the covariates, $\eta_i = g(\mu_i)$, and known link function $g(\cdot)$, the *generalized functional linear model*, or *functional generalized linear model* (both terms used in the literature) is obtained. Instead of the link function, the concrete model can be specified through the inverse link, the so-called response function $h = g^{-1}$, in terms of $\mu_i = h(\eta_i)$. Further extensions that relax the assumption of linearity in (5)–(7) will be discussed in more detail in Section 3.2. A review entirely on SOFR is found in Reiss et al. (2017). For an illustration of the functional GLM (7), we may try to identify the persons’ sex by looking at the hip acceleration profiles as shown in Figure 1 (right). The fitted coefficient function (again using `pfr()` from `refund`) in a functional logit model with response “sex” (female: 0, male: 1) is given in the right panel of Figure 6. Again, we used a P-spline approach with 15 basis functions and a second-order penalty on the basis coefficients. In analogy to the

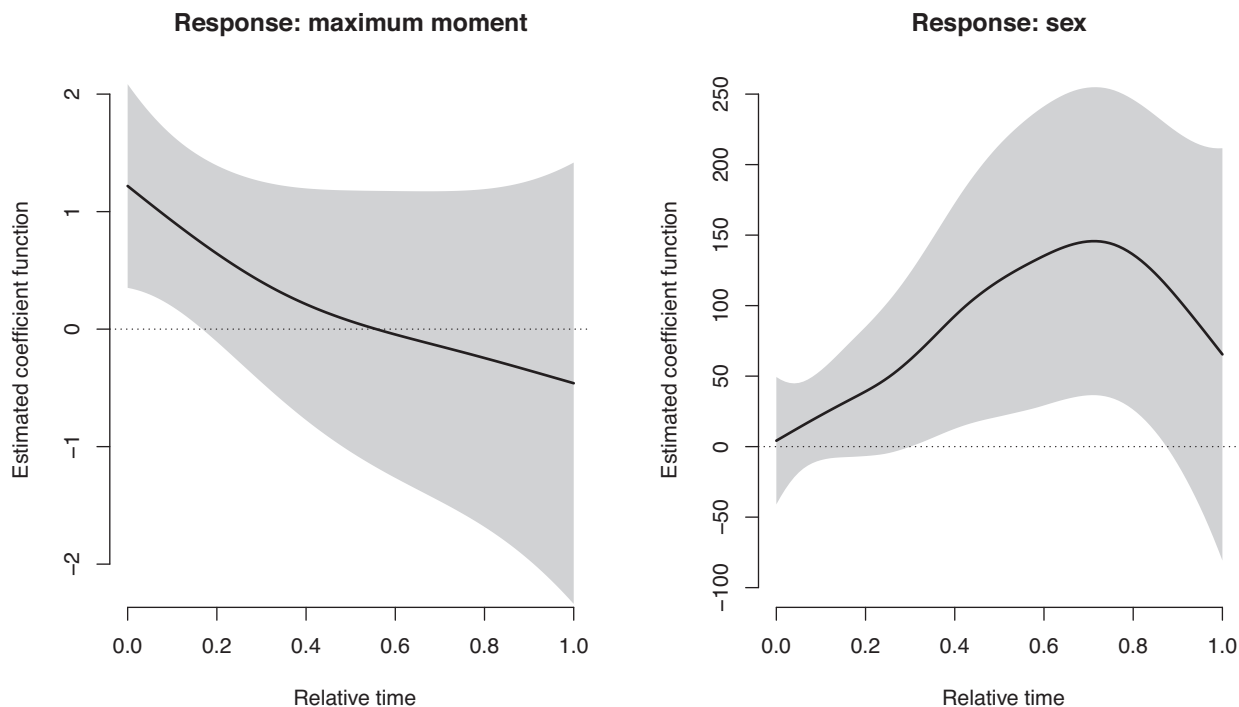


FIGURE 6 | The estimated coefficient function in the functional linear model (left) when predicting the maximum moment using the knee axial rotation acceleration profiles from Figure 1 (left) and additional covariate “sex,” together with approximate, pointwise 95% confidence intervals (shaded region). The estimated coefficient function in the functional logit model with response “sex” and functional covariate hip abduction-adduction acceleration from Figure 1 (right) is given in the right panel.

usual logit model, the functional version is defined by the link function

$$g(\pi_i) = \log \left\{ \frac{\pi_i}{1 - \pi_i} \right\},$$

with $\pi_i = \mu_i = P(Y_i = 1 | X_i = x_i)$ being the conditional probability of person i being male given the covariate curve (hip acceleration) x_i . Equivalently, the model is specified through the response function

$$\pi_i = h(\eta_i) = \frac{\exp(\eta_i)}{1 + \exp(\eta_i)}.$$

As seen in Figure 1 (right), males and females can be separated very well by looking at hip acceleration roughly between time points 0.7 and 0.8. Not surprisingly, the corresponding β -function shown in Figure 6 (right) also has its clear maximum in that area, indicating that subjects with large/positive hip acceleration in that area tend to be men, whereas small/negative values typically correspond to females.

It should be noted, though, that the functional data shown in Figure 1 is very smooth and measured on a dense and regular grid. In practice, however, functional data can be noisy and/or sparsely/irregularly sampled. In those cases, functional regression can be combined with (pre-)smoothing as discussed in Section 2 (see also Section 3.2), or even joint (Bayesian) modeling of both Y_i and $X_i(t)$ (McLean et al. 2013) to take uncertainty in the smooth underlying functions into account.

3.1.2 | Function-on-Scalar Regression

If the response $Y(t)$, $t \in \mathcal{T}$ is functional, but all covariates X_1, \dots, X_p are scalar, the simplest model is the *linear function-on-scalar* model

$$Y_i(t) = \alpha(t) + \sum_{j=1}^p x_{ij} \beta_j(t) + \epsilon_i(t), \quad i = 1, \dots, n, \quad t \in \mathcal{T}, \quad (8)$$

where $\alpha(t)$ and $\beta_j(t)$, $j = 1, \dots, p$ are coefficient functions and $\epsilon_i(t)$ is the error function drawn from a stochastic process with mean zero and covariance function $\Sigma(s, t)$, $s, t \in \mathcal{T}$; compare, for example, Chen, Goldsmith, and Ogden (2016). The latter assumption takes into account that measurements taken from the same individual i at different time points, that is, within-function errors, are typically correlated. This is important for proper statistical inference. Chen, Goldsmith, and Ogden (2016), for instance, apply a “pre-whitening” step before estimating unknown regression parameters from (8) above. Greven and Scheipl (2017), by contrast, decompose $\epsilon_i(t)$ into a smooth, subject-specific functional random effect $E_i(t)$ and remaining white noise error ϵ_{it} . For illustration, let us revisit the data from Figure 1 (right), but now treat hip acceleration as the (functional) response and sex as the (scalar) covariate with “female” as the reference category. Then, model (8) takes the simple form $Y_i(t) = \alpha(t) + x_i \beta(t) + \epsilon_i(t)$, where $\alpha(t)$ is the functional intercept and x_i is a dummy with $x_i = 1$ if subject i is male, and $x_i = 0$ if i is female. So $\alpha(t)$ (see the corresponding estimate in Figure 7, left) can also be interpreted here as the (functional) mean for females and $\beta(t)$ (estimate in Figure 7, right/black dashed) is the effect of being male. The error process $\epsilon_i(t)$ is decomposed into a smooth,

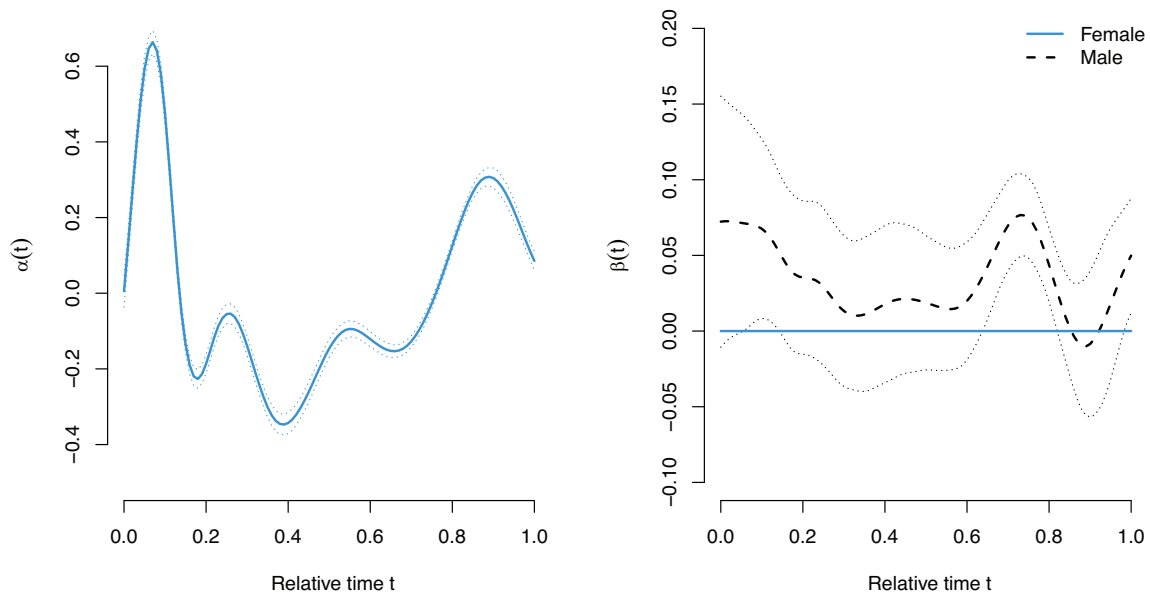


FIGURE 7 | Estimated parameter functions when regressing hip abduction-adduction acceleration from Figure 1 (right) on the person's sex using model (8); shown are the functional intercept (left) and the estimated effects (right) with colors corresponding to subjects' sex, and approximate, pointwise 95% confidence intervals (dotted).

subject-specific functional random effect plus white noise error here (compare Greven and Scheipl 2017). From Figure 7 (right), it is seen that the first peak in the (mean) acceleration profiles is somewhat higher and wider for males and that males' profiles tend to lie above the profiles of females around $t = 0.75$ (also compare Figure 1, right).

In general, besides (8), more complicated models may also be fitted in the semiparametric framework. For instance, Scheipl, Staicu, and Greven (2015) consider additive, smooth effects in terms of

$$Y_i(t) = \alpha(t) + \sum_{j=1}^p \gamma_j(x_{ij}, t) + \epsilon_i(t), \quad i = 1, \dots, n, \quad t \in \mathcal{T}. \quad (9)$$

This means that the effect of covariate X_j is allowed to be nonlinear and varying over t .

3.1.3 | Function-on-Function Regression

If both the response and the covariate(s) are functional and assumed to be observed over the same domain \mathcal{T} , a small modification of (8) gives the so-called functional linear *concurrent* model as

$$Y_i(t) = \alpha(t) + \sum_{j=1}^p x_{ij}(t)\beta_j(t) + \epsilon_i(t), \quad i = 1, \dots, n, \quad t \in \mathcal{T}. \quad (10)$$

As a modification of (6), with potentially different domains, we may introduce linear functional effects such that

$$Y_i(t) = \alpha(t) + \sum_{j=1}^p \int_{S_j} x_{ij}(s)\beta_j(s, t) ds + \epsilon_i(t), \quad i = 1, \dots, n, \quad (11)$$

where $t \in \mathcal{T}$, and S_j denotes the domain of the j th functional covariate. Similarly as done in (9) above, (11) may be relaxed to (Scheipl, Staicu, and Greven 2015)

$$Y_i(t) = \alpha(t) + \sum_{j=1}^p \int_{S_j} F_j(x_{ij}(s), s, t) ds + \epsilon_i(t), \quad i = 1, \dots, n, \quad (12)$$

although this approach requires the estimation of 3D smooths F_j and is thus less often used in practice.

In particular, if the response and the covariates have the same time domain \mathcal{T} , it can also be relevant to allow for more general integration limits than in (11). This allows only (part of) the past of the covariates to affect the current value $Y_i(t)$ but is more general than a concurrent effect as in (10). Common examples of such *historical* effects replace \int_{S_j} in (11) by \int_0^t or \int_{t-h}^t for some lag $h > 0$; see Brockhaus et al. (2017) and Rügamer et al. (2018) for a more detailed discussion of different possible specifications.

3.2 | Semi- and Fully Nonparametric Approaches

We will start by giving some further insight into semiparametric modeling of functional data by use of basis functions, as all models presented so far can be considered in this framework. Fully nonparametric approaches providing even larger flexibility in terms of the association structure between response and covariate(s) will be described in Subsection 3.2.2.

3.2.1 | Semiparametric Models and the Basis Functions Approach

The idea here is very similar to presmoothing of functional data as given in (1). The (unknown) coefficient functions, such as β

in (5), are expanded in basis functions as

$$\beta(t) = \sum_{k=1}^K \phi_k(t) b_k, \quad (13)$$

where ϕ_k could, for example, be B-spline basis functions as shown in Figure 3 or the eigenbasis obtained via FPCA (compare Section 2.4) of some functional data, typically the covariate functions; and b_k are unknown basis coefficients that need to be estimated from the data. The integrals found in Sections 3.1.1 and 3.1.3 are typically calculated/approximated through numerical integration. If considering the functional linear model (5), for instance, we have

$$\int_{\mathcal{T}} x_i(t) \beta(t) dt \approx \sum_{v=1}^V \Delta(t_v) x_i(t_v) \beta(t_v), \quad (14)$$

with a grid t_1, \dots, t_V of evaluation points in \mathcal{T} and suitable integration weights $\Delta(t_v)$. If evaluation points are equidistant and reasonably dense, the simplest (but most often used) approximation is, up to a proportionality factor, given by $\sum_{v=1}^V x(t_v) \beta(t_v)$. Thus, many methods, such as signal regression as considered by Marx and Eilers (1999), can be seen as functional regression. If choosing the basis functions approach in the (generalized) functional linear model (compare (5)–(7)), the functional model turns into a (generalized) linear model with the basis coefficients being the (unknown) regression coefficients that can be estimated using the usual methods for (generalized) linear models. If using basis functions like B-splines, however, typically, a large number of basis functions needs to be used to allow sufficient flexibility concerning the β -functions that can be fitted. With standard methods like least squares or maximum likelihood, the resulting coefficient functions then tend to over-fit the data while showing some erratic, nonsmooth behavior. A popular fix is adding a penalty on the integrated squared (second) derivative(s) of the β -function(s) or a difference penalty on neighboring basis coefficients to encourage smoothness; compare Section 2.2 and below. The observed functional data, such as $x_i(t)$ in (5), might be taken as they are if they are reasonably smooth (compare, e.g., Figure 6), or they may be presmoothed, for example, by use of pre-chosen basis functions, FPCA, or other techniques (compare Section 2). On the one hand, the use of FPCA has the additional advantage that models (5)–(7) then transform into a standard (generalized) linear model with FPCA scores as the covariates; compare, for example, Müller and Stadtmüller (2005) and Müller and Yao (2008). Since often a relatively small number of principal components suffices to capture the main information in the data, the number of unknown regression coefficients in the resulting functional model remains small as well, and further regularization is not necessary. Furthermore, regression on FPCA scores is convenient to generalize. For example, the scores can directly be used in an additive model where the influence of each score is allowed to be nonlinear (Müller and Yao 2008). On the other hand, however, “the principal component scores are computed independently from the response, in an unsupervised fashion, so there is no a priori reason to believe that these scores will correspond to the best dimensions for the regression problem” (Fan, James, and Radchenko 2015), and additionally, the discrete parameter K needs to be carefully chosen. Alternatives to additive regression on FPCA scores that induce nonlinear relationships, for instance, include functional

index models (Fan, James, and Radchenko 2015)

$$Y_i = \alpha + \sum_{j=1}^p g_j \left(\int_{\mathcal{T}_j} x_{ij}(t) \beta_j(t) dt \right) + \epsilon_i, \quad i = 1, \dots, n, \quad (15)$$

where both β - and g -functions are represented via basis functions and estimated in an iterative manner. Basis coefficients for β and g are alternately updated while fixing the other set of coefficients at their current values, cycling through those two steps until convergence. Apparently, (15) includes the functional linear model as a special case if all g -functions are set to be the identity. As before, appropriate penalties can or should be included when fitting the basis coefficients of the β and/or g -functions. Alternative estimators of the components of functional index models can be found, for example, in Ling and Vieu (2021).

Analogously to SOFR, basis/penalty approaches can be used for function-on-scalar regression (FOFR) and function-on-function regression (FOFR) as introduced in Sections 3.1.2 and 3.1.3. In the additive model (9) or the linear model (11) for FOFR, however, the unknown functions (γ_j and β_j , respectively) have two arguments. This means that the basis needs to be chosen appropriately, for example, as a so-called tensor product basis. If considering a single functional covariate in (11), for instance, Equation (13) then turns into

$$\beta(s, t) = \sum_{k=1}^{K_1} \sum_{l=1}^{K_2} \phi_k(s) \phi_l(t) b_{kl}, \quad (16)$$

and smoothing should typically be done in both s - and t -directions.

As already pointed out, if using a rich basis such as a large number of B-splines, smoothing is typically carried out by adding a penalty when fitting the basis coefficients (such as those in (13) and (16)), which prevents overfitting and ensures interpretability of the resulting coefficient functions. In the simplest case, for instance, the functional linear model (5) with reasonably smooth functional covariate curves that are observed on a dense and regular grid t_1, \dots, t_V , fitting can be done by minimizing the penalized quadratic loss

$$Q(\alpha, \mathbf{b}) = (\mathbf{y} - \alpha - \mathbf{X}\Phi\mathbf{b})^\top (\mathbf{y} - \alpha - \mathbf{X}\Phi\mathbf{b}) + \lambda \mathbf{b}^\top \mathbf{\Omega} \mathbf{b} \quad (17)$$

as a function of the basis coefficients $\mathbf{b} = (b_1, \dots, b_K)^\top$ and the intercept $\alpha = (\alpha, \dots, \alpha)^\top$. Here, \mathbf{X} is a $(n \times V)$ data matrix with the entry in row i and column v being $x_i(t_v)$, that is, $(\mathbf{X})_{iv} = x_i(t_v)$; $(V \times K)$ matrix Φ , with $(\Phi)_{vk} = \phi_k(t_v)$, contains the basis functions ϕ_1, \dots, ϕ_K evaluated at t_1, \dots, t_V , \mathbf{y} collects the observed response values, and $\mathbf{\Omega}$ is a penalty matrix specifying the concrete type of quadratic penalty imposed on \mathbf{b} . The strength of the penalty is controlled through tuning parameter λ . Simple matrix algebra shows that, up to integration weights, the i th entry of $\mathbf{X}\Phi\mathbf{b}$ corresponds to (14) for equidistant and dense grids. If using a P-spline approach (Marx and Eilers 1999) with the second-order penalty (as done in Figure 6), for instance, $\mathbf{\Omega}$ takes the form $\mathbf{\Omega} = \mathbf{D}_2^\top \mathbf{D}_2$, where \mathbf{D}_2 produces second-order differences in coefficients as $\mathbf{D}_2 \mathbf{b} = (b_3 - 2b_2 + b_1, \dots, b_K - 2b_{K-1} + b_{K-2})^\top$. In the case of a generalized model, such as (7), the quadratic loss $(\mathbf{y} - \alpha - \mathbf{X}\Phi\mathbf{b})^\top (\mathbf{y} - \alpha - \mathbf{X}\Phi\mathbf{b})$ in (17) is replaced by the negative

log-likelihood. If more than one (functional) covariate is to be considered (compare (6)), the design matrix \mathbf{X} as well as the matrix of basis functions Φ and vector of basis coefficients \mathbf{b} needs to be extended accordingly, and further penalty terms are added. Since suitable values for the associated penalty parameter(s) are typically unknown, those need to be chosen in a data-driven way. Besides generic approaches such as cross-validation or information criteria, a popular and smart way in the case of *quadratic* penalties (as they are typically used with functional data), is offered through the *mixed model* perspective. The latter means that quadratic smoothing penalties are equivalent to interpreting basis coefficients as random effects with normal (prior) distribution and implies a one-to-one correspondence of variance components and smoothing parameters. As a consequence, the latter can be estimated via maximum likelihood or restricted maximum likelihood; compare Wood (2011), Wood, Pya, and Säfken (2016), Wood (2017), and Scheipl, Staicu, and Greven (2015) and Scheipl, Gertheiss, and Greven (2016). In addition to the estimation of smoothing parameters, the mixed model perspective also offers tools for further statistical inference, such as testing and confidence intervals (compare Section 4).

An important point when choosing the final model is variable selection, particularly if the number of potential covariates is large. As before, we need to distinguish between scalar-on-function, function-on-scalar, and function-on-function regression. Most contributions in the literature, however, are on the first two cases, with a very popular approach being sparsity-inducing penalties; see, for example, Matsui and Konishi (2011), Gertheiss, Maity, and Staicu (2013), Feng, Zhang, and Tong (2022) for variable selection in the (generalized) linear model for scalar-on-functions regression, Fan, James, and Radchenko (2015) for the functional index model; and Lian (2013), Chen, Goldsmith, and Ogden (2016), Barber, Reimherr, and Schill (2017), Fan and Reimherr (2017), Parodi and Reimherr (2018) for the function-on-scalar case. Alternatives include functional lars (Cheng, Shi, and Eyre 2020), Bayesian methods (Goldsmith, Huang, and Crainiceanu 2014; Kowal and Bourgeois 2020), and boosting as discussed in Section 6. Also, more classic approaches such as forward/backward selection based on statistical testing as given in Section 4 could be used. A recent review of variable selection in functional regression models is provided by Aneiros, Novo, and Vieu (2022). Finally, variable selection with functional data is sometimes interpreted as selecting the most predictive design (grid) points (Ferraty, Hall, and Vieu 2010; Blanquero et al. 2019) or intervals/regions in the functions' domain (James, Wang, and Zhu 2009; Tutz and Gertheiss 2010; Zhou, Wang, and Wang 2013; Centofanti et al. 2020); which typically requires alternative methods to the basis functions and smoothing approach as described above. For further information, we refer to the literature cited above. The following subsection provides a short introduction to fully nonparametric functional regression.

3.2.2 | Nonparametric Approaches

Let us consider the regression problem with continuous Y_i and a single functional covariate X_i in its most general form, where

$$Y_i = f(X_i) + \epsilon_i, \quad (18)$$

f being an unknown regression function, and ϵ_i some mean zero noise variable, potentially with some further assumptions such as being i.i.d. across subjects $i = 1, \dots, n$. Compared to the semiparametric approaches above, in particular, the functional linear model (5), assumptions with respect to both the form of f and the distribution of ϵ_i are much milder.

For functional as for multivariate covariates, for a new observation with known covariate value x and unknown Y , a kernel-based, nonparametric prediction $\hat{Y} = \hat{f}(x)$ is given by

$$\hat{f}(x) = \frac{\sum_{i=1}^n Y_i K(d(X_i, x)/h_n)}{\sum_{i=1}^n K(d(X_i, x)/h_n)}, \quad (19)$$

with some kernel K , bandwidth $h_n \searrow 0$ (for $n \rightarrow \infty$), and distance measure d appropriate for the type of covariate considered, as further discussed below. From the mean zero error term, it follows that $\hat{f}(x)$ is an estimate of the conditional expectation $\mathbb{E}[Y|X = x]$. For \hat{f} as defined above to be consistent, a necessary condition, informally speaking, is that the probability of a training observation to be in the close neighborhood of x in terms of d is strictly positive. For details on and a discussion of this so-called *small ball probability*, see, for example, Ferraty and Vieu (2006) and Azaïs and Fort (2013). In general, each density function may be used as kernel K , but typically, we restrict ourselves to symmetric kernels with the maximum at zero, such as the (standard) normal density—the so-called *Gaussian* kernel. For details on kernel functions, see, for example, Gasser, Müller, and Mammitzsch (1985).

When dealing with functional covariates, in particular, the choice of d is crucial. A popular choice for functional $X_i, x \in \mathcal{L}^2$, for instance, is the \mathcal{L}^2 norm as introduced in Section 1.2,

$$d(X_i, x) = \|X_i - x\|_{\mathcal{L}^2}. \quad (20)$$

However, restricting d to be a metric such as (20) is sometimes too restrictive in the functional case. So-called *semi-metrics* may also be considered such as

$$d(X_i, x) = \|\dot{X}_i - \dot{x}\|_{\mathcal{L}^2}, \quad (21)$$

where \dot{X}_i, \dot{x} are (first) derivatives; see Ferraty and Vieu (2006) for a deeper insight into this topic. An important difference between metrics and semi-metrics is that in the latter case $d(X_i, x) = 0$ does not necessarily imply $X_i = x$. With (21), for example, $d(X_i, x) = 0$ is also obtained if $x(t) = X_i(t) + c$, for some vertical shift constant $c \neq 0$. In general, the choice of which (semi-)metric to take depends on the shape of the data and the goal of the statistical analysis. For instance, if dimension reduction for functional observations is of interest, one possibility is FPCA as discussed in Section 2.4, in which case options for d include distances on the extracted scores. In general, results can look very different, depending on the chosen measure of proximity. In Chapter 3 of Ferraty and Vieu (2006), examples to illustrate this effect are given. Also, further suggestions for semi-metrics and a survey on which semi-metric may be appropriate for which situation can be found there. For example, semi-metric (21), which is based on the derivatives, is often well suited for smooth data, whereas for rough data a different approach should be considered.

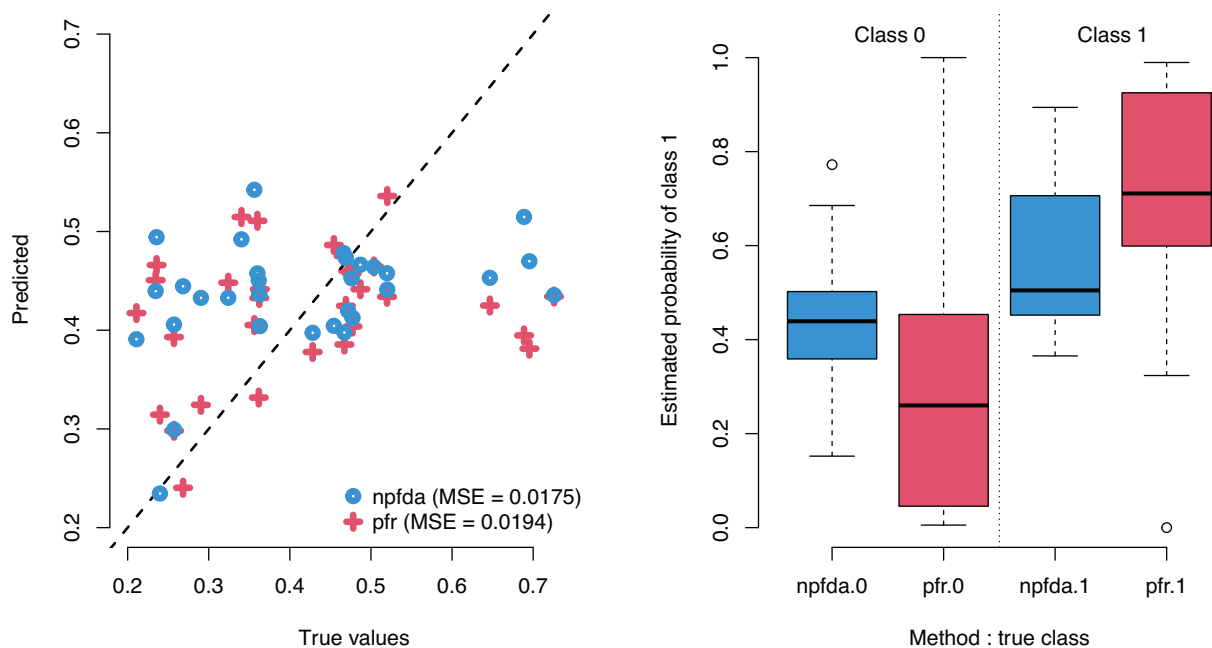


FIGURE 8 | Predicted values versus true values (left) and estimated probabilities of class 1 (i.e., “male”) versus true classes (right; 0 = “female” 1 = “male”) when using nonparametric FDA (“npfda”) or a (generalized) functional linear model fitted by `pfr()` on the data from the left and right of Figure 1, respectively.

For an illustration of nonparametric SOFR (19), consider the knee acceleration data from Figure 1 (left). We use nonparametric regression/prediction (19) with a semi-metric in terms of an (unweighted) Euclidean distance on the scores from an FPCA with five components. The (global) bandwidth is chosen via (inner) cross-validation. Figure 8 (blue circles, left) shows the predicted values obtained via leaving-one-out (outer) cross-validation versus the true values. For comparison, we also give the results for a functional linear model (fitted by `pfr()`) and applying leaving-one-out cross-validation as well. A perfect prediction would correspond to values on the bisecting (dashed black) line. We see that this is a rather difficult regression problem, with nonparametric FDA (“npfda”) performing slightly better in terms of the mean squared error (MSE), which indicates at least some nonlinearity in the association between the response and the covariate (although the larger MSE of “pfr” is mainly caused by the two extreme observations with true values around 0.7 that are fitted much better by “npfda”). A potential issue with nonparametric functional regression as introduced so far, however, is that it is restricted to cases with only one (functional) covariate. An approach how to extend (19) to settings with additional, functional, or scalar covariates (such as “sex”) is presented in Selk and Gertheiss (2023) and Selk (2024).

If using the nonparametric approach (19) on the data from Figure 1 (right) with a binary, 0/1-coded response (female: 0, male: 1), the conditional expectation/regression function $f(x)$ corresponds to the (conditional) probability of class 1 given input x . The corresponding estimates obtained by `npfda` and `pfr` are shown in Figure 8 (right). As we can see, the functional logit model from Section 3.1.1 performs much better here than the nonparametric approach. This can be explained as follows. As noted earlier (compare Figure 1, right), the two classes can be separated very well by focusing on hip acceleration between time points

0.7 and 0.8. The coefficient function fitted by `pfr()` is able to exploit this (compare Figure 6, right), whereas the distance/semi-metric d used here (FPCA) is not. The nonparametric approach might perform better, for instance, by introducing an appropriate weight function in (20). Examples of such functions, including an algorithm for estimating a weight function from the data at hand, are given by Chen, Reiss, and Tarpey (2014). An advantage of the nonparametric approach (19) is that it can be easily used for regression with a multiclass response as well. As this is closely related to the classification of functional data, we refer to Section 5.1 for more details. Furthermore, the kernel-based approach presented in Section 3.2.2 could be combined with the more structured models as presented in Section 3.1.1. For instance, Jeon, Park, and Van Keilegom (2021) proposed a very general approach for additive regression with nonstandard response and covariates, including functional data.

3.3 | Software

Some functions for functional regression analysis, where either the dependent variable or one or more independent variables are functional, are available in R package `fda` (Ramsay and Silverman 2005; Ramsay, Hooker, and Graves 2009); in particular, for the functional linear model for SOFR (6) and the concurrent model (10). A broader class of models, including the models presented in Sections 3.1.1–3.1.3, can be fit by use of `refund` (Goldsmith et al. 2020). The functions implemented there typically use `mgcv` methodology (Wood 2011; Wood, Pya, and Säfken 2016; Wood 2017), and the latter R package can also be used directly to fit at least some of those models, but the use of `refund` may often be easier for the inexperienced user. R functions for nonparametric FDA as presented in Section 3.2.2 and thoroughly discussed in the monograph by Ferraty and Vieu (2006) are found

at the accompanying website <https://www.math.univ-toulouse.fr/~ferraty/SOFTWARES/NPFDa/npfda-routinesR.txt>.

4 | Statistical Inference With Functional Data

4.1 | Functional Analysis of Variance

The setting of one-way univariate functional analysis of variance (ANOVA) is similar to the well-known scalar case. We typically have a grouping variable, a so-called factor, defining groups $1, \dots, g$, and functional data $X_{ij}(t)$, $t \in \mathcal{T}$, coming from these g groups, $j = 1, \dots, g$, $i = 1, \dots, n_j$, with n_j denoting the sample size in group j . The overall sample size is $n = n_1 + \dots + n_g$. Given group membership, functional variables are assumed to be (conditionally) independent. Furthermore, it is typically assumed that functions are square-integrable on \mathcal{T} , that is, elements of $\mathcal{L}^2(\mathcal{T})$ (compare Section 1.2). The null hypothesis to test is

$$H_0 : \mu_1(t) = \dots = \mu_g(t), \quad t \in \mathcal{T}, \quad (22)$$

with $\mu_j(t)$ denoting the mean function in group j ; compare, for example, Zhang (2013). Alternatively, we can use a linear function-on-scalar model with one categorical covariate in terms of

$$X_{ij}(t) = \mu(t) + \alpha_j(t) + \varepsilon_{ij}(t), \quad (23)$$

where $\mu(t)$ is the grand mean (function), $\alpha_j(t) = \mu_j(t) - \mu(t)$ the deviation from the grand mean in group j , and $\varepsilon_{ij}(t)$ a mean zero error/noise process. Furthermore, it is typically assumed that all functions $X_{ij}(t)$, and thus $\varepsilon_{ij}(t)$, share the same covariance function $\Sigma(s, t)$, $s, t \in \mathcal{T}$. This is the functional version of homoscedasticity as commonly assumed in the scalar case (see below for the literature on testing the equality of covariance operators between groups). For identifiability, a constraint such as $\sum_j \alpha_j(t) = 0$ for all t is needed. Then, hypothesis (22) can be rewritten as

$$H_0 : \alpha_1(t) = \dots = \alpha_g(t) = 0, \quad t \in \mathcal{T}. \quad (24)$$

An example is given in Figure 9. Here, the data (dotted lines), a subset of our running example data, comes from $g = 3$ groups defined by different experimental conditions: run at 3 m/s with no backpack (“slowbw”), with 10% body mass backpack (“slowten”), with 20% body mass backpack (“slowtwe”). The empirical, groupwise mean functions are given as solid lines. For simplicity, we restrict ourselves to a small, illustrative dataset comprising 15 males, with five persons in each group.

In what follows, we will give a short overview of some procedures proposed for testing (22) and (24), respectively. Many of these tests are based on the pointwise between-groups sum of squares

$$\text{SSR}(t) = \sum_{j=1}^g n_j (\hat{\mu}_j(t) - \hat{\mu}(t))^2 \quad (25)$$

and the pointwise within-groups/error sum of squares

$$\text{SSE}(t) = \sum_{j=1}^g \sum_{i=1}^{n_j} (X_{ij}(t) - \hat{\mu}_j(t))^2, \quad (26)$$

compare Ramsay and Silverman (2005) and Górecki and Smaga (2019). One way of constructing a test statistic then is the ratio of the integrated between- and within-groups sum of squares, that is

$$\frac{\frac{1}{g-1} \int_{\mathcal{T}} \text{SSR}(t) dt}{\frac{1}{n-g} \int_{\mathcal{T}} \text{SSE}(t) dt}, \quad (27)$$

and the tests proposed in the literature vary concerning the method used for representing the functional data and estimating the unknown parameters, and the way p -values are calculated. For instance, under some assumptions such as Gaussian error processes, Shen and Faraway (2004) derived the test statistics’ null distribution and also provided an approximation based on the Karhunen–Loève expansion of the error process. Zhang (2011) proposed an improved, biased-reduced approximation, and Górecki and Smaga (2015) used a finite-dimensional basis function representation of the observable functional data $X_{ij}(t)$ as the starting point.

As an alternative to (27), we may also calculate the pointwise F-statistic,

$$F(t) = \frac{\frac{1}{g-1} \text{SSR}(t)}{\frac{1}{n-g} \text{SSE}(t)} \quad (28)$$

and integrate/globalize this quantity in terms of $\int_{\mathcal{T}} F(t) dt$ (Zhang and Liang 2014) or use only its maximum, that is, $\sup_{t \in \mathcal{T}} F(t)$ (Zhang et al. 2019).

A somewhat different approach is presented in Cuevas, Febrero, and Fraiman (2004), who used the sum of pairwise \mathcal{L}^2 -distances $\sum_{j < l} \int_{\mathcal{T}} (\hat{\mu}_j(t) - \hat{\mu}_l(t))^2 dt$ as the test statistic and a parametric bootstrap to approximate the null distribution. Cuesta-Albertos and Febrero-Bande (2010), by contrast, considered testing through random projects of the functional data observed. Specifically, a trajectory v is sampled on a fine grid, for example, through a sequence of partial sums of independent normal variables (i.e., a Gaussian random walk). Then, the functional data are projected as $P_{ij} = \int_{\mathcal{T}} X_{ij}(t)v(t) dt$, and a test is performed for mean differences of the projected data between groups. This procedure is repeated k times, that is, for k different v -trajectories, and resulting p -values are corrected using the FDR approach of Benjamini and Yekutieli (2001). For testing mean differences in the (scalar) projections P_{ij} , there are several options, for instance, the approach by Brunner, Dette, and Munk (1997)

For illustration, Figure 10 shows the p -values for five different tests for functional ANOVA on the data shown in Figure 9. Computations were done using the R package `fdanova` (Górecki and Smaga 2019), which provides about a dozen tests for testing (22). Here, we considered a permutation test employing the ratio of the integrated sums of squares (27) and basis representation of the functional data (Górecki and Smaga 2015), globalizing the pointwise F-statistic (28) (Zhang and Liang 2014), using its maximum only (Zhang et al. 2019), \mathcal{L}^2 -distances between groupwise means (Cuevas, Febrero, and Fraiman 2004), and random projections (Cuesta-Albertos and Febrero-Bande 2010). Figure 10 shows large agreement in terms of significant differences between

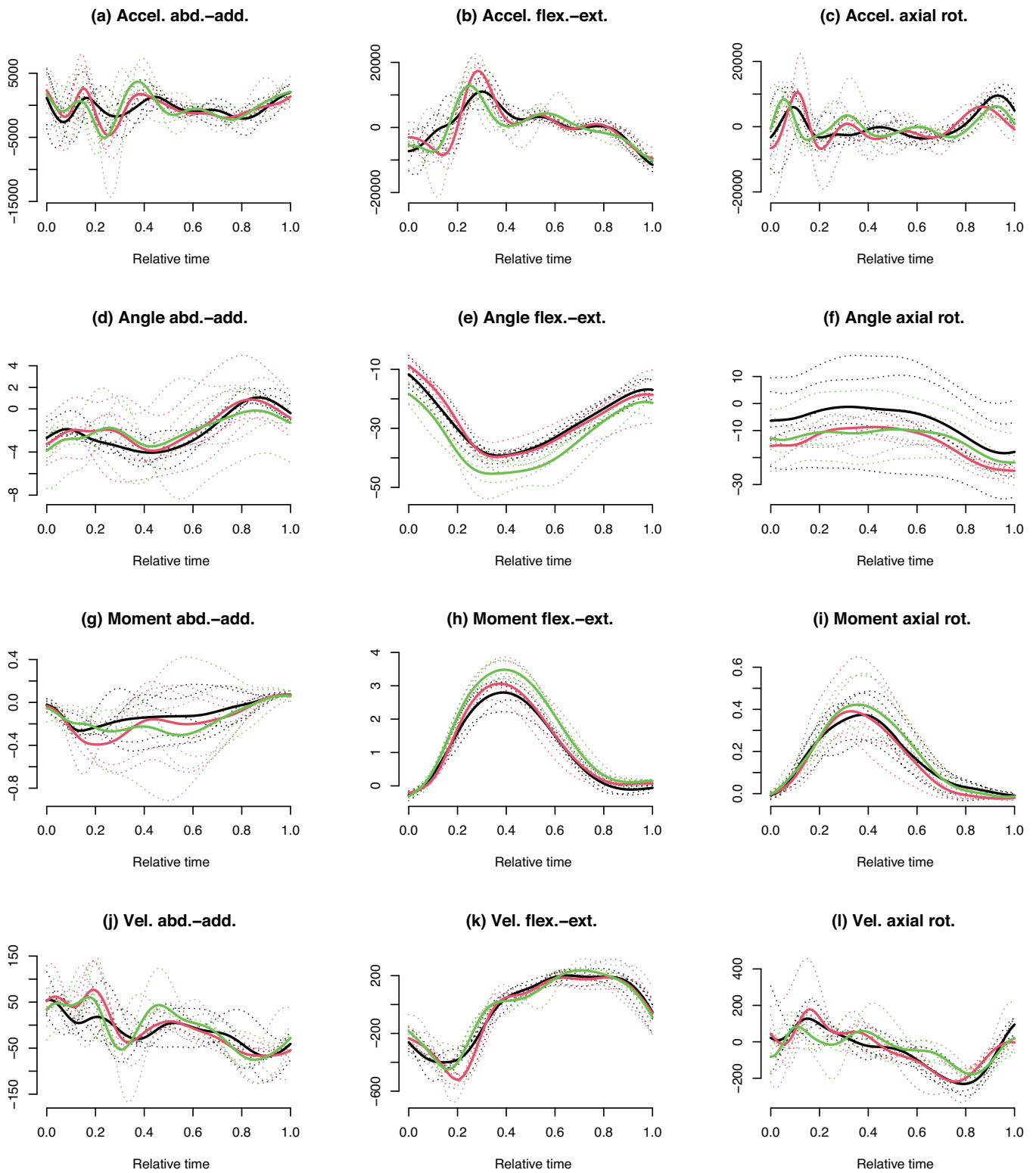


FIGURE 9 | Example knee measurements (dotted lines), compare Table 1, for three experimental conditions: “slowbw” (black), “slowten” (red), and “slowtwe” (green); the solid curves give groupwise means.

groups (“yes” or “no” according to the dashed/dotted lines in Figure 10). In particular, significant differences are detected between the three experimental conditions for (e) “knee flexion-extension angle” and (h) “knee flexion-extension moment”. The different tests, however, are obviously not equivalent. For example, the projection-based approach also identifies significant

differences (at the 5% level) in the mean functions of (c) “knee axial rotation acceleration” and (j) “knee abduction-adduction velocity,” whereas the other tests do not. In light of these findings, and keeping in mind that a large number of tests is available for testing (22)—even many more than considered in Figure 10; see Górecki and Smaga (2019)—it should be pointed out that

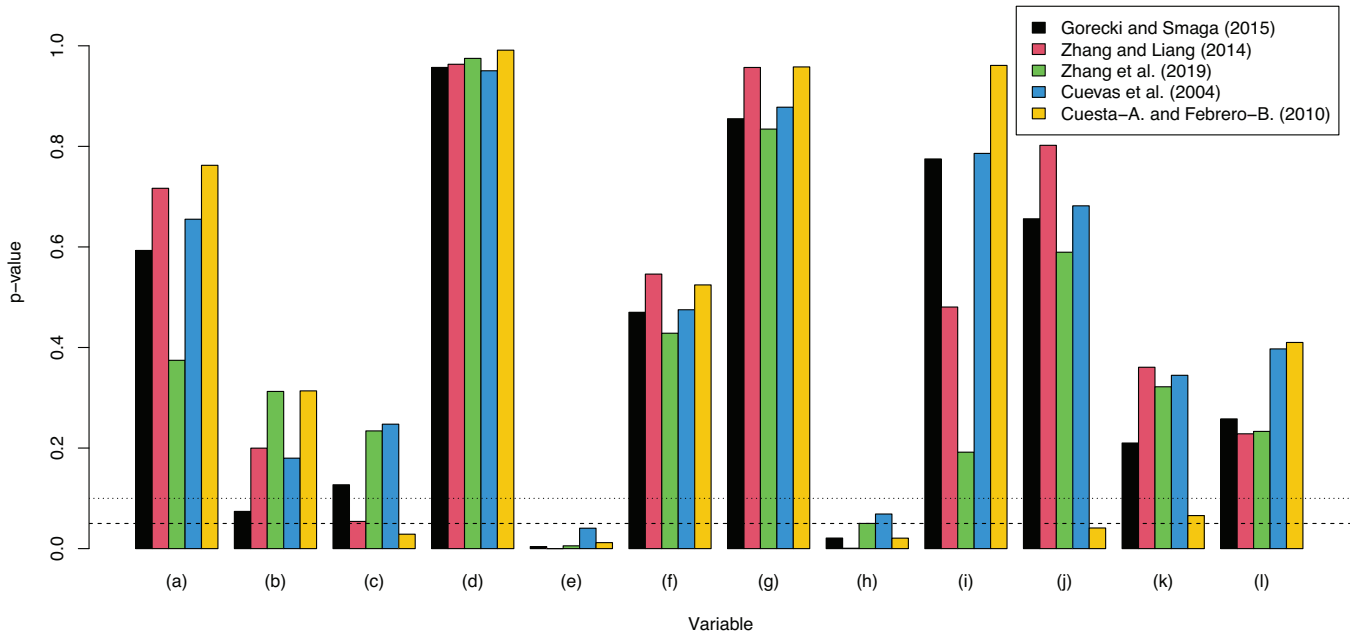


FIGURE 10 | p -Values for five different tests for functional ANOVA on the data shown in Figure 9; significance on the 5% or 10% level is indicated by the dashed and dotted lines, respectively.

running all these tests on the data, and reporting significant differences if any test has a p -value below α , is not admissible, since it may substantially increase the type-I error rate. For illustration, a corresponding resampling study is available as part of the code and data supplement and on [GitHub](#). An obvious solution to this problem is to decide on the test to use *before* analyzing the data. But even if we do, there might still be issues, particularly if we choose to use the approach based on random projections (Cuesta-Albertos and Febrero-Bande 2010). Due to the randomness involved, it can be very difficult to replicate the results obtained, even on the exact same data. For instance, if running the test on (c) “knee axial rotation acceleration” (Figure 9) a hundred times, we obtained p -values smaller than 5% (as seen in Figure 10, yellow) only in about 10% of all runs. In addition, this provides another easy way of “cheating” to researchers who are merely “hunting statistical significance” (Szucs 2016; Simmons, Nelson, and Simonsohn 2011): they may just run the test a few times (e.g., with a sequence of some popular seeds for the random number generator such as “1234,” “2345,” etc. for “reproducibility”) until the desired result appears. Again, this leads to an inflated type-I error rate. Consequently, although it is tempting, researchers should refrain from such behavior.

An important special case of (22) is the two-sample problem, where just two groups, such as case/control or treatment/control, are to be compared; and many testing procedures have been proposed for such situations; see, for example, Ramsay and Silverman (2005), Zhang, Liang, and Xiao (2010), Horváth, Kokoszka, and Reeder (2013) and Ghiglietti, Ieva, and Paganoni (2017). Post hoc comparisons can, for instance, be made through multiple contrast tests recently proposed by Munko et al. (2023). Also, the globalized pointwise F and F_{\max} -test can be used (Smaga and Zhang 2019). Besides mean functions, the distributions of functional data from different groups may also be compared in

a broader sense (e.g., Benko, Härdle, and Kneip 2009; Pomann, Staicu, and Ghosh 2016; Pini and Vantini 2016; Krzysko and Smaga 2021; Wynne and Duncan 2022). Furthermore, there are proposals targeting specifically the covariance functions (e.g., Kraus and Panaretos 2012; Fremdt et al. 2013; Paparoditis and Sapatinas 2016; Guo, Zhou, and Zhang 2018, 2019; Qiu et al. 2024). An entire textbook on inference for functional data is, for instance, provided by Horváth and Kokoszka (2012).

The comparison of mean functions as in testing (22) can be generalized in (at least) two ways. First, we may consider more than one factor in multiway (functional) ANOVA; see, e.g., Cuesta-Albertos and Febrero-Bande (2010). Second, the functional variables to compare between groups may be multivariate, which yields a functional version of multivariate analysis of variance (e.g., Górecki and Smaga 2017).

4.2 | Testing and Confidence Intervals in Functional Regression

In the framework of FOSR, various testing procedures can be found in the literature. For instance, after rewriting the linear model (11) as $Y_i(t) = \mathbf{x}_i^T \boldsymbol{\beta}(t) + \epsilon_i(t)$, \mathcal{L}^2 -norm based (Zhang and Chen 2007) and F-tests (Shen and Faraway 2004; Zhang 2011) have been proposed for testing the general linear hypothesis $H_0 : \mathbf{C}\boldsymbol{\beta}(t) = \mathbf{c}(t)$, where \mathbf{C} is a given matrix of appropriate dimension, and $\mathbf{c}(t)$ is a corresponding vector of given functions. Functional ANOVA, as in (22), can be seen as a special case of this more general testing problem.

The opposite setting is SOFR, where a scalar response is regressed on a functional covariate. In its most general form (18), only a few assumptions are made, while particularly in the functional linear model (5), assumptions are much stronger. The most

interesting hypothesis to test in this setting is typically whether there is an effect of the covariate on the response at all. That means, in (18), the null hypothesis is that the regression function f is constant. In the linear model (5), this translates into the coefficient function being zero. Also, if comparing (5) and (18), an interesting hypothesis to test could be whether the linear model (5) holds. For both problems, a few different testing procedures have been proposed. For instance, Swihart, Goldsmith, and Crainiceanu (2014) and McLean, Hooker, and Ruppert (2015) used the mixed model perspective of functional regression and proposed restricted likelihood ratio tests for checking the significance of a functional covariate and/or linearity of the effect. Kong, Staicu, and Maity (2016) extended four tests common in classical regression (Wald, score, likelihood ratio, and F-tests) to the functional linear model by use of FPCA, intending to test the null hypothesis that there is no association between a scalar response and a functional covariate. A Wald-type test is also discussed in Su, Di, and Hsu (2017), and Yi, Li, and Tang (2022) present an improved F-test. García-Portugués, González-Manteiga, and Febrero-Bande (2014) proposed a goodness-of-fit test based on random projections for the functional linear model with scalar response that can also be used for testing the special case of no association between the functional covariate and the scalar response. A thorough comparison of at least some of those tests is found in Tekbudak et al. (2019).

In addition to the fitted coefficient functions in a (generalized) functional linear model, Figure 6 also provides pointwise 95% confidence intervals as shaded regions. Those were estimated using the `pfr()` function from `refund`, which uses `mgcv` methodology (compare Section 3.3) and are based on the Bayesian view of the smoothing process. For simplicity, let us consider the basis coefficients \mathbf{b} from (17) where only a single functional covariate was present. Given the data and smoothing parameter, the posterior distribution of \mathbf{b} is multivariate normal with some covariance matrix $\mathbf{\Gamma}_b$ (for details on the form of $\mathbf{\Gamma}_b$, see, e.g., Wood 2017). Then, with $\hat{\boldsymbol{\beta}} = (\hat{\beta}(t_1), \dots, \hat{\beta}(t_V))^T = \Phi \mathbf{b}$ denoting the vector that contains the estimated coefficient function at the evaluation points, an approximate pointwise $(1 - \alpha)100\%$ credible interval for $\hat{\beta}(t_v)$ is given by $\hat{\beta}(t_v) \pm z_{1-\alpha/2} \sqrt{\zeta_v}$, where $\boldsymbol{\zeta} = (\zeta_1, \dots, \zeta_V)$ is the diagonal of $\Phi \mathbf{\Gamma}_b \Phi^T$ and $z_{1-\alpha/2}$ is the $(1 - \alpha/2)$ quantile of the standard normal distribution. These intervals also have “surprisingly good frequentist coverage properties” (averaged over t) (Wood 2017). Since $z_{0.975} = 1.96 \approx 2$, $\hat{\beta}(t_v) \pm 2\sqrt{\zeta_v}$ gives an approximate pointwise 95% confidence interval. When calculating confidence intervals in FOSR or FOFR models, it is important to make sure that within-function correlation is accounted for (compare Section 3.1.2). This may, for instance, be done in a conditional model through functional random effects (Scheipl, Staicu, and Greven 2015) or a marginal model through robust standard errors (Chen et al. 2013; Gertheiss et al. 2015).

While pointwise confidence intervals (only) work “pointwise” in terms of coverage properties, so-called simultaneous confidence bands can be interpreted globally. That means the provided bands cover the *entire* (true) function with some prespecified probability such as 95%. Most methods proposed for simultaneous confidence bands for functional data use resampling techniques, particularly different versions of the bootstrap (e.g., Degras 2011; Chang, Lin, and Ogden 2017). Some approaches also use dimension

reduction through FPCA and the Karhunen–Loève expansion; for example, Goldsmith, Greven, and Crainiceanu (2013) and Choi and Reimherr (2018). In a very recent paper, Liebl and Reimherr (2023) present a broad framework based on random process theory. Among other things, the confidence bands proposed also provide some “local” control of the false positive rate over subsets of the bands’ domain.

4.3 | Software

A collection of various tests for functional ANOVA is found in the R package `fdANOVA` (Górecki and Smaga 2019), and all p -values shown in Figure 10 were calculated through the `fanova.tests()` function provided there. The R packages `multiFANOVA` (Ditzhaus et al. 2023) and `fdatest` (Pini and Vantini 2022) implement the multiple contrast tests by Munko et al. (2023) and the interval testing procedure by Pini and Vantini (2016), respectively. Pointwise confidence intervals for terms in SOFR, FOSR, and FOFR models are, for instance, provided by `refund` functions `pfr()` and `pfr()`, respectively. Both are wrappers for `mgcv`’s `gam()` and its siblings (compare Section 3.3).

5 | Classification and Clustering

The purpose of regression, as presented in Section 3, was to explain and/or predict a scalar or functional response variable using one or more scalar/functional explanatory variables. Further approaches with origin in the ML community and a clear focus on prediction will be discussed in Section 6. In Section 5, we will present two further topics beyond regression for functional data from a statistical perspective, namely classification, and clustering. We will start with classification, which shows close links to Section 3.

5.1 | Classification of Functional Data

We assume that functional data comes from distinct classes $1, \dots, G$, and training data with known class labels is available for determining a classification rule such that new (functional) data with unknown group membership can be assigned to its underlying class with a low probability of an error. According to Bayes’ rule, each unit should be classified based on posterior class probabilities, that is, the conditional probabilities of each class given the (functional) data. In practice, however, those conditional probabilities are unknown and need to be estimated from the training data. In the simplest case, functional data only come from two classes. Then, we are in a situation like Figure 6 (right), where we can use the functional logit model to estimate conditional class probabilities and assign each unit to the class of the highest (estimated) probability.

If $G > 2$, the semiparametric approach, in particular the functional logit model, can be generalized in terms of a functional multinomial model (Matsui 2014), increasing, however, the number of parameter functions that need to be estimated. The nonparametric, kernel-based approach as presented in Section 3.2.2, on the other hand, can be applied directly. Specifically, following the idea of nonparametric regression (19), we can estimate the

conditional probability $P_g(x) := P(Y = g|x)$ of class g given a (functional) covariate x with unknown class label Y by

$$\hat{P}_g(x) = \frac{\sum_{i=1}^n I_g(Y_i)K(d(X_i, x)/h_n)}{\sum_{i=1}^n K(d(X_i, x)/h_n)}. \quad (29)$$

Here $I_g(Y_i)$ denotes the indicator function, which equals one if $Y_i = g$ and zero otherwise. As an alternative to (29), we could also use a functional k -nearest-neighbor (kNN) approach where we determine the k nearest neighbors of the new curve x to be classified according to d and assign the new curve to the class that is most frequent among its neighbors. Although specific assumptions need to hold for the functional kNN classifier to be consistent (C erou and Guyader 2006), it is often a useful, simple-to-apply benchmark in functional classification. By using a specific semi-metric, we can also focus on specific features of the functional data when classifying them. By combining several kNN predictions in terms of an ensemble, we can even use several semi-metrics simultaneously and select relevant features (Fuchs, Gertheiss, and Tutz 2015).

Besides the semi- and nonparametric approaches described so far, many further methods could be used to classify functional data. For instance, there is also a functional version of linear discriminant analysis, which in some situations produces “near perfect classification” (Delaigle and Hall 2012); for further information on the latter phenomenon, see also Berrendero, Cuevas, and Torrecilla (2018). Furthermore, after dimension reduction via FPCA, the obtained scores may be used as input to any statistical or ML algorithm designed for the classification of multivariate data, for example, a random forest or a neural net. Also, functional data may be used directly in some situations, for example, as input to a convolutional neural network (CNN). Please refer to Section 6 for details. A recent and more in-depth review of functional data classification is provided by Wang, Huang, and Cao (2023).

5.2 | Clustering

In Section 5.1 above, functional data came from distinct classes $1, \dots, G$, which were known for the training data and unknown for the test data. The purpose was to classify the test data using a classification rule that was learned from the training data. Since there are known class labels for the training data, this type of learning problem is also called *supervised* learning. If those classes are not given but are to be built using the (functional) data only, we call this process “clustering,” which is a so-called *unsupervised* learning problem. Further supervised and unsupervised learning techniques for functional data from a ML perspective are discussed in Section 6. The basic idea of clustering is that data are grouped such that those groups, the “clusters,” are homogeneous in some sense. A survey on and review of functional data clustering is, for instance, provided by Jacques and Preda (2014a) and Zhang and Parnell (2023). Here, we will only give a short and not necessarily exhaustive overview.

In general, most clustering methods (not only functional clustering) can roughly be classified as either “data/observation-based” or “model-based”. In the latter case, it is assumed that the data observed are generated by some underlying stochastic mechanism. Then, observations can, for example, be clustered according

to conditional class probabilities. Data-based methods, by contrast, make no distributional assumptions but start directly from the data. Common strategies involve hierarchical clustering and k -means (or variants thereof, such as k -median or k -medoids). Hierarchical clustering algorithms are typically distance-based and either successively group similar objects into clusters and fuse clusters (“agglomerative” hierarchical clustering) or subsequently split larger into smaller clusters (“divisive” hierarchical clustering). The concrete form of the algorithm results from the chosen distance measures for objects and clusters. In both cases, a “hierarchy” of partitions is created from which clusters might be chosen. For details on hierarchical clustering, see, for example, Murtagh and Contreras (2012). As an alternative, we may try to optimize some criterion across all potential partitions into k clusters. The latter is done with k -means, etc.; and a couple of different algorithms have been proposed for doing so (MacQueen 1967; Hartigan and Wong 1979; to name two of the most prominent references). For an overview of clustering algorithms (including categorizations apart from data vs. model-based), see, for example, Ezugwu et al. (2021) and references therein.

With functional data, we suggest classifying clustering algorithms according to the *two* dimensions “data/observation-based” versus “model-based” and “raw data methods” versus “dimension reduction” because any combination may be possible. For an alternative categorization into clustering methods in a finite-dimensional versus infinite-dimensional space, see Zhang and Parnell (2023). In the case of raw data methods, curves are clustered on the basis of their evaluation points. Within the “dimension reduction” category, we may further distinguish between “filtering methods” and “adaptive methods” (Jacques and Preda 2014a). Filtering methods first approximate the functions by using a set of basis functions and then use the basis coefficients for clustering. Adaptive methods, by contrast, carry out dimension reduction and clustering simultaneously, which may involve that also the type and number of basis functions may depend on the cluster. Model-based functional clustering typically assumes a mixture model that is based on some basis expansion. For instance, James and Sugar (2003) use a spline basis and assume that basis coefficients follow cluster-specific Gaussian distributions. Bouveyron and Jacques (2011) and Jacques and Preda (2013) allow for cluster-specific eigendecompositions (FPCA). Similarly to usual, model-based clustering for multivariate data, various specifications for the underlying Gaussian distributions are possible, leading to different numbers of unknown parameters that need to be estimated, for example, using an expectation-maximization (EM) algorithm; compare Bouveyron and Jacques (2011) for details. A Bayesian version of model-based clustering for functional data is, for instance, used in Heard, Holmes, and Stephens (2006). Extensions to multivariate functional data are, for example, provided by Jacques and Preda (2014b) and Schmutz et al. (2020). More generally speaking, all those approaches correspond to the “adaptive methods” from above and have to be distinguished from the much simpler, two-stage approach of calculating principal component scores on the entire dataset first and then using those values as input to an algorithm for model-based clustering of usual, multivariate data (which would be a “filtering method”). Analogously, distances between functions may be computed/defined via a basis expansion and corresponding basis coefficients, and then be used as input to hierarchical clustering or k -means, which can be described as

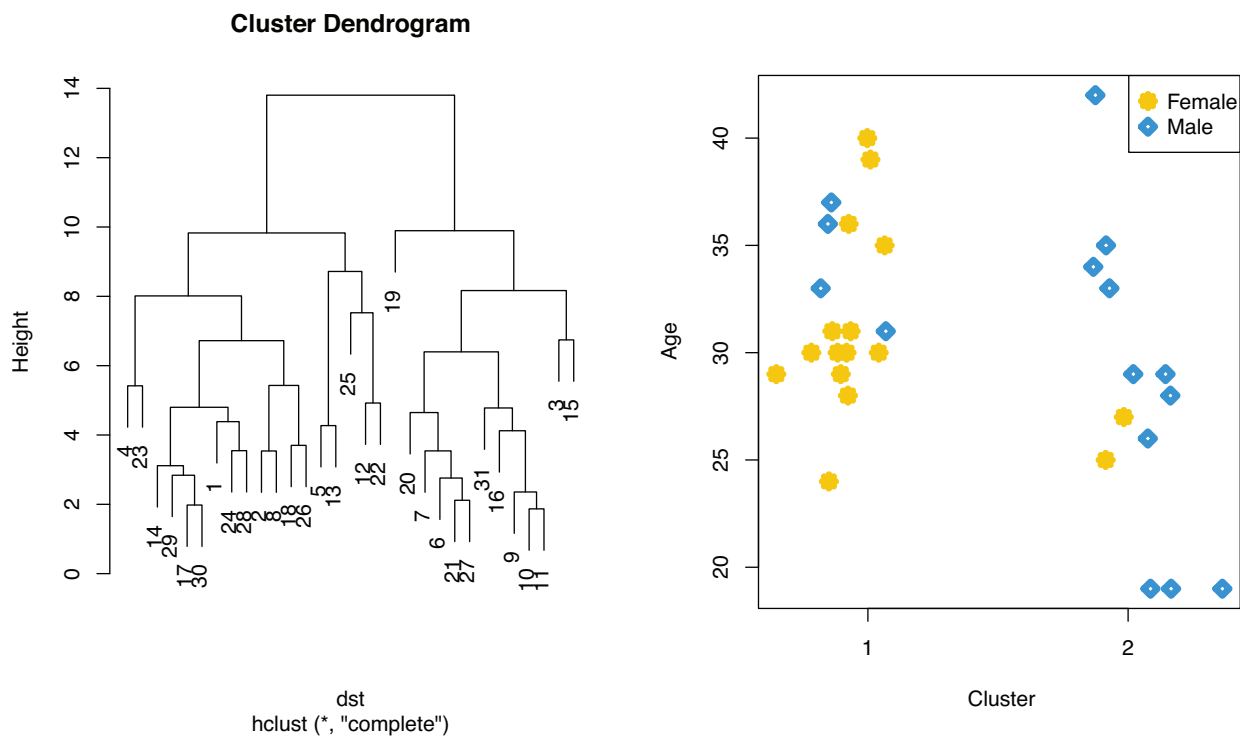


FIGURE 11 | Results for hierarchical clustering with complete linkage of functional data from Figure 1 (left), with dendrogram (left) and visualization of the 2-cluster solution by age and sex (right).

“dimension reduction in combination with data/observation-based clustering.” By contrast, if computing distances directly from raw data, we would end up with a raw, observation-based functional clustering. A discussion of k-means, dimension reduction (particularly FPCA), and specific cases of equivalence between approaches is found in Tarpey and Kinateder (2003). Of course, in addition to dimension reduction, there may be other preprocessing steps before calculating distances between functions; see, for example, Ieva et al. (2013).

For illustration, let us consider the curves from Figure 1 (left). To start with, we use the Euclidean distance on the raw data (compare Section 3.2.2) as input to hierarchical clustering with complete linkage. The resulting dendrogram is shown in Figure 11 (left), where we can nicely see how curves are successively grouped (bottom-up) into clusters. From the dendrogram, we also get the impression that the dataset essentially consists of two clusters. Those two clusters in the space of the first two principal components are found in Figure 12 (left). The distribution of age and sex in the clusters is shown in Figure 11 (right). We see that females are primarily found in cluster 1, whereas cluster 2 mainly consists of males. On average, people in cluster 1 are a little bit older, but the overlap of clusters is much larger for age than sex. Instead of using the raw data, we could also carry out FPCA first. Then, we could use the component scores as input to a standard clustering algorithm for multivariate data, such as hierarchical clustering or k-means. The latter approach minimizes the so-called *within-cluster sum of squares* across all potential partitions into k clusters. With $k = 2$, we obtain the clusters as shown in Figure 12. If looking at the scores of the first two principal components (Figure 12, left), we see that clusters are primarily determined by the first

principal component. The first eigenfunction is depicted in pink in Figure 12 (right), together with the functional data observed and the clustering. In particular, in the plot’s left part, we see how small (i.e., negative) values in the first principal component translate into cluster 2 membership, as cluster 2 mainly consists of curves with a first peak shifted to the right (dashed black). Also, it should be noted that the clusters that are obtained with k-means of FPCA scores (Figure 12) are precisely the same as those obtained with hierarchical clustering/complete linkage on the raw data above (Figure 11). A minimal change in the clusters is observed if using the model-based, adaptive approach by Bouveyron and Jacques (2011), where different FPCA solutions are allowed between clusters. Then, one observation (female, age 24), as marked in blue in Figure 12, moves from cluster 1 into cluster 2. This curve is somewhat peculiar regarding the beginning of the cycle; compare the blue line in Figure 12 (right). Otherwise, however, clustering results are very consistent across different algorithms here.

5.3 | Software

After dimension reduction (FPCA, spline basis, etc.) using already discussed functions, for example, from R packages `fda` or `refund`, estimated component scores or basis coefficients can simply be used as input to standard cluster algorithms such as `hclust()` or `kmeans()` from R’s basic `stats` package (R Core Team 2023); compare Figure 12. Also functions from the `cluster` (Maechler et al. 2021) package can be used. If algorithms work on distance/similarity matrices, the latter may also be computed on raw data, for example, using the `dist()` function in R, as has been done above (compare Figure 11). Model-based

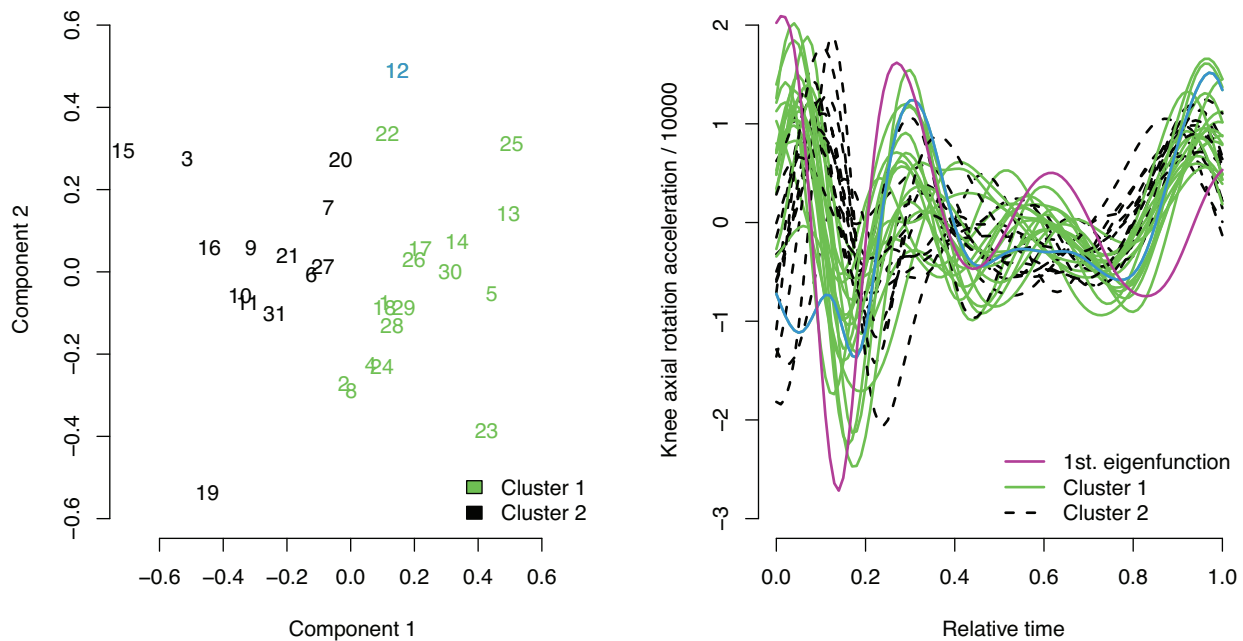


FIGURE 12 | Results for k-means clustering using principal component scores of functional data from Figure 1 (left), with clusters in the space of the first two principal components (left) and visualization of the 2-cluster solution in the functions' space together with the first eigenfunction (right). Note that this gives the same result as hierarchical clustering with complete linkage using the raw data (compare Figure 11, left). The blue color indicates the one curve (12) that moves from cluster 1 into cluster 2 if using model-based, adaptive clustering.

clustering for multivariate data is, for instance, provided by Scrucca et al. (2016) through R-package `mc1ust`. A functional version as employed above (Bouveyron and Jacques 2011; Schmutz et al. 2020) is available in `funHDDC` (Schmutz, Jacques, and Bouveyron 2021).

6 | Machine Learning Approaches

FDA can also be found in machine learning (ML) and its deep learning (DL) subcategory. We here mainly focus on models for supervised learning. In contrast to previously discussed methods, analyses in ML and DL mainly target predictive performance and the incorporation of functional data primarily serves as an appropriate data representation, rather than to allow for an interpretable relationship of (functional) model in- and outputs. We will also outline a few approaches that deal with unsupervised learning or generative approaches and are related to unsupervised approaches described in Section 5.

6.1 | Implicit Approaches

Many different ML approaches only implicitly deal with functional data and do not explicitly account for the functional nature of variables, for example, the smoothness of the inputs. Instead, functional covariates are considered as sequences or (multivariate) time series (see, e.g., Ott et al. 2021). Another common practice is to calculate summary statistics of the functional inputs (the mean of the function, location of the mode, different frequency statistics, etc.) to characterize inputs using scalar values instead of using the actual functions (e.g., Vogel et al. 2022). These statistics can subsequently be used with

classical ML approaches intended for scalar data inputs and preprocessing steps turning functional into scalar values are defined as a (tunable) preprocessing pipeline. Pfisterer et al. (2019) compare such classical ML approaches with functional classification algorithms described in Section 3.1.1. An alternative representation of functional inputs in, for example, the field of biomechanics, is a picture representation of multivariate functions (Liew et al. 2021). Given all functional inputs have the same domain \mathcal{T} , the approach concatenates the p functional observations $(x_{i1}(t_v), \dots, x_{ip}(t_v))$ for $v = 1, \dots, V$ to a $V \times p$ dimensional matrix for every observation i , which is then considered as an “image” in the subsequent analysis. This allows the use of pretrained models (Liew et al. 2024) and is based on the idea of transfer learning (see, e.g., Pan and Yang 2010). For functional outcomes, ML methods are often referred to as trajectory regression (Ott et al. 2021) or sequence-to-sequence models (Sutskever, Vinyals, and Le 2014). Most approaches do not impose smoothness of the output function, but either adjust the loss function to favor smooth(er) outputs or impose regularization to foster generalization and thereby implicitly generate smoother output functions.

In contrast to the previously described implicit approaches, there is also a body of literature in ML and DL that explicitly takes the functional nature of objects into account. We describe several methods in the next subsection.

6.2 | Classical Machine Learning for Functional Data

Among the classic supervised ML approaches, Rossi and Villa (2006) are one of the first to propose an explicit approach for

functional data by adapting support vector machines (SVMs) for FD. SVMs are kernel-based methods, but in contrast to the use of kernels in Section 3.2.2, SVMs utilize kernels $K(\cdot, \cdot)$ as a distance measure between two observations. Rossi and Villa (2006) argue that, although kernels in principle also work for functions in a Hilbert space \mathcal{H} , it is difficult to implement calculations in practice and these kernels do not take advantage of the functional nature of the data. The authors derive a simple solution by defining a map \mathcal{P} from \mathcal{H} to the Euclidean space. In this space classic kernels $K(\cdot, \cdot)$ such as the Gaussian or exponential kernel can be straightforwardly implemented and then used as $K(\mathcal{P}(f), \mathcal{P}(g))$ to evaluate the distance between two functions $f, g \in \mathcal{H}$. Another kernel-based ML approach is Gaussian processes (GPs). An adaptation of GPs called Gaussian process functional regression (GPFR) for functional data has been proposed in Shi and Choi (2011) with implementation in Konzen, Cheng, and Shi (2021). GPFR additively combines the mean estimation of a functional regression model with a multivariate GP. The linear (functional) effects of scalar or functional covariates thereby define the mean of the functional response, while functional covariate(s) over \mathcal{T} or the time information is used for the estimation of the covariance using the GP. Another important class of ML models is boosting. Gradient boosting approaches for functional regression models and various extensions have been proposed by Brockhaus et al. (2015, 2017) and Rügamer et al. (2018). By not relying on regression trees but on single additive predictors, the corresponding boosting implementation (Brockhaus, Rügamer, and Greven 2020) can be seen as a special coordinatewise gradient descent optimization routine. The Gauss–Southwell-type update rule in this case guarantees convergence (Schulte and Rügamer 2024) and additionally induces sparsity due to the coordinatewise nature of the optimization. This, in particular, allows fitting functional regression models in very high dimensions as the optimization routine only optimizes the effect of one (functional) feature at a time. Other extensions of this approach have been proposed recently, including densities-on-scalar regression models using a Bayes Hilbert space representation (Maier et al. 2021) or functional additive models on manifolds of planar shape and form (Stöcker, Steyer, and Greven 2023). Finally, tree-based methods have also been brought forward, both for functional covariates using random forests (Möller, Tutz, and Gertheiss 2016; Rahman et al. 2019) and also via functional random forests for functional outcomes (Fu, Dai, and Liang 2021).

In addition to these supervised learning approaches, several unsupervised methods with explicit function representation exist. The most prominent example is functional data clustering (Tarpey and Kinader 2003) as explained in detail in Section 5.2. An overview can be found in Jacques and Preda (2014a).

6.3 | Deep Learning and Neural Networks for Functional Data

Rossi, Conan-Guez, and Fleuret (2002) are among the first to combine neural networks with functional data by proposing a multilayer perceptron (MLP) for functional data and also extending the universal approximation theorem of neural networks (Hornik, Stinchcombe, and White 1989) to functional inputs. An

MLP for scalar values is defined by multiple neurons gathered in layers that are in turn stacked onto each other. Every neuron (or unit) can be seen as a function $h : \mathbb{R} \rightarrow \mathbb{R}, x \mapsto \tau(b + wx)$ that takes a scalar value x and computes a hidden state value of the form $\tau(b + wx)$ with bias (intercept) b , weight (coefficient) w , and activation (response) function τ . Similar to SOFR, a functional analog for functional inputs x can be defined by $\tau(b + \int w(t)x(t)dt)$ and approximated at given measurement points by finding a suitable approximation for $w(t)$ and the integral over t . This will result in units $h : \mathcal{H} \rightarrow \mathbb{R}$ that take functions as input and output a scalar value. Thus, the MLP is only functional in its first layer, while the following layers consist of conventional layers and neurons. In recent years, various other authors have proposed extensions and unifications (Guss 2016; Wang et al. 2019; Thind, Multani, and Cao 2023; Rao and Reimherr 2023). For functional response variables, the functional MLP can be extended in an analogous manner to the extension from SOFR to FOFR. The k th output neuron $h_k^{(l)}$ of a functional layer $l = 1, \dots, L$ in a functional MLP with L layers, each with J_l output neurons, is recursively defined by its previous layers as

$$h_k^{(l)}(t) = \tau^{(l)}\left(b_k^{(l)}(t) + \sum_{j=1}^{J_{l-1}} \int w_{j,k}^{(l)}(s, t) h_j^{(l-1)}(s) ds\right), \quad (30)$$

where $b_k^{(l)} \in \mathcal{L}^2(\mathcal{T})$ and $w_{j,k}^{(l)} \in \mathcal{L}^2(\mathcal{T} \times \mathcal{T})$. In addition, special roles are given to the input layer by defining $h_j^{(0)}(\cdot) = x_j(\cdot)$ for all covariates $j = 1, \dots, p$, and the very last layer consisting of only one unit $J_L = 1$ defined as $h^{(L)}(t) = \mathbb{E}(Y(t)|\mathbf{X})$, where \mathbf{X} collects all p functional covariates. Similar to the mean function, it is also possible to learn the covariance kernel using deep neural networks (Sarkar and Panaretos 2022). An extension to neural networks that model functional data both with structured predictors and arbitrary deep neural network architectures has been recently proposed by Rügamer et al. (2024).

In un- or self-supervised learning, one of the more prominent techniques is autoencoders (AEs; Kingma and Welling 2013), which try to encode a high-dimensional input into a small number of hidden variables and then decode these variables back to the original input size with the goal to recover the original input. A recent adaption of AEs is given by Hsieh et al. (2021), proposing a functional AE. The proposed approach can be seen as a nonlinear extension of FPCA. A similar method is proposed in Shi et al. (2024). Another method that can be seen as an unsupervised technique is function registration, which has recently been combined with deep neural networks (Chen and Srivastava 2021).

6.4 | Benchmark Comparison

To demonstrate the performance of different methods presented in this section, we conduct a small benchmark comparison between selected methods with available software implementations. We want to emphasize that all ML and DL methods are not tuned but either used with default settings or trained with values from preliminary experiments. We use two implicit neural network approaches as proposed in Liew et al. (2021) that represent functional covariates as images. One network is

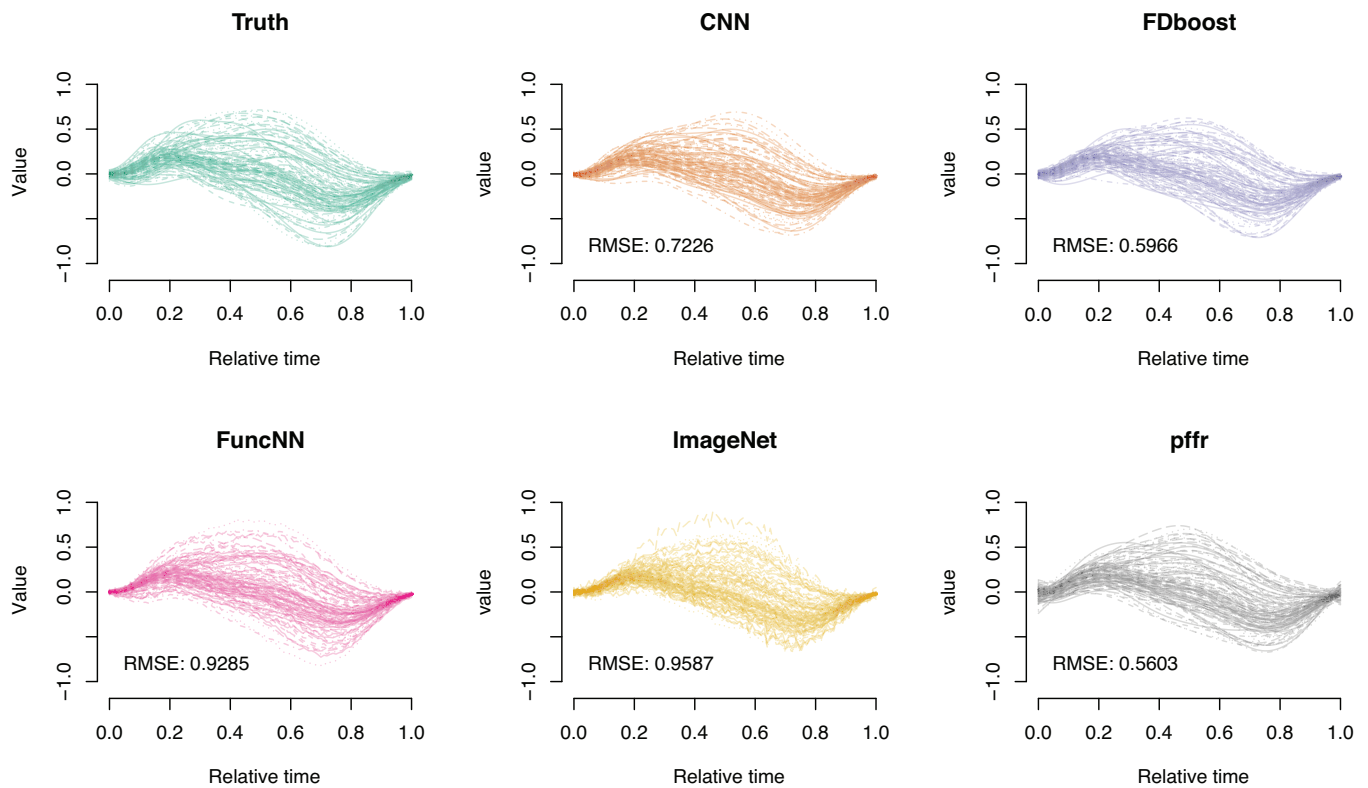


FIGURE 13 | Comparison of different ML and DL methods (facets) to fit a function-on-function relationship on the running data with respective test RMSE values for every method in the bottom left corner.

a classic CNN with several convolution blocks to process the “image” and predict a vector of the length of the discretized output function $y(t)$. The second implicit neural network approach uses a pretrained deep neural network (more specifically, the ImageNet architecture, Deng et al. 2009) and applies transfer learning for the given “images.” We also use an explicit functional neural network as proposed by Thind, Multani, and Cao (2023) (FuncNN) and the boosting approach FDboost by Brockhaus, Rügamer, and Greven (2020) and compare these approaches against a statistical approach, the pffr function for penalized FOFR in Goldsmith et al. (2020); see also Section 3. We split the data into a train and test set, using the ankle moment (inversion-eversion) as the outcome function and the different joint accelerations, velocities, and angles of all three body parts as functional covariates. For all methods except for the CNNs, we additionally use the scalar covariates age, height, weight, and sex as scalar features. For structured models (pffr, FDboost), we also incorporate random effects for the running condition and the study. Figure 13 depicts the true functions in the test set as well as predictions of the five methods. In addition, test performance results for the average point-wise root mean squared error (RMSE) between actual and predicted functions are given for every method. Results suggest that all methods work well in general, capturing the overall trend in the data. The approaches estimating an additive model (FDboost, pffr) tend to better capture some nuances in the data compared to the different neural network approaches. This is also confirmed by the RMSE values of 0.5966 and 0.5603 for the additive models, whereas neural network approaches yield values between 0.7226 and 0.9587.

6.5 | Software

Software packages that model functional data explicitly using ML and DL include FDboost (Brockhaus et al. 2017) building on a model-based boosting framework, FuncNN (Thind, Multani, and Cao 2023) for functional neural networks, funGp (Betancourt et al. 2023) using GPs for functional inputs or GPFDA (Shi, Cheng, and Konzen 2023) to combine functional regression with GPs. Implicit applications of ML can, for example, be implemented by using appropriate functional data transformations from the package scikit-fda (Ramos-Carreño et al. 2022).

7 | Outlook

7.1 | Functional Data and Related Fields

While we discussed the differences between FDA, longitudinal data (LD) analysis, and time series analysis in the introduction, developments in FDA have also led to fruitful cross-fertilization in these other two areas. Depending on the goal of the analysis and the data structure, especially the semi-/nonparametric analysis methods of FDA can also be of interest in these two fields. In particular, LD can be viewed as sparse (and noisy) functional data. FDA then provides new perspectives: the commonly used linear mixed models for LD are parametric, with, for example, a random intercept random slope model being linear over time per subject (observational unit). On the other hand, FPCA with its subject-specific random scores for the general principal component

(PC) functions, can be viewed as a linear mixed model with data-driven PC basis functions replacing the constant and linear basis functions, yielding larger flexibility and a better semiparametric fit to the LD (e.g., Yao, Müller, and Wang 2005; Goldsmith, Greven, and Crainiceanu 2013). Similar approaches have also been successfully incorporated in joint models for longitudinal and time-to-event data to allow for more flexible modeling of the univariate or multivariate longitudinal trajectories (e.g., Köhler et al. 2017; Köhler, Umlauf, and Greven 2018; Li, Xiao, and Luo 2022; Volkmann, Umlauf, and Greven 2023). While time series analysis usually focuses on one time series, there are also cases where a sample of time series is collected and questions such as time series classification are of interest, in which case functional data classification methods can be utilized as discussed in Sections 5.1 and 6. Generally, there are often different ways to look at the same data depending on the research questions of interest. In addition to longitudinal and time series data, this also holds, for example, for spatiotemporal data, which depending on the setting and questions of interest could also often be viewed as repeated observations of spatial random fields over time, or spatially correlated curves (time series or functional data) over time (e.g., Delicado et al. 2010; Kokoszka 2012). Besides spatial correlation, functional data can appear with other additional structures known from scalar data, such as, for example, functional time series (e.g., Hyndman and Shang 2009; Kokoszka 2012; Aue, Norinho, and Hörmann 2015) or longitudinal functional data (e.g., Greven et al. 2010; Park and Staicu 2015). Koner and Staicu (2023) call multivariate functional data, longitudinal functional data, functional time series, and spatially correlated functional data “second-generation functional data” and provide a review.

7.2 | Beyond One-Dimensional Functions

FDA is still an active field of research with many current developments. In recent years, an important focus of the field has been to broaden the view and see functional data as just one example of more general “object data,” that is, of data where each observation constitutes an object in a more complex space than the usual \mathbb{R}^d space for vectors. This area is sometimes called object (oriented) data analysis (Marron and Alonso 2014). While the univariate functional data considered here are often viewed as objects in the $\mathcal{L}^2(\mathcal{T})$ of square-integrable functions on some interval $\mathcal{T} \subset \mathbb{R}$, images and surface-valued data (e.g., Goldsmith, Huang, and Crainiceanu 2014; Happ and Greven 2018) can be viewed as $\mathcal{L}^2(\mathcal{T})$ functions on a higher dimensional domain $\mathcal{T} \subset \mathbb{R}^d$, $d > 1$, while multivariate functions are often defined on an interval \mathcal{T} but to be vector-valued in \mathbb{R}^d , $d > 1$ (e.g., Chiou, Chen, and Yang 2014; Górecki et al. 2018), with a more general setting considered in Happ and Greven (2018). Examples are brain scans for the former, and movement trajectories, outlines, or several recorded functions for the latter (e.g., Steyer, Stöcker, and Greven 2023a, 2023b; Volkmann, Stöcker, et al. 2023). Many methods developed for univariate functional data can be suitably extended to images and multivariate functional data such as, for example, FPCA (e.g., Berrendero, Justel, and Svarc 2011; Chiou, Chen, and Yang 2014; Happ and Greven 2018) or regression models (e.g., Goldsmith, Huang, and Crainiceanu 2014; Chiou, Yang, and Chen 2016; Volkmann, Stöcker, et al. 2023). For both settings, phase variation can also be important, for image analysis

due to necessary registration. For multivariate functional data, it can also be relevant to have the analysis be invariant to warping, that is, only analyze the image of a curve, such as when an outline of an object such as a cell or bone is recorded and the parameterization of the curve along that outline is essentially arbitrary. In this case, the object of interest is the equivalence class of such curves with respect to reparameterization, and methods such as distance and mean computation as well as regression have also been developed for such elements of quotient spaces (e.g., Srivastava et al. 2010; Srivastava and Klassen 2016; Steyer, Stöcker, and Greven 2023a, 2023b). Similarly, the shape/form of an object (an outline or vector of landmark coordinates) is considered to be the equivalence class under rotation, translation, and (for shape) scale (Dryden and Mardia 2016), and is of interest if the coordinate system of the recorded object is arbitrary or not relevant. Such shapes/forms live on certain manifolds (e.g., Huckemann, Hotz, and Munk 2009; Thomas Fletcher 2013; Stöcker, Steyer, and Greven 2023) or in more complex quotient spaces, for example, if parametrization invariance is simultaneously considered (e.g., Stöcker et al. 2022; Steyer, Stöcker, and Greven 2023a). Shape analysis (Dryden and Mardia 2016) is an important field in itself and has more recently intersected with FDA in functional shape analysis (Srivastava and Klassen 2016) when the shape of a curve is of interest, for example, as the outcome in a regression model (Stöcker, Steyer, and Greven 2023). Similarly, compositional data analysis (van den Boogaart and Tolosana-Delgado 2013; Pawlowsky-Glahn, Egozcue, and Tolosana-Delgado 2015; Filzmoser, Hron, and Templ 2018) is a field with a long tradition, looking at multivariate vectors of nonnegative entries that add up to a constant such as 1 or 100%. It takes into account the special geometry of the simplex, which compositions are elements of. More recently, this field has also intersected with FDA when looking at probability densities, which can be viewed as functional data with particular constraints or as infinite compositions. More generally, continuous, discrete, mixed, bivariate, etc. densities can be considered (e.g., Maier et al. 2021), with one possible geometry given by the so-called Bayes–Hilbert-space approach (van den Boogaart, Egozcue, and Pawlowsky-Glahn 2010, 2014) and compositions seen as the special case of discrete densities. Other possible approaches include Wasserstein spaces (Panaretos and Zemel 2020). Further examples of object data include the so-called generalized functional data (e.g., Hall, Müller, and Yao 2008; Goldsmith, Zipunnikov, and Schrack 2015; Scheipl, Gertheiss, and Greven 2016; Greven and Scheipl 2017; Gertheiss, Goldsmith, and Staicu 2017), where conditional on a smooth mean curve over \mathcal{T} as in functional data, observed values are viewed as realizations from some non-Gaussian (e.g., binary or count) distribution, directional data (Mardia 1975, 1988), covariance-(matrix- or operator-)valued data (e.g., Masarotto, Panaretos, and Zemel 2019; Lin, Müller, and Park 2023), point-process-valued data (e.g., Panaretos and Zemel 2016), and tree- or network-valued data (e.g., Feragen et al. 2013; Duncan, Klassen, and Srivastava 2018; Calissano, Feragen, and Vantini 2022, 2024). In contrast to neighboring fields such as graphical models, the focus is not on modeling, for example, a single network, but always on the setting where a sample of networks or more generally objects is observed. The objects constitute the unit of analysis, for example, the response or covariate in a regression model. Many developments are expected to occur in this exciting field in the coming years.

Acknowledgments

We thank the associate editor and two anonymous reviewers for their very thoughtful and constructive comments that helped us to improve the original version of this paper. Sonja Greven gratefully acknowledges funding by grants GR 3793/2-2, GR 3793/3-1, and GR 3793/8-1 from the German Research Foundation (DFG). Bernard Liew is supported by The Academy of Medical Sciences, UK, Springboard Award SBF006/1019.

Open access funding enabled and organized by Projekt DEAL.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The R code and data needed to reproduce the analyses presented in the paper are found at https://github.com/davidruegamer/FDA_tutorial.

Open Research Badges



This article has earned an Open Data badge for making publicly available the digitally-shareable data necessary to reproduce the reported results. The data is available in the [Supporting Information](#) section.

This article has earned an open data badge “**Reproducible Research**” for making publicly available the code necessary to reproduce the reported results. The results reported in this article were reproduced partially due to partially non-executable code.

References

- Aneiros, G., S. Novo, and P. Vieu. 2022. “Variable Selection in Functional Regression Models: A Review.” *Journal of Multivariate Analysis* 188: 104871.
- Arribas-Gil, A., and J. Romo. 2014. “Shape Outlier Detection and Visualization for Functional Data: The Outliergram.” *Biostatistics* 15, no. 4: 603–619.
- Aue, A., D. D. Norinho, and S. Hörmann. 2015. “On the Prediction of Stationary Functional Time Series.” *Journal of the American Statistical Association* 110, no. 509: 378–392.
- Azaïs, J.-M., and J.-C. Fort. 2013. “Remark on the Finite-Dimensional Character of Certain Results of Functional Statistics.” *Comptes Rendus de l'Académie des Sciences - Series I - Mathematics* 351: 139–141.
- Barber, R. F., M. Reimherr, and T. Schill. 2017. “The Function-on-Scalar LASSO With Applications to Longitudinal GWAS.” *Electronic Journal of Statistics* 11, no. 1: 1351–1389.
- Benjamini, Y., and D. Yekutieli. 2001. “The Control of the False Discovery Rate in Multiple Testing Under Dependency.” *The Annals of Statistics* 29: 1165–1188.
- Benko, M., W. Härdle, and A. Kneip. 2009. “Common Functional Principal Components.” *The Annals of Statistics* 37: 1–34.
- Berrendero, J. R., A. Justel, and M. Svarc. 2011. “Principal Components for Multivariate Functional Data.” *Computational Statistics & Data Analysis* 55, no. 9: 2619–2634.
- Berrendero, J. R., A. Cuevas, and J. L. Torrecilla. 2018. “On the Use of Reproducing Kernel Hilbert Spaces in Functional Classification.” *Journal of the American Statistical Association* 113: 1210–1218.
- Betancourt, J., F. Bachoc, T. Klein, and J. Rohmer. 2023. “funGp: Gaussian Process Models for Scalar and Functional Inputs.” R Package Version 0.3.2. <https://CRAN.R-project.org/package=funGp>.
- Blanquero, R., E. Carrizosa, A. Jiménez-Cordero, and B. Martín-Barragán. 2019. “Variable Selection in Classification for Multivariate Functional Data.” *Information Sciences* 481: 445–462.
- Bosq, D. 2000. *Linear Processes in Function Spaces: Theory and Applications*, Lecture Notes in Statistics, vol. 149. Berlin: Springer Science & Business Media.
- Bouveyron, C., and J. Jacques. 2011. “Model-Based Clustering of Time Series in Group-Specific Functional Subspaces.” *Advances in Data Analysis and Classification* 5, no. 4: 281–300.
- Brockhaus, S., F. Scheipl, T. Hothorn, and S. Greven. 2015. “The Functional Linear Array Model.” *Statistical Modelling* 15, no. 3: 279–300.
- Brockhaus, S., M. Melcher, F. Leisch, and S. Greven. 2017. “Boosting Flexible Functional Regression Models With a High Number of Functional Historical Effects.” *Statistics and Computing* 27, no. 4: 913–926.
- Brockhaus, S., D. Rügamer, and S. Greven. 2020. “Boosting Functional Regression Models With Fdboost.” *Journal of Statistical Software* 94, no. 10: 1–50.
- Brunner, E., H. Dette, and A. Munk. 1997. “Box-Type Approximations in Nonparametric Factorial Designs.” *Journal of the American Statistical Association* 92: 1494–1502.
- Calissano, A., A. Feragen, and S. Vantini. 2022. “Graph-Valued Regression: Prediction of Unlabelled Networks in a Non-Euclidean Graph Space.” *Journal of Multivariate Analysis* 190: 104950.
- Calissano, A., A. Feragen, and S. Vantini. 2024. “Populations of Unlabelled Networks: Graph Space Geometry and Generalized Geodesic Principal Components.” *Biometrika* 111, no. 1: 147–170.
- Cardot, H., P. Cénac, and P.-A. Zitt. 2013. “Efficient and Fast Estimation of the Geometric Median in Hilbert Spaces With an Averaged Stochastic Gradient Algorithm.” *Bernoulli* 19, no. 1: 18–43.
- Cederbaum, J., F. Scheipl, and S. Greven. 2018. “Fast Symmetric Additive Covariance Smoothing.” *Computational Statistics & Data Analysis* 120: 25–41.
- Centofanti, F., M. Fontana, A. Lepore, and S. Vantini. 2020. “Smooth Lasso Estimator for the Function-on-Function Linear Regression Model.” Preprint. <https://doi.org/10.48550/arXiv.2007.00529>.
- Chakraborty, A., and P. Chaudhuri. 2014. “On Data Depth in Infinite Dimensional Spaces.” *Annals of the Institute of Statistical Mathematics* 66: 303–324.
- Chang, C., X. Lin, and R. Ogden. 2017. “Simultaneous Confidence Bands for Functional Regression Models.” *Journal of Statistical Planning and Inference* 188: 67–81.
- Chen, C., and A. Srivastava. 2021. “Srvfregnet: Elastic Function Registration Using Deep Neural Networks.” In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 4462–4471. Piscataway, NJ: IEEE.
- Chen, H., Y. Wang, M. C. Paik, and H. A. Choi. 2013. “A Marginal Approach to Reduced-Rank Penalized Spline Smoothing With Application to Multi-Level Functional Data.” *Journal of the American Statistical Association* 108: 1216–1229.
- Chen, H., P. T. Reiss, and T. Tarpey. 2014. “Optimally Weighted L2 Distance for Functional Data.” *Biometrics* 70, no. 3: 516–525.
- Chen, K., and H.-G. Müller. 2012. “Modeling Repeated Functional Observations.” *Journal of the American Statistical Association* 107, no. 500: 1599–1609.
- Chen, Y., J. Goldsmith, and R. T. Ogden. 2016. “Variable Selection in Function-on-Scalar Regression.” *Stat* 5, no. 1: 88–101.
- Cheng, Y., J. Q. Shi, and J. Eyre. 2020. “Nonlinear Mixed-Effects Scalar-on-Function Models and Variable Selection.” *Statistics and Computing* 30, no. 1: 129–140.
- Chiou, J.-M., Y.-T. Chen, and Y.-F. Yang. 2014. “Multivariate Functional Principal Component Analysis: A Normalization Approach.” *Statistica Sinica* 24: 1571–1596.

- Chiou, J.-M., Y.-F. Yang, and Y.-T. Chen. 2016. "Multivariate Functional Linear Regression and Prediction." *Journal of Multivariate Analysis* 146: 301–312.
- Choi, H., and M. Reimherr. 2018. "A Geometric Approach to Confidence Regions and Bands for Functional Parameters." *Journal of the Royal Statistical Society Series B: Statistical Methodology* 80: 239–260.
- Cuesta-Albertos, J. A., and M. Febrero-Bande. 2010. "A Simple Multiway ANOVA for Functional Data." *Test* 19: 537–557.
- Cuevas, A. 2014. "A Partial Overview of the Theory of Statistics With Functional Data." *Journal of Statistical Planning and Inference* 147: 1–23.
- Cuevas, A., M. Febrero, and R. Fraiman. 2004. "An ANOVA Test for Functional Data." *Computational Statistics & Data Analysis* 47: 111–122.
- Cérou, F., and A. Guyader. 2006. "Nearest Neighbor Classification in Infinite Dimension." *ESAIM: Probability and Statistics* 10: 340–355.
- Dai, W., and M. G. Genton. 2018. "Multivariate Functional Data Visualization and Outlier Detection." *Journal of Computational and Graphical Statistics* 27, no. 4: 923–934.
- Dauxois, J., A. Pousse, and Y. Romain. 1982. "Asymptotic Theory for the Principal Component Analysis of a Vector Random Function: Some Applications to Statistical Inference." *Journal of Multivariate Analysis* 12, no. 1: 136–154.
- Degras, D. 2011. "Simultaneous Confidence Bands for Nonparametric Regression With Functional Data." *Statistica Sinica* 21: 1735–1765.
- Delaigle, A., and P. Hall. 2012. "Achieving Near Perfect Classification for Functional Data." *Journal of the Royal Statistical Society Series B: Statistical Methodology* 74: 267–286.
- Delicado, P., R. Giraldo, C. Comas, and J. Mateu. 2010. "Statistics for Spatial Functional Data: Some Recent Contributions." *Environmetrics* 21, no. 3-4: 224–239.
- Deng, J., W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. 2009. "ImageNet: A Large-Scale Hierarchical Image Database." In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 248–255. Piscataway, NJ: IEEE.
- Di, C.-Z., C. M. Crainiceanu, B. S. Caffo, and N. M. Punjabi. 2009. "Multilevel Functional Principal Component Analysis." *The Annals of Applied Statistics* 3, no. 1: 458–488.
- Ding, F., S. He, D. E. Jones, and J. Z. Huang. 2022. "Functional PCA With Covariate-Dependent Mean and Covariance Structure." *Technometrics* 64, no. 3: 335–345.
- Ditzhaus, M., M. Munko, M. Pauly, L. Smaga, and J.-T. Zhang. 2023. "multiFANOVA: Multiple Contrast Tests for Functional Data." R Package Version 0.1.0. <https://CRAN.R-project.org/package=multiFANOVA>.
- Dryden, I. L., and K. V. Mardia. 2016. *Statistical Shape Analysis: With Applications in R*, Wiley Series in Probability and Statistics, vol. 995. Hoboken, NJ: John Wiley & Sons.
- Duncan, A., E. Klassen, and A. Srivastava. 2018. "Statistical Shape Analysis of Simplified Neuronal Trees." *The Annals of Applied Statistics* 12, no. 3: 1385–1421.
- Eilers, P. H., and B. D. Marx. 2021. *Practical Smoothing: The Joys of P-Splines*. Cambridge, UK: Cambridge University Press.
- Ezugwu, A. E., A. K. Shukla, M. B. Agbaje, O. N. Oyelade, A. José-García, and J. O. Agushaka. 2021. "Automatic Clustering Algorithms: A Systematic Review and Bibliometric Analysis of Relevant Literature." *Neural Computing and Applications* 33, no. 11: 6247–6306.
- Fan, Y., G. M. James, and P. Radchenko. 2015. "Functional Additive Regression." *The Annals of Statistics* 43: 2296–2325.
- Fan, Z., and M. Reimherr. 2017. "High-Dimensional Adaptive Function-on-Scalar Regression." *Econometrics and Statistics* 1: 167–183.
- Feng, S., M. Zhang, and T. Tong. 2022. "Variable Selection for Functional Linear Models With Strong Heredity Constraint." *Annals of the Institute of Statistical Mathematics* 74, no. 2: 321–339.
- Feragen, A., M. Owen, J. Petersen, et al. 2013. "Tree-Space Statistics and Approximations for Large-Scale Analysis of Anatomical Trees." In *Information Processing in Medical Imaging: Proceedings of the 23rd International Conference on Information Processing in Medical Imaging (IPMI 2013)*, Asilomar, CA, June 28–July 3, 2013, 74–85. Berlin: Springer.
- Ferraty, F., and P. Vieu. 2006. *Nonparametric Functional Data Analysis*. Springer Series in Statistics. New York: Springer.
- Ferraty, F., P. Hall, and P. Vieu. 2010. "Most-Predictive Design Points for Functional Data Predictors." *Biometrika* 97, no. 4: 807–824.
- Filzmoser, P., K. Hron, and M. Templ. 2018. *Applied Compositional Data Analysis—With Worked Examples in R*. Cham, Switzerland: Springer.
- Fremdt, S., J. Steinebach, L. Horváth, and P. Kokoszka. 2013. "Testing the Equality of Covariance Operators in Functional Samples." *Scandinavian Journal of Statistics* 40: 138–152.
- Fu, G., X. Dai, and Y. Liang. 2021. "Functional Random Forests for Curve Response." *Scientific Reports* 11, no. 1: 1–14.
- Fuchs, K., J. Gertheiss, and G. Tutz. 2015. "Nearest Neighbor Ensembles for Functional Data With Interpretable Feature Selection." *Chemometrics and Intelligent Laboratory Systems* 146: 186–197.
- Fukuchi, R. K., C. A. Fukuchi, and M. Duarte. 2017. "A Public Dataset of Running Biomechanics and the Effects of Running Speed on Lower Extremity Kinematics and Kinetics." *PeerJ* 5: e3298.
- García-Portugués, E., W. González-Manteiga, and M. Febrero-Bande. 2014. "A Goodness-of-Fit Test for the Functional Linear Model With Scalar Response." *Journal of Computational and Graphical Statistics* 23, no. 3: 761–778.
- Gasser, T., H.-G. Müller, and V. Mammitzsch. 1985. "Kernels for Nonparametric Curve Estimation." *Journal of the Royal Statistical Society Series B: Statistical Methodology* 47, no. 2: 238–252.
- Gertheiss, J., A. Maity, and A.-M. Staicu. 2013. "Variable Selection in Generalized Functional Linear Models." *Stat* 2, no. 1: 86–101.
- Gertheiss, J., V. Maier, E. F. Hessel, and A.-M. Staicu. 2015. "Marginal Functional Regression Models for Analyzing the Feeding Behavior of Pigs." *Journal of Agricultural, Biological and Environmental Statistics* 20: 353–370.
- Gertheiss, J., J. Goldsmith, and A.-M. Staicu. 2017. "A Note on Modeling Sparse Exponential-Family Functional Response Curves." *Computational Statistics & Data Analysis* 105: 46–52.
- Ghiglietti, A., F. Ieva, and A. Paganoni. 2017. "Statistical Inference for Stochastic Processes: Two-Sample Hypothesis Tests." *Journal of Statistical Planning and Inference* 180: 49–68.
- Gijbels, I., and S. Nagy. 2017. "On a General Definition of Depth for Functional Data." *Statistical Science* 32, no. 4: 630–639.
- Goldsmith, J., J. Bobb, C. M. Crainiceanu, B. Caffo, and D. Reich. 2011. "Penalized Functional Regression." *Journal of Computational and Graphical Statistics* 20, no. 4: 830–851.
- Goldsmith, J., S. Greven, and C. Crainiceanu. 2013. "Corrected Confidence Bands for Functional Data Using Principal Components." *Biometrics* 69, no. 1: 41–51.
- Goldsmith, J., L. Huang, and C. M. Crainiceanu. 2014. "Smooth Scalar-on-Image Regression via Spatial Bayesian Variable Selection." *Journal of Computational and Graphical Statistics* 23, no. 1: 46–64.
- Goldsmith, J., V. Zippunnikov, and J. Schrack. 2015. "Generalized Multi-level Function-on-Scalar Regression and Principal Component Analysis." *Biometrics* 71, no. 2: 344–353.
- Goldsmith, J., F. Scheipl, L. Huang, et al. 2020. "refund: Regression With Functional Data." R Package Version 0.1-23.
- Górecki, T., M. Krzyśko, Ł. Waszak, and W. Wołyński. 2018. "Selected Statistical Methods of Data Analysis for Multivariate Functional Data." *Statistical Papers* 59: 153–182.

- Greven, S., and F. Scheipl. 2017. "A General Framework for Functional Regression Modelling." *Statistical Modelling* 17, no. 1–2: 1–35.
- Greven, S., C. Crainiceanu, B. Caffo, and D. Reich. 2010. "Longitudinal Functional Principal Component Analysis." *Electronic Journal of Statistics* 4: 1022–1054.
- Guo, J., B. Zhou, and J.-T. Zhang. 2018. "Testing the Equality of Several Covariance Functions for Functional Data: A Supremum-Norm Based Test." *Computational Statistics & Data Analysis* 124: 15–26.
- Guo, J., B. Zhou, and J.-T. Zhang. 2019. "New Tests for Equality of Several Covariance Functions for Functional Data." *Journal of the American Statistical Association* 114: 1251–1263.
- Guss, W. H. 2016. "Deep Function Machines: Generalized Neural Networks for Topological Layer Expression." Preprint. <https://doi.org/10.48550/arXiv.1612.04799>.
- Górecki, T., and L. Smaga. 2015. "A Comparison of Tests for the One-Way ANOVA Problem for Functional Data." *Computational Statistics* 30: 987–1010.
- Górecki, T., and L. Smaga. 2017. "Multivariate Analysis of Variance for Functional Data." *Journal of Applied Statistics* 44: 2172–2189.
- Górecki, T., and L. Smaga. 2019. "fdANOVA: An R Software Package for Analysis of Variance for Univariate and Multivariate Functional Data." *Computational Statistics* 34, no. 2: 571–597.
- Hadjipantelis, P. Z., J. A. Aston, H.-G. Müller, and J. Moriarty. 2014. "Analysis of Spike Train Data: A Multivariate Mixed Effects Model for Phase and Amplitude." *Electronic Journal of Statistics* 8, no. 2: 1797–1807.
- Hall, H., P. Müller, and F. Yao. 2008. "Modelling Sparse Generalized Longitudinal Observations With Latent Gaussian Processes." *Journal of the Royal Statistical Society, Series B Statistical Methodology* 70: 703–723.
- Happ, C., and S. Greven. 2018. "Multivariate Functional Principal Component Analysis for Data Observed on Different (Dimensional) Domains." *Journal of the American Statistical Association* 113, no. 522: 649–659.
- Happ, C., F. Scheipl, A.-A. Gabriel, and S. Greven. 2019. "A General Framework for Multivariate Functional Principal Component Analysis of Amplitude and Phase Variation." *Stat* 8, no. 1: e220.
- Happ-Kurz, C. 2022. "MFPCA: Multivariate Functional Principal Component Analysis for Data Observed on Different Dimensional Domains." R Package Version 1.3-10. <https://CRAN.R-project.org/package=MFPCA>.
- Hartigan, J. A., and M. A. Wong. 1979. "Algorithm AS 136: A k-Means Clustering Algorithm." *Journal of the Royal Statistical Society Series C: Applied Statistics* 28, no. 1: 100–108.
- Heard, N. A., C. C. Holmes, and D. A. Stephens. 2006. "A Quantitative Study of Gene Regulation Involved in the Immune Response of Anopheles Mosquitoes." *Journal of the American Statistical Association* 101, no. 473: 18–29.
- Hörmann, S., and F. Jammoul. 2022. "Preprocessing Noisy Functional Data: A Multivariate Perspective." *Electronic Journal of Statistics* 16, no. 2: 6232–6266.
- Hornik, K., M. Stinchcombe, and H. White. 1989. "Multilayer Feedforward Networks are Universal Approximators." *Neural Networks* 2, no. 5: 359–366.
- Horváth, L., and P. Kokoszka. 2012. *Inference for Functional Data With Applications*. New York: Springer.
- Horváth, L., P. Kokoszka, and R. Reeder. 2013. "Estimation of the Mean of Functional Time Series and a Two-Sample Problem." *Journal of the Royal Statistical Society Series B: Statistical Methodology* 75: 103–122.
- Hsieh, T.-Y., Y. Sun, S. Wang, and V. G. Honavar. 2021. "Functional Autoencoders for Functional Data Representation Learning." In *Proceedings of the SIAM Conference on Data Mining*, 666–674. Philadelphia, PA: SIAM.
- Hsing, T., and R. Eubank. 2015. *Theoretical Foundations of functional Data analysis, With an Introduction to Linear Operators*, Wiley Series in Probability and Statistics, vol. 997. Hoboken, NJ: John Wiley & Sons.
- Huckemann, S., T. Hotz, and A. Munk. 2009. "Intrinsic MANOVA for Riemannian Manifolds With an Application to Kendall's Space of Planar Shapes." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 32, no. 4: 593–603.
- Hyndman, R. J., and H. L. Shang. 2009. "Forecasting Functional Time Series." *Journal of the Korean Statistical Society* 38: 199–211.
- Hyndman, R. J., and H. L. Shang. 2010. "Rainbow Plots, Bagplots, and Boxplots for Functional Data." *Journal of Computational and Graphical Statistics* 19, no. 1: 29–45.
- Ieva, F., A. M. Paganoni, D. Pigoli, and V. Vitelli. 2013. "Multivariate Functional Clustering for the Morphological Analysis of Electrocardiograph Curves." *Journal of the Royal Statistical Society Series C: Applied Statistics* 62, no. 3: 401–418.
- Jacques, J., and C. Preda. 2013. "Funclust: A Curves Clustering Method Using Functional Random Variables Density Approximation." *Neurocomputing* 112: 164–171.
- Jacques, J., and C. Preda. 2014a. "Functional Data Clustering: A Survey." *Advances in Data Analysis and Classification* 8, no. 3: 231–255.
- Jacques, J., and C. Preda. 2014b. "Model-Based Clustering for Multivariate Functional Data." *Computational Statistics & Data Analysis* 71: 92–106.
- James, G. M., and C. A. Sugar. 2003. "Clustering for Sparsely Sampled Functional Data." *Journal of the American Statistical Association* 98, no. 462: 397–408.
- James, G. M., J. Wang, and J. Zhu. 2009. "Functional Linear Regression That's Interpretable." *The Annals of Statistics* 37, no. 5A: 2083–2108.
- Jeon, J., B. U. Park, and I. Van Keilegom. 2021. "Additive Regression for Non-Euclidean Responses and Predictors." *The Annals of Statistics* 49: 2611–2641.
- Karhunen, K. 1947. "Über Lineare Methoden in der Wahrscheinlichkeitsrechnung." *Annales Academiae Scientiarum Fennicae* 37: 1–79.
- Kingma, D. P., and M. Welling. 2013. "Auto-Encoding Variational Bayes." Preprint. <https://doi.org/10.48550/arXiv.1312.6114>.
- Köhler, M., N. Umlauf, A. Beyerlein, C. Winkler, A.-G. Ziegler, and S. Greven. 2017. "Flexible Bayesian Additive Joint Models With an Application to Type 1 Diabetes Research." *Biometrical Journal* 59, no. 6: 1144–1165.
- Köhler, M., N. Umlauf, and S. Greven. 2018. "Nonlinear Association Structures in Flexible Bayesian Additive Joint Models." *Statistics in Medicine* 37, no. 30: 4771–4788.
- Kokoszka, P. 2012. "Dependent Functional Data." *International Scholarly Research Network, ISRN Probability and Statistics* 2012: 958254.
- Kokoszka, P., and M. Reimherr. 2017. *Introduction to Functional Data Analysis*. Boca Raton, FL: CRC Press.
- Koner, S., and A.-M. Staicu. 2023. "Second-Generation Functional Data." *Annual Review of Statistics and Its Application* 10: 547–572.
- Kong, D., A.-M. Staicu, and A. Maity. 2016. "Classical Testing in Functional Linear Models." *Journal of Nonparametric Statistics* 28: 813–838.
- Konzen, E., Y. Cheng, and J. Q. Shi. 2021. "Gaussian Process for Functional Data Analysis: The Gpfdpa Package for R." <https://doi.org/10.48550/arXiv.2102.00249>.
- Kowal, D. R., and D. C. Bourgeois. 2020. "Bayesian Function-on-Scalars Regression for High-Dimensional Data." *Journal of Computational and Graphical Statistics* 29, no. 3: 629–638.
- Kraus, D., and V. Panaretos. 2012. "Dispersion Operators and Resistant Second-Order Analysis of Functional Data." *Biometrika* 99: 813–832.
- Krzysko, M., and L. Smaga. 2021. "Two-Sample Tests for Functional Data Using Characteristic Functions." *Austrian Journal of Statistics* 50: 53–64.
- Li, C., L. Xiao, and S. Luo. 2022. "Joint Model for Survival and Multivariate Sparse Functional Data With Application to a Study of Alzheimer's Disease." *Biometrics* 78, no. 2: 435–447.

- Li, Y., Y. Qiu, and Y. Xu. 2022. "From Multivariate to Functional Data Analysis: Fundamentals, Recent Developments, and Emerging Areas." *Journal of Multivariate Analysis* 188: 104806.
- Lian, H. 2013. "Shrinkage Estimation and Selection for Multiple Functional Regression." *Statistica Sinica* 23, no. 1: 51–74.
- Liebl, D., and M. Reimherr. 2023. "Fast and Fair Simultaneous Confidence Bands for Functional Parameters." *Journal of the Royal Statistical Society Series B: Statistical Methodology* 85, no. 3: 842–868.
- Liew, B. X., S. Morris, J. W. Keogh, B. Appleby, and K. Netto. 2016. "Effects of two Neuromuscular Training Programs on Running Biomechanics With Load Carriage: A Study Protocol for a Randomised Controlled Trial." *BMC Musculoskeletal Disorders* 17, no. 1: 1–10.
- Liew, B. X., S. Morris, and K. Netto. 2016. "The Effects of Load Carriage on Joint Work at Different Running Velocities." *Journal of Biomechanics* 49, no. 14: 3275–3280.
- Liew, B. X., D. Rügamer, A. Stocker, and A. M. De Nunzio. 2020. "Classifying Neck Pain Status Using Scalar and Functional Biomechanical Variables—Development of a Method Using Functional Data Boosting." *Gait & Posture* 76: 146–150.
- Liew, B. X., D. Rügamer, X. Zhai, Y. Wang, S. Morris, and K. Netto. 2021. "Comparing Shallow, Deep, and Transfer Learning in Predicting Joint Moments in Running." *Journal of Biomechanics* 129: 110820.
- Liew, B. X., F. Pfisterer, D. Rügamer, and X. Zhai. 2024. "Strategies to Optimise Machine Learning Classification Performance When Using Biomechanical Features." *Journal of Biomechanics* 165: 111998.
- Lin, Z., H.-G. Müller, and B. Park. 2023. "Additive Models for Symmetric Positive-Definite Matrices and Lie Groups." *Biometrika* 110, no. 2: 361–379.
- Ling, N., and P. Vieu. 2021. "On Semiparametric Regression in Functional Data Analysis." *WIREs Computational Statistics* 13: e1538.
- Loève, M. 1945. "Fonctions Aléatoires du Second Ordre." *Comptes Rendus Académie des Sciences* 220: 380.
- López-Pintado, S., and J. Romo. 2009. "On the Concept of Depth for Functional Data." *Journal of the American Statistical Association* 104, no. 486: 718–734.
- Lu, Y., R. Herbei, and S. Kurtek. 2017. "Bayesian Registration of Functions With a Gaussian Process Prior." *Journal of Computational and Graphical Statistics* 26, no. 4: 894–904.
- MacQueen, J. 1967. "Some Methods for Classification and Analysis of Multivariate Observations." In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, edited by L. M. Le Cam and J. Neyman, editors, vol. 1, 281–297. Berkeley, CA: University of California Press.
- Maechler, M., P. Rousseeuw, A. Struyf, M. Hubert, and K. Hornik. 2021. "cluster: Cluster Analysis Basics and Extensions." R Package Version 2.1.1. <https://CRAN.R-project.org/package=cluster>.
- Maier, E.-M., A. Stöcker, B. Fitzenberger, and S. Greven. 2021. "Additive Density-on-Scalar Regression in Bayes Hilbert Spaces With an Application to Gender Economics." Preprint. <https://doi.org/10.48550/arXiv.2110.11771>.
- Mardia, K. V. 1988. "Directional Data Analysis: An Overview." *Journal of Applied Statistics* 15, no. 2: 115–122.
- Mardia, K. V. 1975. "Statistics of Directional Data." *Journal of the Royal Statistical Society Series B: Statistical Methodology* 37, no. 3: 349–371.
- Marron, J. S., and A. M. Alonso. 2014. "Overview of Object Oriented Data Analysis." *Biometrical Journal* 56, no. 5: 732–753.
- Marron, J. S., J. O. Ramsay, L. M. Sangalli, and A. Srivastava. 2015. "Functional Data Analysis of Amplitude and Phase Variation." *Statistical Science* 30, no. 4: 468–484.
- Martelli, S., M. E. Kersh, A. G. Schache, and M. G. Pandey. 2014. "Strain Energy in the Femoral Neck During Exercise." *Journal of Biomechanics* 47, no. 8: 1784–1791.
- Marx, B. D., and P. H. C. Eilers. 1999. "Generalized Linear Regression on Sampled Signals and Curves: A p-spline Approach." *Technometrics* 41, no. 1: 1–13.
- Masarotto, V., V. M. Panaretos, and Y. Zemel. 2019. "Procrustes Metrics on Covariance Operators and Optimal Transportation of Gaussian Processes." *Sankhya A* 81: 172–213.
- Matsui, H. 2014. "Variable and Boundary Selection for Functional Data via Multiclass Logistic Regression Modeling." *Computational Statistics & Data Analysis* 78: 176–185.
- Matsui, H., and S. Konishi. 2011. "Variable Selection for Functional Regression Models via the l1 Regularization." *Computational Statistics & Data Analysis* 55, no. 12: 3304–3310.
- Matuk, J., K. Bharath, O. Chkrebti, and S. Kurtek. 2022. "Bayesian Framework for Simultaneous Registration and Estimation of Noisy, Sparse, and Fragmented Functional Data." *Journal of the American Statistical Association* 117, no. 540: 1964–1980.
- McLean, M. W., F. Scheipl, G. Hooker, S. Greven, and D. Ruppert. 2013. "Bayesian Functional Generalized Additive Models With Sparsely Observed Covariates." Preprint. <https://doi.org/10.48550/arXiv.1305.3585>.
- McLean, M. W., G. Hooker, and D. Ruppert. 2015. "Restricted Likelihood Ratio Tests for Linearity in Scalar-on-Function Regression." *Statistics and Computing* 25, no. 5: 997–1008.
- Mercer, J. 1909. "Functions of Positive and Negative Type, and Their Connection With the Theory of Integral Equations." *Philosophical Transactions of the Royal Society of London. Series A* 209, no. 441–458: 415–446.
- Möller, A., G. Tutz, and J. Gertheiss. 2016. "Random Forests for Functional Covariates." *Journal of Chemometrics* 30, no. 12: 715–725.
- Montagnani, E., S. C. Morrison, M. Varga, and C. Price. 2021. "Pedobarographic Statistical Parametric Mapping of Plantar Pressure Data in New and Confident Walking Infants: A Preliminary Analysis." *Journal of Biomechanics* 129: 110757.
- Morris, J. S., and R. J. Carroll. 2006. "Wavelet-Based Functional Mixed Models." *Journal of the Royal Statistical Society Series B: Statistical Methodology* 68, no. 2: 179–199.
- Munko, M., M. Ditzhaus, M. Pauly, L. Smaga, and J.-T. Zhang. 2023. "General Multiple Tests for Functional Data."
- Murtagh, F., and P. Contreras. 2012. "Algorithms for Hierarchical Clustering: An Overview." *WIREs Data Mining and Knowledge Discovery* 2, no. 1: 86–97.
- Müller, H.-G., and U. Stadtmüller. 2005. "Generalized Functional Linear Models." *The Annals of Statistics* 33: 774–805.
- Müller, H.-G., and F. Yao. 2008. "Functional Additive Models." *Journal of the American Statistical Association* 103: 1534–1544.
- Ojo, O. T., R. E. Lillo, and A. Fernandez Anta. 2023. "fdaoutlier: Outlier Detection Tools for Functional Data Analysis." R Package Version 0.2.1. <https://CRAN.R-project.org/package=fdaoutlier>.
- Ott, F., D. Rügamer, L. Heublein, B. Bischl, and C. Mutschler. 2021. "Joint Classification and Trajectory Regression of Online Handwriting Using a Multi-Task Learning Approach." In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 1244–1254. Piscataway, NJ: IEEE.
- Pan, S. J., and Q. Yang. 2010. "A Survey on Transfer Learning." *IEEE Transactions on Knowledge and Data Engineering* 22, no. 10: 1345–1359.
- Panaretos, V. M., and S. Tavakoli. 2013. "Cramér–Karhunen–Loève Representation and Harmonic Principal Component Analysis of Functional Time Series." *Stochastic Processes and Their Applications* 123, no. 7: 2779–2807.
- Panaretos, V. M., and Y. Zemel. 2016. "Amplitude and Phase Variation of Point Processes." *Annals of Statistics* 44: 771–812.
- Panaretos, V. M., and Y. Zemel. 2020. *An Invitation to Statistics in Wasserstein Space*. Cham, Switzerland: Springer.

- Paparoditis, E., and T. Sapatinas. 2016. "Bootstrap-Based Testing of Equality of Mean Functions or Equality of Covariance Operators for Functional Data." *Biometrika* 103: 727–733.
- Park, S. Y., and A.-M. Staicu. 2015. "Longitudinal Functional Data Analysis." *Stat* 4, no. 1: 212–226.
- Parodi, A., and M. Reimherr. 2018. "Simultaneous Variable Selection and Smoothing for High-Dimensional Function-on-Scalar Regression." *Electronic Journal of Statistics* 12, no. 2: 4602–4639.
- Pataky, T. C. 2010. "Generalized N-Dimensional Biomechanical Field Analysis Using Statistical Parametric Mapping." *Journal of Biomechanics* 43, no. 10: 1976–1982.
- Pataky, T. C., J. Vanrenterghem, and M. A. Robinson. 2015. "Zero-vs. One-Dimensional, Parametric vs. Non-Parametric, and Confidence Interval vs. Hypothesis Testing Procedures in One-Dimensional Biomechanical Trajectory Analysis." *Journal of Biomechanics* 48, no. 7: 1277–1285.
- Pataky, T. C., J. Vanrenterghem, and M. A. Robinson. 2016. "The Probability of False Positives in Zero-Dimensional Analyses of One-Dimensional Kinematic, Force and EMG Trajectories." *Journal of Biomechanics* 49, no. 9: 1468–1476.
- Pawlowsky-Glahn, V., J. J. Egozcue, and R. Tolosana-Delgado. 2015. *Modeling and Analysis of Compositional Data*. New York: John Wiley & Sons.
- Pfisterer, F., L. Beggel, X. Sun, F. Scheipl, and B. Bischl. 2019. "Benchmarking Time Series Classification—Functional Data vs Machine Learning Approaches." Preprint. <https://doi.org/10.48550/arXiv.1911.07511>.
- Pini, A., and S. Vantini. 2016. "The Interval Testing Procedure: A General Framework for Inference in Functional Data Analysis." *Biometrics* 72: 835–845.
- Pini, A., and S. Vantini. 2022. "fdatest: Interval Testing Procedure for Functional Data." R Package Version 2.1.1. <https://CRAN.R-project.org/package=fdatest>.
- Pomann, G.-M., A.-M. Staicu, and S. Ghosh. 2016. "A Two-Sample Distribution-Free Test for Functional Data With Application to a Diffusion Tensor Imaging Study of Multiple Sclerosis." *Journal of the Royal Statistical Society Series C: Applied Statistics* 65, no. 3: 395–414.
- Qiu, Z., J. Fan, J.-T. Zhang, and J. Chen. 2024. "Tests for Equality of Several Covariance Matrix Functions for Multivariate Functional Data." *Journal of Multivariate Analysis* 199: 105243.
- R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Rahman, R., S. R. Dhruva, S. Ghosh, and R. Pal. 2019. "Functional Random Forest With Applications in Dose-Response Predictions." *Scientific Reports* 9, no. 1: 1–14.
- Ramos-Carreño, C., J. L. Torrecilla, Y. Hong, and A. Suárez. 2022. "Scikit-fda: Computational Tools for Machine Learning With Functional Data." In *2022 IEEE 34th International Conference on Tools With Artificial Intelligence (ICTAI)*, 213–218. Piscataway, NJ: IEEE.
- Ramsay, J. O., and X. Li. 1998. "Curve Registration." *Journal of the Royal Statistical Society Series B: Statistical Methodology* 60, no. 2: 351–363.
- Ramsay, J. O., and B. W. Silverman. 2005. *Functional Data Analysis*. Springer Series in Statistics. New York: Springer.
- Ramsay, J. O., G. Hooker, and S. Graves. 2009. *Functional Data Analysis With R and MATLAB*. New York: Springer.
- Ramsay, J. O., S. Graves, and G. Hooker. 2024. "fda: Functional Data Analysis." R Package Version 6.1.8. <https://CRAN.R-project.org/package=fda>.
- Rao, A. R., and M. Reimherr. 2023. "Nonlinear Functional Modeling Using Neural Networks." *Journal of Computational and Graphical Statistics* 32, no. 4: 1248–1257.
- Reiss, P. T., J. Goldsmith, H. L. Shang, and R. T. Ogden. 2017. "Methods for Scalar-on-Function Regression." *International Statistical Review* 85, no. 2: 228–249.
- Rice, J. A., and B. W. Silverman. 1991. "Estimating the Mean and Covariance Structure Nonparametrically When the Data are Curves." *Journal of the Royal Statistical Society Series B: Statistical Methodology* 53, no. 1: 233–243.
- Robinson, M. A., J. Vanrenterghem, and T. C. Pataky. 2015. "Statistical Parametric Mapping (spm) for Alpha-Based Statistical Analyses of Multi-Muscle EMG Time-Series." *Journal of Electromyography and Kinesiology* 25, no. 1: 14–19.
- Rossi, F., and N. Villa. 2006. "Support Vector Machine for Functional Data Classification." *Neurocomputing* 69, no. 7–9: 730–742.
- Rossi, F., B. Conan-Guez, and F. Fleuret. 2002. "Functional Data Analysis With Multi Layer Perceptrons." In *Proceedings of the 2002 International Joint Conference on Neural Networks. IJCNN'02*, vol. 3, 2843–2848. Piscataway, NJ: IEEE.
- Rügamer, D., S. Brockhaus, K. Gentsch, K. Scherer, and S. Greven. 2018. "Boosting Factor-Specific Functional Historical Models for the Detection of Synchronization in Bioelectrical Signals." *Journal of the Royal Statistical Society Series C: Applied Statistics* 67, no. 3: 621–642.
- Rügamer, D., B. Liew, Z. Altai, and A. Stöcker. 2024. "A Functional Extension of Semi-Structured Networks." Preprint (under review).
- Sarkar, S., and V. M. Panaretos. 2022. "Covnet: Covariance Networks for Functional Data on Multidimensional Domains." *Journal of the Royal Statistical Society Series B: Statistical Methodology* 84, no. 5: 1785–1820.
- Scheipl, F., A.-M. Staicu, and S. Greven. 2015. "Functional Additive Mixed Models." *Journal of Computational and Graphical Statistics* 24, no. 2: 477–501.
- Scheipl, F., J. Gertheiss, and S. Greven. 2016. "Generalized Functional Additive Mixed Models." *Electronic Journal of Statistics* 10: 1455–1492.
- Schmutz, A., J. Jacques, C. Bouveyron, L. Chéze, and P. Martin. 2020. "Clustering Multivariate Functional Data in Group-Specific Functional Subspaces." *Computational Statistics* 35, no. 3: 1101–1131.
- Schmutz, A., J. Jacques, and C. Bouveyron. 2021. "funHDDC: Univariate and Multivariate Model-Based Clustering in Group-Specific Functional Subspaces." R Package Version 2.3.1. <https://CRAN.R-project.org/package=funHDDC>.
- Schulte, R., and D. Rügamer. 2024. "On the Interconnections Between Gradient Boosting and Gradient Optimization Methods." Preprint (under review).
- Scrucca, L., M. Fop, T. B. Murphy, and A. E. Raftery. 2016. "Mclust 5: Clustering, Classification and Density Estimation Using Gaussian Finite Mixture Models." *The R Journal* 8, no. 1: 289–317.
- Selk, L. 2024. "Uniform Convergence Rates and Automatic Variable Selection in Nonparametric Regression With Functional and Categorical Covariates." *Journal of Nonparametric Statistics* 36: 264–286.
- Selk, L., and J. Gertheiss. 2023. "Nonparametric Regression and Classification With Functional, Categorical, and Mixed Covariates." *Advances in Data Analysis and Classification* 17: 519–543.
- Shang, H. L., and R. Hyndman. 2024. "rainbow: Bagplots, Boxplots and Rainbow Plots for Functional Data." R Package Version 3.8. <https://CRAN.R-project.org/package=rainbow>.
- Shen, Q., and J. Faraway. 2004. "An F Test for Linear Models With Functional Responses." *Statistica Sinica* 14: 1239–1257.
- Shi, J. Q., and T. Choi. 2011. *Gaussian Process Regression Analysis for Functional Data*. Boca Raton, FL: CRC Press.
- Shi, J. Q., Y. Cheng, and E. Konzen. 2023. "GPFDA: Gaussian Process for Functional Data Analysis." R Package Version 4.0.0. <https://CRAN.R-project.org/package=GPFDA>.
- Shi, Z., J. Fan, L. Song, D.-X. Zhou, and J. A. Suykens. 2024. "Nonlinear Functional Regression by Functional Deep Neural Network With Kernel Embedding." Preprint. <https://doi.org/10.48550/arXiv.2401.02890>.

- Shou, H., V. Zipunnikov, C. M. Crainiceanu, and S. Greven. 2015. "Structured Functional Principal Component Analysis." *Biometrics* 71, no. 1: 247–257.
- Simmons, J. P., L. D. Nelson, and U. Simonsohn. 2011. "False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant." *Psychological Science* 22: 1359–1366.
- Smaga, L., and J.-T. Zhang. 2019. "Linear Hypothesis Testing With Functional Data." *Technometrics* 61: 99–110.
- Srivastava, A., and E. P. Klassen. 2016. *Functional and Shape Data Analysis*. New York: Springer.
- Srivastava, A., E. Klassen, S. H. Joshi, and I. H. Jermyn. 2010. "Shape Analysis of Elastic Curves in Euclidean Spaces." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 33, no. 7: 1415–1428.
- Steyer, L. 2024. "elasdics: Elastic Analysis of Sparse, Dense and Irregular Curves." R Package Version 1.1.3. <https://CRAN.R-project.org/package=elasdics>.
- Steyer, L., A. Stöcker, and S. Greven. 2023a. "Elastic Analysis of Irregularly or Sparsely Sampled Curves." *Biometrics* 79: 2103–2115.
- Steyer, L., A. Stöcker, and S. Greven. 2023b. "Regression in Quotient Metric Spaces With a Focus on Elastic Curves." Preprint. <https://doi.org/10.48550/arXiv.2305.02075>.
- Stöcker, A., M. Pfeuffer, L. Steyer, and S. Greven. 2022. "Elastic Full Procrustes Analysis of Plane Curves via Hermitian Covariance Smoothing." Preprint. <https://doi.org/10.48550/arXiv.2203.10522>.
- Stöcker, A., L. Steyer, and S. Greven. 2023. "Functional Additive Models on Manifolds of Planar Shapes and Forms." *Journal of Computational and Graphical Statistics* 32: 1600–1612.
- Su, Y.-R., C.-Z. Di, and L. Hsu. 2017. "Hypothesis Testing in Functional Linear Models." *Biometrics* 73, no. 2: 551–561.
- Sun, Y., and M. G. Genton. 2011. "Functional Boxplots." *Journal of Computational and Graphical Statistics* 20, no. 2: 316–334.
- Sutskever, I., O. Vinyals, and Q. V. Le. 2014. "Sequence to Sequence Learning With Neural Networks." *Advances in Neural Information Processing Systems* 27: 3104–3112.
- Swihart, B. J., J. Goldsmith, and C. M. Crainiceanu. 2014. "Restricted Likelihood Ratio Tests for Functional Effects in the Functional Linear Model." *Technometrics* 56, no. 4: 483–493.
- Szucs, D. 2016. "A Tutorial on Hunting Statistical Significance by Chasing n." *Frontiers in Psychology* 7: 1444.
- Sørensen, H., J. Goldsmith, and L. M. Sangalli. 2013. "An Introduction With Medical Applications to Functional Data Analysis." *Statistics in Medicine* 32, no. 30: 5222–5240.
- Tarpey, T., and K. K. Kinader. 2003. "Clustering Functional Data." *Journal of Classification* 20, no. 1: 93–114.
- Tekbudak, M. Y., M. Alfaro-Córdoba, A. Maity, and A.-M. Staicu. 2019. "A Comparison of Testing Methods in Scalar-on-Function Regression." *AStA Advances in Statistical Analysis* 103, no. 3: 411–436.
- Thind, B., K. Multani, and J. Cao. 2023. "Deep Learning With Functional Inputs." *Journal of Computational and Graphical Statistics* 32, no. 1: 171–180.
- Thomas Fletcher, P. 2013. "Geodesic Regression and the Theory of Least Squares on Riemannian Manifolds." *International Journal of Computer Vision* 105: 171–185.
- Tucker, J. D. 2024. "fdasrvf: Elastic Functional Data Analysis." R Package Version 2.2.0. <https://CRAN.R-project.org/package=fdasrvf>.
- Tutz, G., and J. Gertheiss. 2010. "Feature Extraction in Signal Regression: A Boosting Technique for Functional Data Regression." *Journal of Computational and Graphical Statistics* 19, no. 1: 154–174.
- Ullah, S., and C. F. Finch. 2013. "Applications of Functional Data Analysis: A Systematic Review." *BMC Medical Research Methodology* 13: 43.
- van den Boogaart, K. G., and R. Tolosana-Delgado. 2013. *Analyzing Compositional Data With R*. Heidelberg: Springer.
- van den Boogaart, K. G., J. J. Egozcue, and V. Pawlowsky-Glahn. 2010. "Bayes Linear Spaces." *SORT: Statistics and Operations Research Transactions* 34, no. 4: 201–222.
- van den Boogaart, K. G., J. J. Egozcue, and V. Pawlowsky-Glahn. 2014. "Bayes Hilbert Spaces." *Australian & New Zealand Journal of Statistics* 56, no. 2: 171–194.
- Vogel, F., N. M. Vahle, J. Gertheiss, and M. J. Tomasik. 2022. "Supervised Learning for Analysing Movement Patterns in a Virtual Reality Experiment." *Royal Society Open Science* 9: 211594.
- Volkman, A., A. Stöcker, F. Scheipl, and S. Greven. 2023. "Multivariate Functional Additive Mixed Models." *Statistical Modelling* 23, no. 4: 303–326.
- Volkman, A., N. Umlauf, and S. Greven. 2023. "Flexible Joint Models for Multivariate Longitudinal and Time-to-Event Data Using Multivariate Functional Principal Components." Preprint. <https://doi.org/10.48550/arXiv.2311.06409>.
- Wang, Q., S. Zheng, A. Farahat, S. Serita, T. Saeki, and C. Gupta. 2019. "Multilayer Perceptron for Sparse Functional Data." In *2019 International Joint Conference on Neural Networks (IJCNN)*, 1–10. Piscataway, NJ: IEEE.
- Wang, S., Y. Huang, and G. Cao. 2023. "Review on Functional Data Classification." *WIREs Computational Statistics* 16: e1638.
- Warmenhoven, J., A. Harrison, M. A. Robinson, et al. 2018. "A Force Profile Analysis Comparison Between Functional Data Analysis, Statistical Parametric Mapping and Statistical Non-Parametric Mapping in On-Water Single Sculling." *Journal of Science and Medicine in Sport* 21, no. 10: 1100–1105.
- Wood, S. N. 2011. "Fast Stable Restricted Maximum Likelihood and Marginal Likelihood Estimation of Semiparametric Generalized Linear Models." *Journal of the Royal Statistical Society Series B: Statistical Methodology* 73: 3–36.
- Wood, S. N., N. Pya, and B. Säfken. 2016. "Smoothing Parameter and Model Selection for General Smooth Models." *Journal of the American Statistical Association* 111: 1548–1575.
- Wood, S. N. 2017. *Generalized Additive Models: An Introduction With R*, 2nd ed. Boca Raton, FL: CRC Press.
- Wynne, G., and A. B. Duncan. 2022. "A Kernel Two-Sample Test for Functional Data." *The Journal of Machine Learning Research* 23, no. 1: 3159–3209.
- Xiao, L., Y. Li, and D. Ruppert. 2013. "Fast Bivariate P-Splines: The Sandwich Smoother." *Journal of the Royal Statistical Society Series B: Statistical Methodology* 75, no. 3: 577–599.
- Yao, F., H.-G. Müller, and J.-L. Wang. 2005. "Functional Data Analysis for Sparse Longitudinal Data." *Journal of the American Statistical Association* 100, no. 470: 577–590.
- Yi, M., Z. Li, and Y. Tang. 2022. "F-Type Testing in Functional Linear Models." *Stat* 11, no. 1: e420.
- Zhang, J.-T. 2011. "Statistical Inferences for Linear Models With Functional Responses." *Statistica Sinica* 21: 1431–1451.
- Zhang, J.-T. 2013. *Analysis of Variance for Functional Data*. Boca Raton, FL: CRC Press.
- Zhang, J.-T., and J. Chen. 2007. "Statistical Inferences for Functional Data." *The Annals of Statistics* 35, no. 3: 1052–1079.
- Zhang, J.-T., and X. Liang. 2014. "One-way ANOVA for Functional Data via Globalizing the Pointwise F-Test." *Scandinavian Journal of Statistics* 41: 51–71.

Zhang, J.-T., X. Liang, and S. Xiao. 2010. "On the Two-Sample Behrens-Fisher Problem for Functional Data." *Journal of Statistical Theory and Practice* 4: 571–587.

Zhang, J.-T., M. Y. Cheng, H. T. Wu, and B. Zhou. 2019. "A New Test for Functional One-Way ANOVA With Applications to Ischemic Heart Screening." *Computational Statistics & Data Analysis* 132: 3–17.

Zhang, M., and A. Parnell. 2023. "Review of Clustering Methods for Functional Data." *ACM Transactions on Knowledge Discovery From Data* 17, no. 7: 91.

Zhou, J., N.-Y. Wang, and N. Wang. 2013. "Functional Linear Model With Zero-Value Coefficient Function at Sub-Regions." *Statistica Sinica* 23, no. 1: 25–50.

Zhou, Y., S. Bhattacharjee, C. Carroll, et al. 2022. "fdapace: Functional Data Analysis and Empirical Dynamics." R Package Version 0.5.9. <https://CRAN.R-project.org/package=fdapace>.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.