JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION
2021, VOL. 00, NO. 0, 1–17: Theory and Methods
https://doi.org/10.1080/01621459.2021.1893179

Taylor & Francis
Taylor & Francis Group



# Bayesian Framework for Simultaneous Registration and Estimation of Noisy, Sparse, and Fragmented Functional Data

James Matuk[a], Karthik Bharath[b], Oksana Chkrebtii[a], and Sebastian Kurtek[a]

[a]Department of Statistics, The Ohio State University, Columbus, OH; [b]School of Mathematical Sciences, University of Nottingham, Nottingham, UK

**ABSTRACT**

In many applications, smooth processes generate data that are recorded under a variety of observational regimes, including dense sampling and sparse or fragmented observations that are often contaminated with error. The statistical goal of registering and estimating the individual underlying functions from discrete observations has thus far been mainly approached sequentially without formal uncertainty propagation, or in an application-specific manner by pooling information across subjects. We propose a unified Bayesian framework for simultaneous registration and estimation, which is flexible enough to accommodate inference on individual functions under general observational regimes. Our ability to do this relies on the specification of strongly informative prior models over the amplitude component of function variability using two strategies: a data-driven approach that defines an empirical basis for the amplitude subspace based on training data, and a shape-restricted approach when the relative location and number of extrema is well-understood. The proposed methods build on the elastic functional data analysis framework to separately model amplitude and phase variability inherent in functional data. We emphasize the importance of uncertainty quantification and visualization of these two components as they provide complementary information about the estimated functions. We validate the proposed framework using multiple simulation studies and real applications.

**ARTICLE HISTORY**
Received December 2019
Accepted February 2021

**KEYWORDS**
Amplitude and phase variability; Bayesian inference; Function estimation; Function registration; Square-root velocity function

## 1. Introduction and Motivation

Functional data analysis (FDA) is an area of statistics in which the primary objects of interest are more naturally understood as functions rather than vectors (Ramsay and Silverman 2005; Ferraty and Vieu 2006; Srivastava and Klassen 2016). This perspective is advantageous in a wide range of application domains including biology, medicine, environmental science and engineering, where the underlying evolution of variables of interest is often smooth. Though the goals of FDA, including sample summarization and visualization, inference, and prediction, are similar to those of multivariate statistics, they are more challenging due to the inherent difficulty of reconciling finite-dimensionality of observed data with the inherent infinite-dimensionality of function spaces representing the data space.

In particular, functional data objects are observed on a finite grid over of an interval $[a, b] \subset \mathbb{R}$, resulting in a variety of possible observational regimes: (i) *densely observed* functional data, wherein a large number of observations per function is available; (ii) *sparse* functional data, comprised of a small number of nonuniformly spaced observations per function; (iii) *fragmented* functional data, where observations of each function are unavailable over certain subsets of $[a, b]$. While Scenario (i) is often assumed for modeling (Ramsay and Silverman 2005; Srivastava and Klassen 2016), Scenarios (ii) and (iii) are very common in biomedical and industrial applications. Additionally, in each of these scenarios, observations are frequently measured with noise or error. Further complicating matters is the fact that functional data often exhibit two latent types of structural variability: amplitude variability, akin to differences in magnitude within variables in the multivariate setting, and phase or warping variability, related to differences in the timing of amplitude features, conspicuously absent in the multivariate setting. Failure to account for these sources of variability in function estimation and modeling can result in misleading summaries and biased inference (Marron et al. 2015).

Figure 1 illustrates such data (throughout this article and in the figures, we transform the domain to the unit interval without loss of generality). Panels (a) and (b) show fractional anisotropy (FA) data, extracted from diffusion tensor-magnetic resonance images (DT-MRIs), that are (a) fragmented early in the domain and (b) densely observed. Panel (c) shows densely measured, but noisy, growth velocity curves for a group of children, and panel (d) shows both fragmented and sparse observations of bone mineral density (BMD). These examples vary in the amount of sparsity, the degree of fragmentation, and the amount of noise.

Estimating the underlying functions, as well as their amplitude and phase components, in such diverse observational regimes is a challenging problem. While the problem has been tackled for specific observational regimes, to the best of our knowledge, it has not been studied under a unified methodological framework. Such a framework should exploit

**CONTACT** Sebastian Kurtek *kurtek.1@stat.osu.edu* Department of Statistics, The Ohio State University, Columbus, OH 43210.
Supplementary materials for this article are available online. Please go to *www.tandfonline.com/r/JASA*.

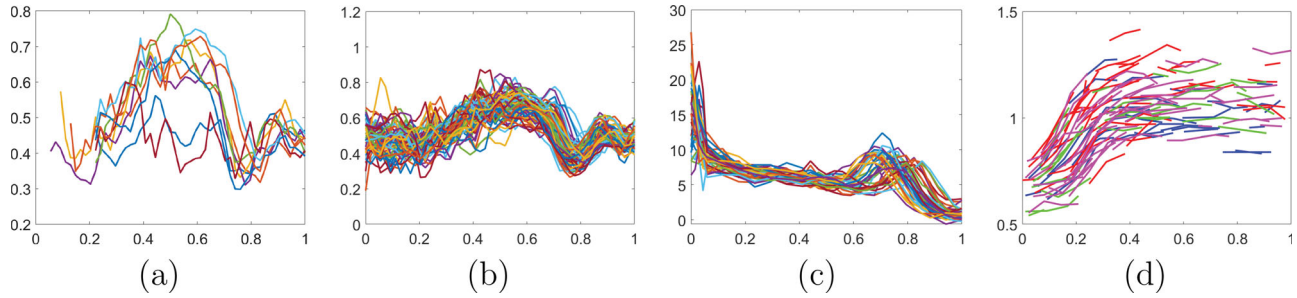


**Figure 1.** (a) Fragmented and (b) densely observed functional data from fractional anisotropy (FA) measurements. (c) Noisy growth velocity functions. (d) Fragmented and sparsely observed functions of bone mineral density (BMD).

the assumption that functional data share certain features, such as number of modes or inflection points. For example, in panels (a), (b), and (c) of Figure 1, the different functional observations contain similar numbers of modes, although their times of occurrence vary (phase variation) relative to some common state (template). The functions in panel (d) are generally known to be increasing and have an inflection point whose location varies across observations. Moreover, such a framework should automatically provide a calibrated assessment of uncertainty, and thus, quality of registration under different observational regimes. We propose a Bayesian approach that satisfies these desiderata for simultaneous estimation and registration of functional data.

The survey of FDA approaches in Section 1.1 highlights that existing methods frequently follow a sequence of estimation steps, which perform the necessary tasks of (i) smoothing (e.g., projecting data to a lower-dimensional function space), (ii) registration of smoothed functions (separation of amplitude and phase variability), and (iii) summarization, modeling and inference. While conceptually straightforward, this pipeline generally suffers from two drawbacks: difficulty in formally propagating uncertainty between stages of estimation leading to over-confidence in the results, and lack of flexibility under different observational regimes. We argue that both of these issues can be resolved by a unified Bayesian inferential framework that performs estimation, registration and inference simultaneously. Within this framework, data- or information-driven prior choices are necessary to extract meaningful observation-level information regardless of the collection protocol and/or issues with data quality, which often occur in practice.

## 1.1. Related Work

Both estimation and registration are crucial components in FDA. In most methods, the problems of estimation and registration are addressed separately, without formal uncertainty propagation from the estimation to the registration stage. Past methods that attempt to address this issue through a unified approach are either too rigid, or are unable to accommodate general observational regimes. Next, we review several different methods for function estimation and/or registration that are relevant to the proposed approach; for ease of exposition, we separate our discussion into three parts: methods that primarily focus on (i) estimation, (ii) registration, and (iii) joint estimation and registration.

### 1.1.1. Estimation

In FDA, most approaches assume that the observations used for statistical summarization, modelling and inference are finely sampled and smooth (Ramsay and Silverman 2005). However, in most real data scenarios this is not the case. The data comes in the form of discrete observations, possibly sparse or fragmented, and subject to measurement error. Consequently, a crucial step in FDA is to estimate a functional form from the discrete observations while accounting for observation noise, which relies on nonparametric function estimation. Namely, kernel smoothing and basis expansion methods, combined with penalized least squares, are extremely popular approaches for estimating underlying functions based on discrete observations (Wahba 1990; Ramsay and Silverman 2005; Srivastava and Klassen 2016). The primary challenge in these approaches is the selection of an appropriate smoothing parameter, which is commonly done through cross-validation. However, in high noise scenarios, individual estimation of functions using these methods tends to overlook population-level features, and fails to reflect the uncertainty in the observations in subsequent analyses.

To address these common issues with kernel smoothing and basis expansion approaches, methods have been developed to simultaneously infer functions underlying the observations, as well as population quantities of interest, for example, a population mean function. Functional mixed effects models have been effective in modeling discrete observations with a fixed mean function, random subject-specific functions, and pointwise error (Shi, Weiss, and Taylor 1996; Rice and Wu 2001; Yang et al. 2016). Descary and Panaretos (2019) modeled discrete, noisy observations as the sum of a stochastic process representing smooth, large scale-variation, and an uncorrelated, zero-mean stochastic process representing rough, small-scale variation. Inference in this setting enables joint covariance estimation for the processes as well as estimation of smooth functions, based on the discrete observations, through the Karhunen–Loéve expansion using the covariance of the smooth process. Shape-constrained estimation methods impose structure more directly, by incorporating prior knowledge of population-level features, for example, the number of peaks and valleys, to constrain the resulting shape of the estimated functions (Wheeler, Dunson, and Herring 2017). Similar methods have also been developed to estimate and model probability density functions under shape constraints (Dasgupta, Pati, and Srivastava 2019, 2020).

Estimation of functions based on sparse and fragmented observations, often arising in longitudinal studies, require

specific attention, because they violate the common assumption that a fine or dense sampling of discrete observations is available (Wang, Chiou, and Müller 2016). Estimation of subject-specific functions, along with the population mean and covariance, for fragmented functional data was studied in Delaigle and Hall (2016) and Kraus (2015). While these works implicitly assumed a missing-at-random mechanism for the unobserved parts of the function, recent work in Liebl (2019) performed estimation under a missing-not-at-random assumption. The work of Rice and Wu (2001) explicitly discusses fitting functional mixed effects models to sparse data. Principal Analysis through Conditional Expectation (PACE) (Yao, Müller, and Wang 2005), a framework for modeling sparse functional data, pools discrete observations across all samples to estimate a mean function and a covariance function; this enables individual function estimation based on sparse observations via the Karhunen–Loéve expansion.

#### 1.1.2. Registration

Works that focus solely on function estimation tend to overlook the distinction between phase and amplitude variability in functional data. As a result, registration is often considered as a sequential, unrelated step following estimation (Ramsay and Silverman 2005). Popular methods for registration of smooth, densely sampled functional data consist of landmark-, metric-, and model-based procedures. Landmark registration relies on the ability to specify landmarks, a set of points representing important function features, for example, peaks and valleys, and attempts to align these features (Kneip and Gasser 1992; Ramsay and Silverman 2005). A major disadvantage of these procedures is the need to select the landmark points, which can be subjective, inaccurate or even intractable (Marron et al. 2015). There have been some recent efforts to automatically identify, and quantify uncertainty in, mathematical landmarks (Strait, Chkrebtii, and Kurtek 2019). On the other hand, metric-based procedures formulate an optimality criterion, which is then maximized/minimized to estimate amplitude and phase components (Wang and Gasser 1997; Ramsay and Li 1998; Srivastava et al. 2011). The choice of optimality criterion is crucial and can drastically influence the registration result (Marron et al. 2015; Srivastava and Klassen 2016). Model-based registration procedures can be viewed as extensions of metric-based approaches where registration is determined through a formal statistical model (Tang and Müller 2008; Claeskens, Silverman, and Slaets 2010). These procedures have an added benefit in that they can be directly linked to common inferential tasks. Lu, Herbei, and Kurtek (2017) proposed a model that pools information from all available samples to estimate a common template, while registering the observations to this template. Kneip and Ramsay (2008) coupled registration with functional principal component analysis (fPCA) that enables dimension reduction.

#### 1.1.3. Joint Estimation and Registration

Joint estimation and registration procedures have been developed to propagate uncertainty from the estimation stage to the registration stage. Shape-invariant models (Lawton, Sylvestre, and Maggio 1972; Telesca and Inoue 2008; Fu and Heckman 2019) jointly estimate a population template function and subject-specific scaling factors, translations and warpings, based on discrete, noisy observations. While this approach does produce joint estimation and registration, while accounting for diverse observational regimes, the model is extremely rigid, as all estimated functions are required to have the same underlying shape, that is, the amplitude components are only scaled and translated versions of the template function. Shape-constrained approaches relax this assumption. The work of Rakêt, Sommer, and Markussen (2014) attempts to modify functional mixed effects models to address joint estimation and registration. In this approach, only the population mean component of the model is subject to warping, and assumes that all observations are recorded on a common time grid.

### 1.2. Contributions and Organization

The issues raised in Section 1.1 may be addressed by simultaneously combining coherent uncertainty propagation between estimation, registration and inference, enabled by prior specification for phase and amplitude components of individual functions. In particular, we argue that an inferential framework under a general observational regime must include an automatic method to appropriately restrict the model space (Rakêt, Sommer, and Markussen 2014; Wheeler, Dunson, and Herring 2017), based on either a physical understanding of the underlying process or on previously observed data. Bayesian inference is a natural framework for introducing such hard and soft constraints through the choice of a prior distribution. We focus on two automatic approaches for prior specification on subject-specific amplitude components to carry out this restriction:

1. When densely-observed training data is available (e.g., Figures 1(a) and (b)), a data-driven prior is constructed to capture important modes of amplitude variation.
2. When knowledge on the data-generating mechanism is available, such as the distribution of peaks and valleys (e.g., Figures 1(c) and (d)), we define a prior that informs the shape and smoothness by controlling the number of extrema.

We choose to model phase variation using a recently proposed point process-based prior on warping functions, compatible with the discrete nature of sparse and fragmented data (Bharath and Kurtek 2020). We acknowledge that prior training data or knowledge of extrema may not be available for all applications. In the absence of this information, additional hierarchical layers can be added to the proposed modelling framework to broaden the applicability. This is left as future work and discussed in Section 5.

While Bayesian models for functional data with phase variation have been considered under the sparse and dense observational regimes (e.g., Telesca and Inoue 2008; Zhou et al. 2014), we are unaware of any work that deals with estimation for fragmented functional data in the presence of phase variation. Moreover, the proposed model for handling amplitude and phase variation in the dense and sparse regimes adds flexibility relative to current approaches, by allowing for subject-specific templates rather than assuming a common shape template.

The remainder of this article is organized as follows. Section 2 describes an approach for prior modeling of amplitude and phase variability in densely observed functional data, which is used to inform the Bayesian model described in Section 3. We validate the proposed framework on simulated and real data in Section 4, and close with a brief discussion and some directions for future work in Section 5. The supplementary materials include (i) an additional simulation for the shape-restricted amplitude prior model, (ii) an additional real data example that considers estimation of CD4 cell count functions for HIV patients, (iii) an additional discussion of the BMD modeling results presented in Section 4.2.4, (iv) the Markov chain Monte Carlo (MCMC) algorithm used to sample from the posterior distribution and other implementation details, (v) MCMC diagnostics, and (vi) detailed sensitivity analyses with respect to various modeling choices.

## 2. Phase-Amplitude Separation via EFDA

Functional samples from a population often share common features, such as the number and relative location of peaks and valleys. Differences in magnitude and location of these features along the domain are commonly referred to as amplitude and phase variability, respectively, as formalized in Ramsay and Silverman (2005), Srivastava and Klassen (2016), and Marron et al. (2015). It has been shown in multiple places (Srivastava and Klassen 2016; Srivastava et al. 2011; Tucker, Wu, and Srivastava 2013) that modeling functional data without appropriately accounting for phase variability can result in inaccurate descriptive statistics and biased or misleading inference. Decoupling these two sources of variation is known as registration or alignment of functional data. For a sample of densely observed functions, $f_i : [0,1] \to \mathbb{R}$, $i = 1, \ldots, k$, a registration procedure yields corresponding amplitude functions $\tilde{f}_i : [0,1] \to \mathbb{R}$ and phase functions $\gamma_i$. Their composition $f_i = \tilde{f}_i \circ \gamma_i$, $i = 1, \ldots, k$ reconstructs the original functions uniquely. A formal registration procedure must thus define a representation space of phase (these can vary in flexibility from linear to diffeomorphic transforms) and an appropriate optimality criterion.

To perform simultaneous estimation and registration, coupled with informative shape restriction, we choose a warping functional form and optimality criterion based on the elastic functional data analysis (EFDA) framework of Srivastava et al. (2011). Indeed, its ability to define amplitude purely in terms of the shape of a function, that is, the number and (relative) heights of peaks and valleys independent of their locations on $[0,1]$, is consistent with our goal of semiparametric shape restriction. Phase variability is modeled via smooth, monotone transformations of $[0,1]$: $\Gamma = \{\gamma : [0,1] \to [0,1] \,|\, \gamma(0) = 0, \gamma(1) = 1, \dot{\gamma} > 0\}$, where $\dot{\gamma}$ is the time derivative. Though most other methods use the $\mathbb{L}^2$ metric as the optimality criterion for registration, this choice suffers from major deficiencies including the pinching effect (distorted amplitude) and asymmetry in registration (ill-defined amplitude) (Marron et al. 2015; Srivastava and Klassen 2016). This is because the $\mathbb{L}^2$ metric is not invariant to simultaneous warping of functions, making its use inconsistent with the goal of registration. Instead, EFDA uses an extension of the Fisher–Rao Riemannian metric as a foundation for registration and statistical modeling (Srivastava et al. 2011); we omit its formal definition here for brevity. While this choice has useful mathematical properties, including the critical invariance to identical warping, it is difficult to use directly. However, the simple transformation described below reduces this metric to the standard $\mathbb{L}^2$ metric, facilitating simplified computation.

Let $\mathcal{F} = \{f : [0,1] \to \mathbb{R} \,|\, f \text{ is absolutely continuous}\}$ denote the function space of interest. For any $f \in \mathcal{F}$, define its square-root velocity function (SRVF) using the mapping

$$Q : \mathcal{F} \to \mathbb{L}^2([0,1], \mathbb{R}), \quad Q(f) = \text{sign}(\dot{f})\sqrt{|\dot{f}|} := q\,.$$

The space of SRVFs corresponding to $\mathcal{F}$ is denoted as $\mathcal{Q}$. Because the SRVF is invertible up to a translation, the original function can be reconstructed from its SRVF using $f(t) = Q^{-1}(q, f(0)) = f(0) + \int_0^t q(s)|q(s)|ds$. The extended Fisher–Rao (eFR) distance on $\mathcal{F}$ simplifies to the $\mathbb{L}^2$ distance on $\mathcal{Q}$: the eFR distance between two functions $f_1, f_2 \in \mathcal{F}$ is computed as $d_{eFR}(f_1, f_2) := ||q_1 - q_2|| = \left[\int_0^1 (q_1(t) - q_2(t))^2 dt\right]^{1/2}$.

EFDA defines the amplitude of a function $f$ as an equivalence class $[f] := \{f \circ \gamma \,|\, \gamma \in \Gamma\}$ under the equivalence relation $f \sim g$ if there exists a warping $\gamma \in \Gamma$ such that $f \circ \gamma = g$. Equivalently, on the SRVF space $\mathcal{Q}$ we have $[q] := \{(q, \gamma) \,|\, \gamma \in \Gamma\}$, where $(q, \gamma) = (q \circ \gamma)\sqrt{\dot{\gamma}}$ is the corresponding action of $\Gamma$ on $\mathcal{Q}$ under the SRVF map $f \mapsto q$. The set of equivalence classes forms a partition of $\mathcal{Q}$, and is referred to as the quotient space $\mathcal{Q}/\Gamma$, that is, $\mathcal{Q}/\Gamma$ defines the amplitude space. Therefore, registration under this framework requires the determination of an average or mean equivalence class, and alignment of all functions to one of its elements. Let $f_1, \ldots, f_k$ denote a sample of densely observed functions, and $q_1, \ldots, q_k$ their corresponding SRVFs. The sample mean equivalence class is taken to be the Karcher mean: $[\hat{\mu}_q] = \underset{[q] \in \mathcal{Q}/\Gamma}{\text{argmin}} \sum_{i=1}^k d([q], [q_i])^2 = \underset{[q] \in \mathcal{Q}/\Gamma}{\text{argmin}} \sum_{i=1}^k \min_{\gamma \in \Gamma} ||q - (q_i, \gamma)||^2$. To ensure identifiability, a representative element of the mean equivalence class, $\hat{\mu}_q \in [\hat{\mu}_q]$, is chosen such that the average of the optimal warpings of all functions to $\mu_q$ is the identity warping $\gamma_{id}(t) = t$. In addition to the mean amplitude function, $\hat{\mu}_f = Q^{-1}(\hat{\mu}_q, \bar{f}(0))$ where $\bar{f}(0) = \frac{1}{k}\sum_{i=1}^k f_i(0)$, this registration procedure also yields (i) phase functions, $\gamma_i^* = \underset{\gamma \in \Gamma}{\text{argmin}} ||\hat{\mu}_q - (q_i, \gamma)||$, $i = 1, \ldots, k$, (ii) registered SRVFs, $(q_i, \gamma_i^*)$, $i = 1, \ldots, k$, and (iii) amplitude functions, $\tilde{f}_i = Q^{-1}((q_i, \gamma_i^*), f_i(0))$, $i = 1, \ldots, k$.

fPCA is a decomposition of the total sample variation into orthogonal modes of variability. We propose to perform fPCA on amplitudes to construct informative empirical priors over this component of variation. Also called vertical fPCA, it was first used within the EFDA framework by Tucker, Wu, and Srivastava (2013) in the context of building generative models for functional data. fPCA on the amplitude component corresponds to an eigendecomposition of the sample covariance function across the aligned SRVFs, $\widehat{K(s,t)} = \frac{1}{k-1}\sum_{i=1}^k ((q_i, \gamma_i^*)(s) - \hat{\mu}_q(s))((q_i, \gamma_i^*)(t) - \hat{\mu}_q(t)) = \sum_{b=1}^\infty \hat{\lambda}_b \hat{\phi}_b(s)\hat{\phi}_b(t)$ where $\hat{\phi}_b$, $b = 1, 2, \ldots$ are fPCs that form an orthogonal basis for the space of aligned SRVFs. With this data-driven basis, a finite representation of aligned SRVFs can be obtained through

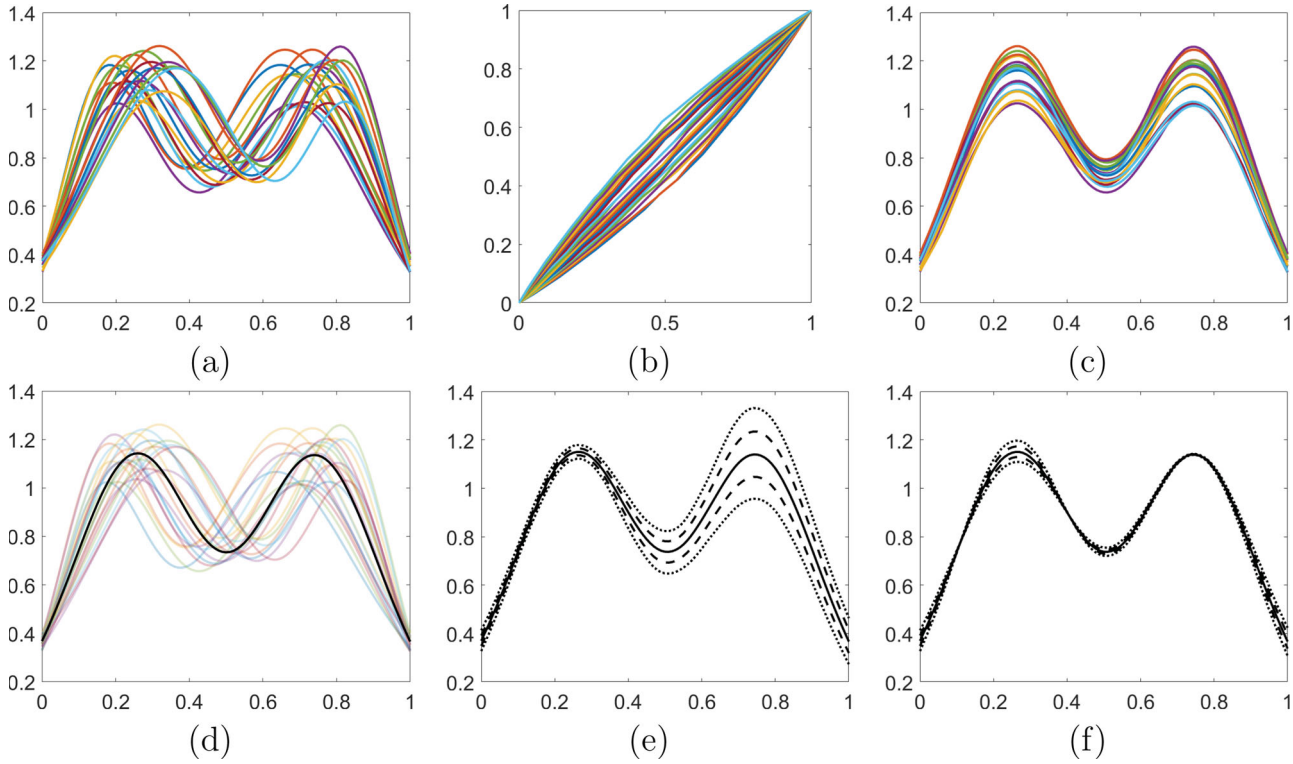


**Figure 2.** (a) Simulated data, (b) phase, and (c) amplitude functions. (d) Same as (a) with mean amplitude in bold. (e) First and (f) second principal modes of amplitude variability.

truncation as $(q_i, \gamma_i^*) \approx \hat{\mu}_q + \sum_{b=1}^{B} \hat{c}_{i,b}\hat{\phi}_b$, where $\hat{c}_{i,b} = \int_0^1 ((q_i, \gamma_i^*)(s) - \hat{\mu}_q(s))\hat{\phi}_b(s)ds$; $\hat{\mathbf{c}} = (\hat{c}_{i,1}, \ldots, \hat{c}_{i,B})^\top$ is a $B$-dimensional Euclidean representation of $(q_i, \gamma_i^*)$. Then, a generative model consists of (i) drawing a random coefficient vector $\mathbf{c} \sim \text{MVN}_B(0, \text{diag}(\hat{\lambda}_1, \ldots, \hat{\lambda}_B))$, (ii) constructing a random SRVF as $q_r = \hat{\mu}_q + \sum_{b=1}^{B} c_b\hat{\phi}_b$, and (iii) computing the corresponding function $f_r = Q^{-1}(q_r, T)$, where $T \sim N(\bar{f}(0), \hat{\tau}^2)$ is a random translation, and $\hat{\tau}^2$ is the sample variance of the function starting points $f_1(0), \ldots, f_k(0)$. Thus, the resulting random function $f_r$ is necessarily aligned to the mean amplitude function $\hat{\mu}_f$.

To illustrate EFDA, we consider a sample of simulated functions shown in Figure 2(a). The data are generated through the composition of the following amplitude and phase functions: for $i = 1, \ldots, 21$ and $t \in [0, 1]$,

$$\tilde{f}_i(t) = z_{1,i} \exp\{-((t - .25)^2/72)\} + z_{2,i} \exp\{-((t - .75)^2/72)\}, \quad \text{and} \tag{1}$$

$$\gamma_i(t) = \begin{cases} \frac{\exp(a_i t) - 1}{\exp(a_i) - 1} & a_i \neq 0, \\ t & \text{otherwise}, \end{cases} \tag{2}$$

where $z_{1,i}, z_{1,i} \overset{\text{iid}}{\sim} N(1, 0.25^2)$, and $a_i$ are equally spaced on $[-1, 1]$. Such simulated functions are commonly used to assess performance of models for functional data in the presence of phase variation (Kneip and Ramsay 2008; Srivastava et al. 2011; Tucker, Wu, and Srivastava 2013), and are used in simulation experiments in this work throughout Section 4, and in Section 6 of the supplementary materials. The phase and amplitude functions extracted via the eFR-based registration procedure are shown in panels (b) and (c), respectively, with the amplitude mean shown in bold in (d) on top of the original functions. The first two principal modes of variability for the amplitude component are given in panels (e) and (f), respectively. That is, we plot $Q^{-1}(\hat{\mu}_q + k\hat{\lambda}_b^{1/2}\hat{\phi}_b, T)$ for $b \in \{1, 2\}$, $k \in \{-2, -1, 0, 1, 2\}$; $T$ is a small translation that is included for improved display. The primary mode of variability captures differences in the height of the right peak, while the second mode describes variability in the left peak, and to a lesser extent, the relative heights of the left and right peaks. The remaining modes of variability, which are not displayed here, capture a negligible amount variability.

## 3. Model-Based Estimation and Registration

We summarize the notation used thus far and introduce notation for discretely observed functional observations, which become key at the modeling and implementation stages. Let $\mathcal{F}$ be the set of absolutely continuous functions $f : [0, 1] \to \mathbb{R}$, $\mathcal{Q}$ the corresponding SRVF space under the mapping $f \mapsto Q(f) = \text{sign}(\dot{f})\sqrt{|\dot{f}|} := q$, and $T \in \mathbb{R}$ a scalar translation. The function $Q^{-1} : \mathcal{Q} \times \mathbb{R} \to \mathcal{F}$ takes in an SRVF and a translation, and maps them to the corresponding point $f$ in the function space $\mathcal{F}$. A warping function is denoted by $\gamma \in \Gamma$, where $\Gamma = \{\gamma : [0, 1] \to [0, 1] \,|\, \gamma(0) = 0, \gamma(1) = 1, \dot{\gamma} > 0\}$.

A model for functional data can be defined either on the original function space $\mathcal{F}$ or on the SRVF space $\mathcal{Q}$. While modeling phase variation through the usual warping action $f \mapsto f \circ \gamma$ is simpler on $\mathcal{F}$, the simple $\mathbb{L}^2$ geometry of $\mathcal{Q}$ makes dimension reduction through a (possibly data-driven) basis expansion possible. We choose a marriage of the two

options, and assume that functional observations are generated under the model,

$$y_i(t) = (f_i \circ \gamma_i)(t) + \epsilon_i(t),$$
$$y_i(t) = (Q^{-1}(q_i, T_i) \circ \gamma_i)(t) + \epsilon_i(t), \quad (3)$$

where $T_i \in \mathbb{R}$ accounts for function translation, $\epsilon_i$ denotes the error process, and the subscript $i$ indexes each sample element (e.g., subject). Thus, individual curves $y_i$ are viewed as perturbations of a function $f_i := Q^{-1}(q_i, T_i)$ warped by $\gamma_i$. *This allows us to specify structured prior distributions on the amplitude component based on a judicious choice of shape-defining basis functions, while specifying a prior distribution on warping functions* $\Gamma$ *that is naturally compatible with possible fragmentation and sparsity.* This model is characterized by subject-specific amplitude (and phase) components, in contrast to a common amplitude template as done in Telesca and Inoue (2008) and Zhou et al. (2014). We let $\mathbf{y} = (y_1, \ldots, y_m)^\top$ and $\mathbf{t} = (t_1 < \cdots < t_m)^\top$ denote a vector of noisy, discrete functional observations and the corresponding time grid on which this data was observed, respectively. Then, for a function $f$, $f(\mathbf{t}) = (f(t_1), \ldots, f(t_m))^\top$ is a vector of function evaluations on this same time grid.

### 3.1. Likelihood Specification

We now formulate error model (3) for discretely observed functions, as is typical in practice. The vector of observations of $f_i$, denoted by $\mathbf{y}_i$, is obtained from evaluations over a subject-specific grid of time points $\mathbf{t}_i = (t_{i,1} < \cdots < t_{i,m_i})^\top$ of an appropriately translated (via $T_i$) and warped (via $\gamma_i$) SRVF $q_i$ specifying the overall shape, as

$$\mathbf{y}_i = (Q^{-1}(q_i, T_i) \circ \gamma_i)(\mathbf{t}_i) + \epsilon_i(\mathbf{t}_i), \quad i = 1, \ldots, n. \quad (4)$$

This formulation allows general observational regimes, such as when $\mathbf{t}_i$ is sparse or fragmented. Under the assumption $\epsilon_i(\mathbf{t}_i) \overset{\text{ind}}{\sim} \text{MVN}_{m_i}(0, \sigma_i^2 I_{m_i})$ (in most cases, one can fix all $\sigma_i^2$ to a common value), the likelihood for a single observation $\mathbf{y}_i$ under our model is given by

$$\mathbf{y}_i | q_i, \gamma_i, T_i, \sigma_i^2 \sim \text{MVN}_{m_i}\left((Q^{-1}(q_i, T_i) \circ \gamma_i)(\mathbf{t}_i), \sigma_i^2 I_{m_i}\right), \quad i = 1, \ldots, n. \quad (5)$$

### 3.2. Prior Models for Translation, Phase, and Error Variance

We first focus on model components whose probable values a-priori may be assumed relatively similar between problems. We assume that the translation parameters, $T_i$, $i = 1, \ldots, n$, and error variances, $\sigma_i^2$, $i = 1, \ldots, n$ are a-priori independent with $T_i \sim N(\mu_T, \tau^2)$ and $\sigma_i^2 \sim \text{Inverse-Gamma}(\alpha_\sigma, \beta_\sigma)$. The choice of hyperparameters $\tau^2$, $\alpha_\sigma$ and $\beta_\sigma$ is largely problem-specific. In all applications, we use $\mu_T = 0$.

Defining a prior process on the set $\Gamma$ of warping functions is more challenging due to its restricted functional form. Common approaches in the literature model phase functions using basis expansions with constrained coefficients (Telesca and Inoue 2008; Fu and Heckman 2019) or directly via the Riemannian geometry of the group of warping functions (Lu, Herbei, and Kurtek 2017). In contrast, we use a prior based on a piecewise linear process proposed in Bharath and Kurtek (2020), consisting of random phase increments $p(\gamma_i) = (\gamma_i(t_1), \ldots, \gamma_i(t_j) - \gamma_i(t_{j-1}), \ldots, 1 - \gamma_i(t_{m_\gamma}))^\top$ over $m_\gamma$ successive time points $\mathbf{t}_\gamma = (t_1 < \cdots < t_{m_\gamma})^\top$ on the input domain $[0, 1]$. The partition size is a user-specified value $m_\gamma \leq \min_i m_i$, whose magnitude provides a trade-off between model flexibility and computation time. For each $i = 1, \ldots, n$, independently, the finite-dimensional vector of phase increments follows a Dirichlet distribution, $p(\gamma_i) \sim \text{Dirichlet}(\theta_\gamma \mathbf{u})$, where $\mathbf{u} = (u_{(1)}, \ldots, u_{(j)} - u_{(j-1)}, \ldots, 1 - u_{(m_\gamma)})$ and $0 =: u_{(0)} < u_{(1)} < \cdots < u_{(m_\gamma)} < 1$ is a vector of order statistics corresponding to a random sample of size $m_\gamma$ from a uniform distribution on $[0, 1]$, and $\theta_\gamma$ acts as a precision parameter. This finite-dimensional model specification is computationally convenient; furthermore, Bharath and Kurtek (2020) show that it has desirable asymptotic properties by relating it to a stochastic process with (measurably) increasing sample paths as $m_\gamma \to \infty$.

In this prior specification, the choice of the order statistics dictates where the prior process is centered. For example, uniform order statistics result in a prior distribution over $\Gamma$ centered at the identity warping $\gamma_{id}$; order statistics from an arbitrary distribution $G$ on $[0, 1]$ would result in the prior distribution over $\Gamma$ being centered at $G^{-1}$. We choose the uniform order statistics to regularize the prior model toward the identity warping. We evaluate the prior approximately by assigning the random order statistics to a uniform spacing of the domain grid, since the successive spacings in both cases are essentially of order $O(1/m_\gamma)$. The precision hyperparameter $\theta_\gamma$ controls the spread of the prior around its mean $G^{-1}$: a small value results in a diffuse-type prior, whereas a large value concentrates the prior around $G^{-1}$. When the observed data contains information about the location of prominent function features, for example, peaks and valleys, the posterior distribution is not very sensitive to the choice of this parameter. On the other hand, when the available data contains less information about the underlying signal, a large $\theta_\gamma$ induces regularization of estimated phase to the identity element. In other words, the choice of $\theta_\gamma$ is problem specific and should in part depend on the observational regime, that is, when the observed data is densely sampled, $\theta_\gamma$ can be small, but when faced with sparse and/or highly fragmented data, a considerable amount of regularization is required to identify subject-specific model components; see Section 6 in the supplementary materials for detailed sensitivity analyses. The role of the parameter $\theta_\gamma$ in this phase prior is detailed in Bharath and Kurtek (2020).

### 3.3. Shape-Informed Prior Models for Amplitude

Besides allowing direct interpretation of phase and amplitude uncertainty, modeling underlying functions in model (4) via the EFDA framework allow us to define the amplitude model on the mathematically convenient SRVF representation space. Choice of the prior model over the registered SRVFs, representing amplitude, is both problem-specific and has a potentially strong impact on posterior inference under general observational regimes. We propose two semiautomatic approaches to

model prior information about amplitude in different data collection scenarios. The first type is appropriate when, along with data that is sparsely observed or fragmented, we also have access to functional observations of different subjects from the same population that are neither fragmented nor sparse; we refer to this data as training and define an empirical prior model based on statistics computed from this data. The second type of prior is appropriate when information is available about the number and relative locations of local extrema that underlie the functional observations. Importantly in practice, this prior model does not require the location of the extrema on the domain. We will also explain how this second scenario has connections to landmark registration (Kneip and Gasser 1992).

#### 3.3.1. Empirical Amplitude Prior

When densely observed training data $f_1, \ldots, f_k$ is available, we can construct a shape-informed prior for the amplitude of $n$ new, partially observed functions on the subspace spanned by the empirical basis constructed from the amplitudes of training data SRVFs $q_1, \ldots, q_k$. Elastic fPCA is carried out by first jointly performing registration and computation of the Karcher mean $\hat{\mu}_q$, and then decomposing the sample covariance of the amplitude components of the training data to obtain eigenfunctions $\{\hat{\phi}_b,\ b = 1, \ldots, B\}$ and corresponding eigenvalues $\hat{\lambda}_b,\ b = 1, \ldots, B$. The basis functions $\hat{\phi}_b$ represent amplitude variation about the Karcher mean $\hat{\mu}_q$ so that the SRVFs $q_i,\ i = 1, \ldots, n$ in model (3) can be represented as

$$q_i(t) = \hat{\mu}_q(t) + \sum_{i=b}^{B} c_{i,b}\hat{\phi}_b(t), \quad i = 1, \ldots, n. \tag{6}$$

Thus, a prior process over the amplitude component $q_i$ can be defined through a prior distribution over the coefficient vector $\mathbf{c}_i$, such as $\mathbf{c}_i \sim \text{MVN}_B(0_B, \text{diag}(\hat{\lambda}_1, \ldots, \hat{\lambda}_B))$. This prior specification provides a data-driven way of imposing structure on the amplitude model based on the training data. Note that the training observations must be representative of the important features of amplitudes in the population, and their number must be sufficient to be able to estimate these. A similar approach was recently used to model shapes of three-dimensional curves representing fiber tracts extracted from diffusion tensor-magentic resonance imaging (DT-MRI) (Zhang, Descoteaux, and Dunson 2019).

The proposed prior model is illustrated in Figures 2 and 3 for a simulated training dataset. Figure 3(a) shows the sample mean SRVF (black) and fPCA basis computed using the amplitude functions in Figure 2(c). SRVF draws from the prior, and corresponding amplitude functions obtained by the transformation $Q^{-1}$, are shown in Figures 3(b) and (c), respectively. We illustrate visually that these random functions are registered to the amplitude functions in the training data, as required.

#### 3.3.2. Shape-Restricted Amplitude Prior

When reliable information about the number and relative ordering of extrema of underlying functions is available, a prior on the amplitude $q_i$ can be specified by choosing a set of basis functions that reflects this information. In contrast to the empirical amplitude prior, such a prior does not require training data. We assume the following form for the SRVFs $q_i$:

$$q_i(t) = \sum_{b=1}^{B} c_{i,b}U_b^*(t), \quad i = 1 \ldots, n. \tag{7}$$

The basis functions are defined as $U_b^*(t) = M(\prod_{h=1}^{H}(t - \alpha_h))U_b(t),\ b = 1, \ldots, B$, where $U_b$ are B-spline basis functions. This basis system is based on a modification of the shape-restricted B-splines developed by Wheeler, Dunson, and Herring (2017). These bases relate to the derivative of the function since amplitudes are defined on the SRVF space rather than the original function space. This basis system forces the SRVFs to be exactly zero at the change point locations $\alpha_1, \ldots, \alpha_H$, corresponding to extreme values at these locations in the corresponding amplitude functions. The constant $M \in \{-1, 1\}$ is application-specific and defines the order of extreme values.

Our use of this basis system differs from Wheeler, Dunson, and Herring (2017) as our model accounts for phase variability explicitly, that is, we treat the change points as constants that determine the common locations of extreme values of the underlying amplitude functions, unlike the original work where the change points are also inferred. We use a diffuse exponential prior model on the coefficient vector $\mathbf{c_i}$ to ensure that each basis coefficient is positive; this, in turn, guarantees that all of the amplitude functions have the same ordering of extrema. Specifically, the prior model in this case is $c_{i,b} \overset{\text{ind}}{\sim} \text{Exp}(\beta_c)$, with a large $\beta_c$. For simplicity in implementation of this model,

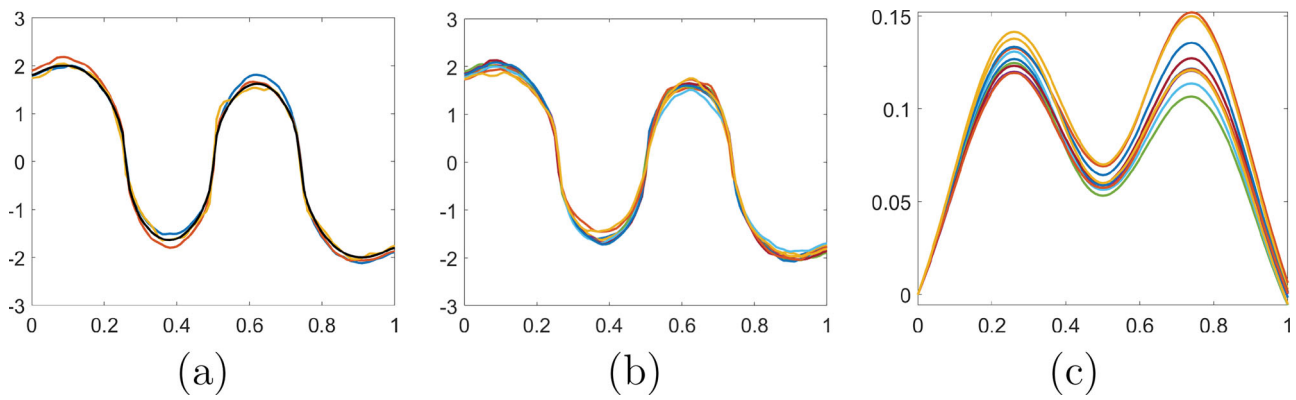


**Figure 3.** (a) Sample mean SRVF (black) and fPCA basis elements computed using the training data in Figure 2. (b) SRVF draws, and (c) corresponding amplitude functions transformed under $Q^{-1}$, generated from the proposed empirical amplitude prior model.

the size of the discretization grid for the phase functions, $m_\gamma$, is set equal to $H$, that is, the number of extrema. This ensures that the derivative changes in the phase functions occur only at the extrema. To provide maximum separation of extrema in the estimated amplitude functions, we recommend choosing the change points $\alpha$ as equally spaced over the entire domain. In some settings, one can alternatively use the cross-sectional average of previously observed data, akin to training data, to approximately determine good values of $\alpha$.

The shape-restricted amplitude prior model has a clear connection to landmark-based registration (Kneip and Gasser 1992) in which one must first identify a set of common landmarks on each function in a dataset. The landmarks either correspond to mathematical features of the data, for example, extrema, or application-specific, interpretable features. Given a set of landmarks on each function, the functions are registered to each other via a piecewise linear warping that aligns the landmark locations exactly. If only extrema are considered as landmarks, then the change points in the shape-restricted amplitude prior act as domain locations for landmark alignment. In general, selecting appropriate landmarks can be challenging, especially when the number of functions and/or landmarks is large. The process of selecting nonmathematical landmarks can also be highly subjective. Recently, Strait, Chkrebtii, and Kurtek (2019) developed an automated approach for mathematical landmark selection that alleviates the aforementioned issues.

Figure 4 illustrates the shape-restricted amplitude prior and the structure it enforces on the amplitude component in our observation model. Panel (a) depicts the basis system that is used to represent the SRVFs. Panels (b) and (c) show several prior draws of SRVFs and corresponding amplitude functions, respectively. The basis system is constrained to take zero values at the change point locations. Combined with the appropriate restrictions on $M$ and the basis coefficients, this pattern propagates to the generated SRVFs. Zeros of SRVFs map to extrema of amplitude functions; the relative heights of the extrema are flexible.

Both methods proposed here for informing the amplitude model consist of constructing a set of meaningful basis elements, either empirically based on training data or from prior shape information concerning the number and relative location of extrema. The shape-restricted amplitude prior relies on the practitioner to specify the number and pattern of extreme values, while the empirical amplitude prior achieves this automatically, but requires training data. Another difference is that functions in the amplitude space spanned by the empirical basis are aligned with respect to the eFR metric, while elements of the shape-restricted amplitude space are only registered based on the extrema. In this sense, the empirical basis is far more informative than the shape-restricted basis as illustrated in Figures 3 and 4.

## 4. Simulations and Real Data Applications

In this section, we discuss several simulated and real data examples, which illustrate the performance of the proposed framework. We compare our methodology to PACE, implemented via a publicly available software package (Yao, Müller, and Wang 2005; Tang and Müller 2008). The PACE framework is the most appropriate for comparison to our approach as it is able to accommodate sparsity, pointwise noise, and phase and amplitude variability in functional data. In brief, the PACE approach uses three steps: (i) estimation of population parameters, such as the mean and principal components, based on noisy and sparse observations, (ii) estimation of fitted functional observations on a common domain grid based on the parameter estimates from (i) (Yao, Müller, and Wang 2005), and (iii) registration of the estimated functions via a penalized $\mathbb{L}^2$ metric criterion (Tang and Müller 2008). There are two versions of the procedure, one for observations recorded over a common grid of domain points and one for observations recorded on different grids. When observations are recorded on a common grid, the PACE procedure begins with Steps (i)–(iii) (Step (iii) results in the final phase functions). Then, the estimate from Step (ii) is disregarded; instead, the phase functions are applied to the original noisy observations followed again by Steps (i) and (ii) to estimate the final amplitude functions. In this setting, a single estimate of the original observations is generated through composition of the final phase and amplitude estimates. On the other hand, when observations are recorded on different domain grids, following a first iteration through Steps (i)–(iii) (Step (iii) here results in a first estimate of amplitude functions

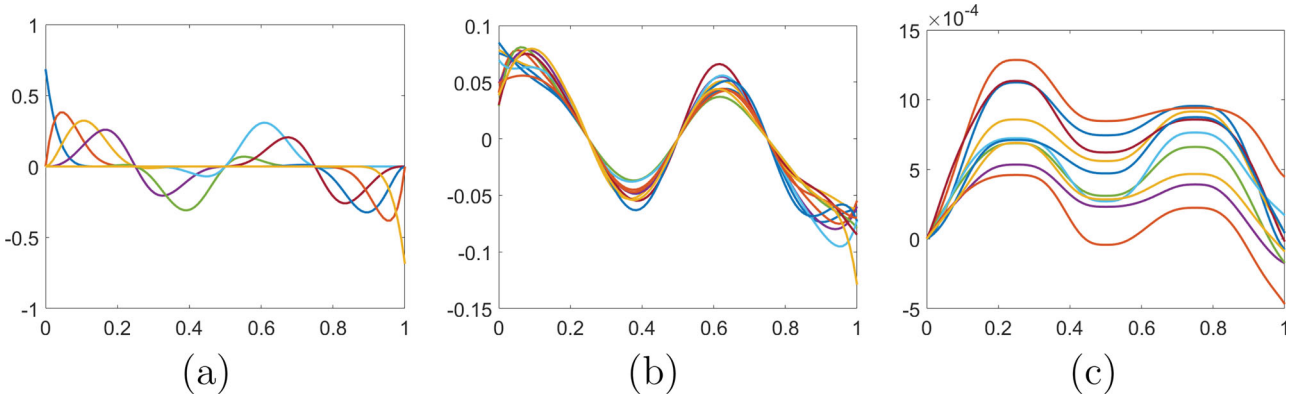


**Figure 4.** (a) Shape-restricted spline basis. (b) SRVF draws, and (c) corresponding amplitude functions, generated from the proposed shape-restricted amplitude prior model with $H = 3$, $\alpha = (0.25, 0.5, 0.75)$, $M = -1$.

and the final phase functions), Steps (i) and (ii) are applied again to the registered estimated functions to extract a second estimate of amplitude. Thus, one can actually obtain two different estimates: the first through composition of the final phase functions and the first estimate of amplitude functions (termed PACE in results), and the second through composition of the final phase functions and the second estimate of amplitude functions (termed WPACE in results). All tuning parameters are set by default in the package. In the simulation examples, we consider three estimators of a sparsely observed or fragmented function $f$ from a posterior MCMC sample $\{(q^{[j]}, T^{[j]}, \gamma^{[j]}),\ j = 1, \ldots, N\}$ after burn-in:

1. The plug-in estimator $\hat{f}_{\text{plug-in}} = \left[\frac{1}{N}\sum_{j=1}^{N} Q^{-1}(q^{[j]}, T^{[j]})\right] \circ \left[\frac{1}{N}\sum_{j=1}^{N} \gamma^{[j]}\right]$.
2. The maximum a-posteriori (MAP) estimator $\hat{f}_{\text{MAP}} = Q^{-1}(q^{[p]}, T^{[p]}) \circ \gamma^{[p]}$ where $p$ is the index with the largest unnormalized log-posterior density.
3. The pointwise estimator $\hat{f}_{\text{pointwise}} = \frac{1}{N}\sum_{j=1}^{N}\left[Q^{-1}(q^{[j]}, T^{[j]}) \circ \gamma^{[j]}\right]$.

The performance of each estimator is compared with the PACE and WPACE estimators.

### 4.1. Bayesian Model With Empirical Amplitude Prior

We begin the results section with a few examples that use the empirical amplitude prior in the proposed framework to estimate and register fragmented and sparse functions.

#### 4.1.1. Simulated Example 1: Fragmented Simulated Functions

Figures 5(a) and (b) show simulated training data and a fragmented functional observation (black points) that we wish to infer, respectively. Within our unified framework, we employ an empirical amplitude prior with $\theta_\gamma = .1$, $m_\gamma = 8$ and $B = 8$, to infer the full underlying function. In this case, we use a diffuse prior over the phase component by setting $\theta_\gamma$ to a small value. Figure 5 shows posterior marginal samples over the amplitude (c) and phase (d) components, and their composition (e); the solid black line in each panel represents the pointwise posterior mean. This display allows us to assess uncertainty in the different components underlying the estimated function. Additionally, in (e), we show the PACE and WPACE estimates in red and blue, respectively, for comparison. All three approaches provide reasonable estimates with a peak along the missing portion of the function; this is in agreement with the training data, and the estimated peaks across the three different methods are in roughly similar regions of the domain.

A major advantage of the proposed approach over PACE and WPACE is its ability to assess uncertainty in the estimated function. As evidenced by the credible bands, uncertainty is smaller along the portion of the function where data was observed than the portion that was not observed, both in terms of the height and location of the missing peak.

To provide a quantitative comparison of our method with PACE/WPACE, we simulated 100 missing portions of different functions from the complete data in Figure 5(a). In each simulation, one function from the complete dataset was selected at random and fragmented from $t = 0$ up to a random point

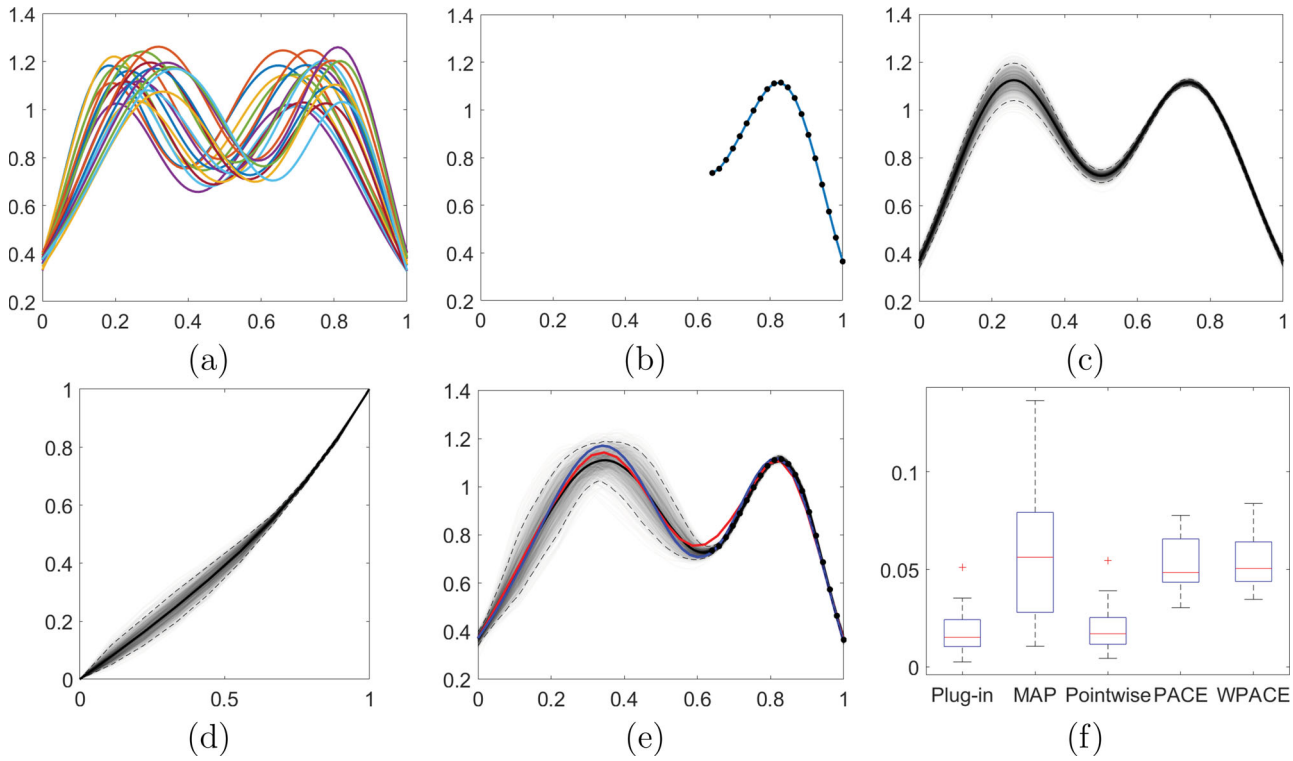


**Figure 5.** (a) Simulated training data. (b) Fragmented simulated function with observations shown as black points. Posterior draws (transparent), pointwise mean (solid black) and 95% credible interval (dashed) for the (c) amplitude and (d) phase components. (e) Composition of amplitude and phase, and PACE (red) and WPACE (blue) estimates. (f) Boxplots of $\mathbb{L}^2$ distances between a true function and corresponding estimated function. Each boxplot corresponds to a different approach: (i) plug-in, (ii) MAP and (iii) pointwise estimators based on posterior samples from the proposed Bayesian model, and (iv) PACE and (v) WPACE.

on [0, 1] drawn from a Beta(25, 25) distribution; the remaining functions were treated as training data. The observation set in each simulation consisted of 20 evenly spaced points along the nonfragmented part of the randomly selected function. As our inferential approach is based on the posterior distribution over unknown model components, we consider the three estimators $\hat{f}_{\text{plug-in}}$, $\hat{f}_{\text{MAP}}$, and $\hat{f}_{\text{pointwise}}$ described above. We compare our performance to PACE/WPACE by computing the $\mathbb{L}^2$ distance between the true function and the estimate. Boxplots of these distances for each method are shown in Figure 5(f). The MAP estimator appears to be comparable in performance to PACE and WPACE. However, both the plug-in and pointwise estimators outperform PACE and WPACE. The plug-in estimator outperforms PACE and WPACE in 99% and 100% of the simulations, respectively, while the pointwise estimator outperforms both PACE and WPACE in 99% of the simulations. We remark that the behavior of the MAP estimate can be sensitive to estimation of the phase component. In the fragmented region, where no values of the function are observed, the posterior distribution is driven largely by the diffuse phase prior. Consequently, the MAP estimate generally fits the data very well in the nonfragmented region, but exhibits random warping around the identity element of $\Gamma$ in the fragmented region; this results in misalignment between the estimated and true functions in the fragmented region, which is greatly penalized by the $\mathbb{L}^2$ distance (Marron et al. 2015). That said, the estimated *amplitude* portion of the MAP samples is very accurate and reflects the shape of the true function in terms of the heights of the peaks and valley. This observation suggests that one should use the plug-in and pointwise estimators for the phase component in the presence of significant fragmentation in the observed data.

#### 4.1.2. Simulated Example 2: Sparse ECG Signals

The electrocardiogram (ECG) is an important diagnostic tool for many conditions including myocardial infarction. It records fluctuations in electrical potential of the heart muscle on the body surface. Often, one studies the shape of PQRST complexes extracted from a long ECG signal, which can be associated with abnormal heart function (Kurtek et al. 2013). The letters PQRST refer to the first peak (P wave), the shallow, deep valley followed by the sharp second peak and another valley (QRS complex), and finally the third peak (T wave). In this simulated example, we study the performance of the proposed Bayesian model in the context of PQRST complex estimation from sparse observations. Figure 6 shows a set of training signals in (a) as well as a sparse set of 10 evenly spaced data points extracted from a known PQRST complex (not part of the training data) in (b); such a setup allows us to again assess the performance of our approach qualitatively and quantitatively. We employ an empirical amplitude prior under the settings $\theta_\gamma = 100$, $m_\gamma = 8$, and $B = 10$. Here, due to the high level of sparsity in the observed data, we use a phase prior that is tightly concentrated around the identity element of $\Gamma$. We display marginal posterior samples in Figure 6 for amplitude (c), phase (d) and their composition (e), describing posterior uncertainty in the unknown PQRST complex function. It turns out that the posterior distribution

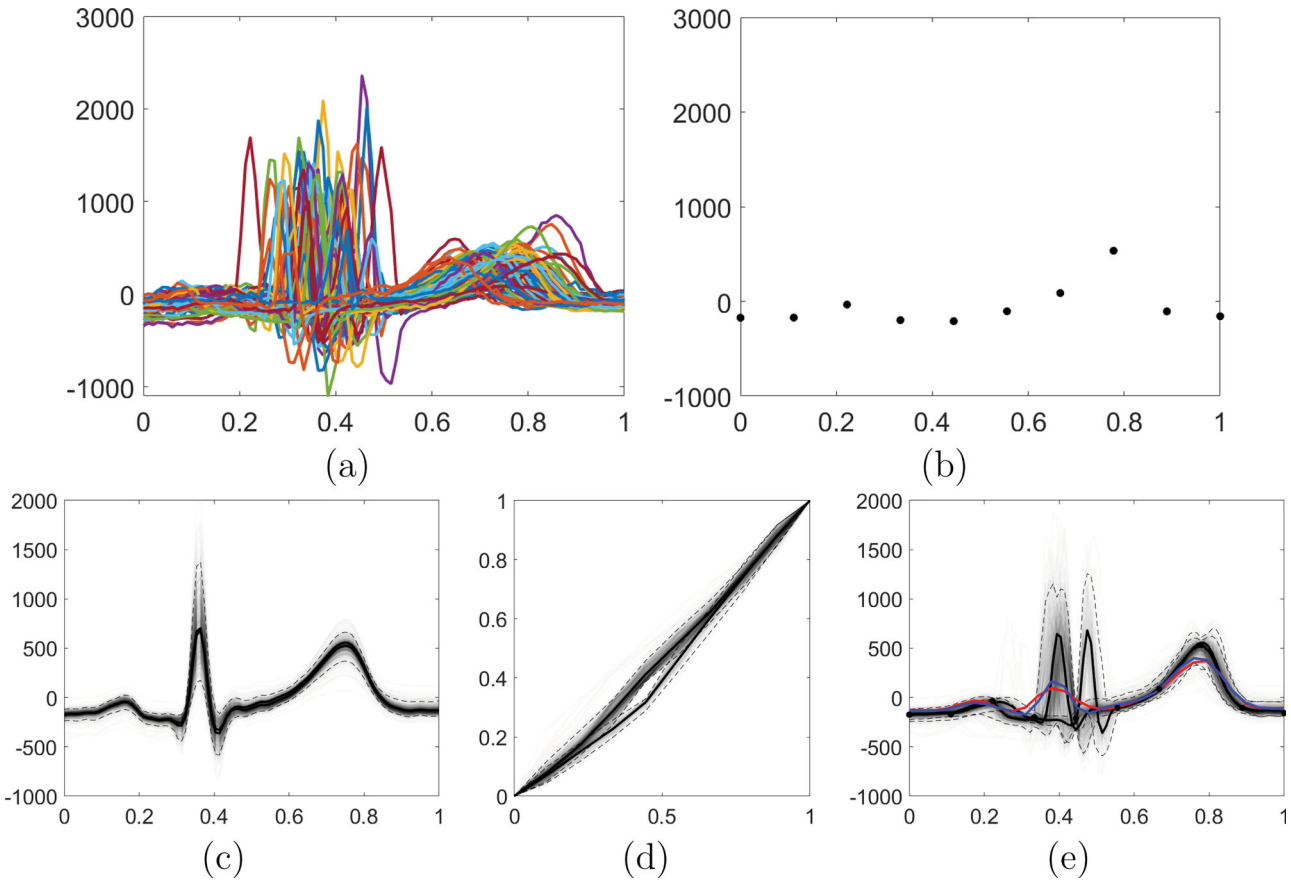


**Figure 6.** (a) ECG training data. (b) Sparsely observed ECG signal. Posterior draws (transparent), pointwise mean (solid black) and 95% credible interval (dashed) for the (c) amplitude and (d) phase components. (e) Composition of amplitude and phase, and PACE (red) and WPACE (blue) estimates.

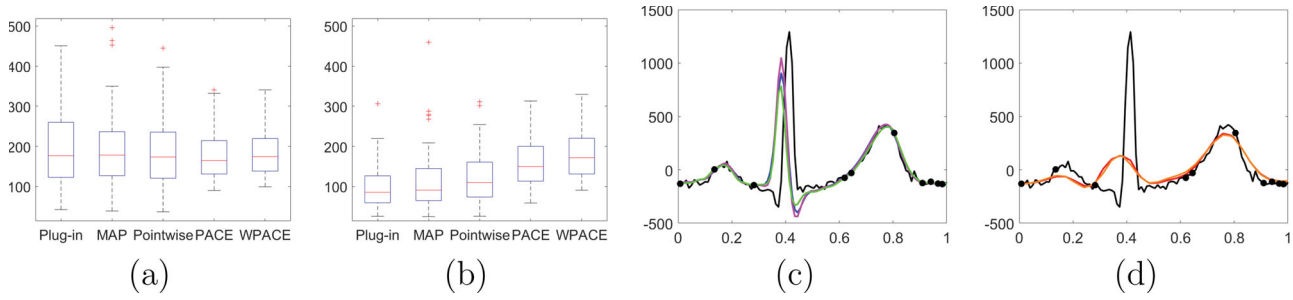


**Figure 7.** (a) Boxplots of $\mathbb{L}^2$ distances between a true function and estimated function. Each boxplot corresponds to a different approach: (i) plug-in, (ii) MAP and (iii) pointwise estimators based on posterior samples from the proposed Bayesian model, and (iv) PACE and (v) WPACE. (b) Same as (a), but for the amplitude component only. (c) The true function (black) and corresponding sparse observations (black points), and plug-in (magenta), MAP (green) and pointwise estimates (blue). (d) Same as (c), but with PACE (red) and WPACE (orange) estimates.

in this case is bimodal, with two modes formed by the phase functions. Thus, we display modewise summaries for the phase sample in panel (d) and the composition of amplitude and phase in (e). Again, the means are shown in bold black and the 95% confidence bands as dashed lines. The two modes correspond to two plausible locations of the QRS complex given the data. Indeed, none of the sparse observations cover the sharp R peak making its location difficult to estimate. Importantly, the structure of the estimated PQRST complex based on each mode of the posterior distribution is valid. In contrast, the QRS complex in the PACE and WPACE estimates is highly distorted.

As with the preceding example, we numerically assess estimation performance of all methods. In this simulation, for each of 100 iterations, we select a PQRST complex from the training data at random, and artificially subsample it to 10 observations chosen independently and uniformly along the domain; the remaining PQRST complexes are treated as training data to generate the empirical prior for the amplitude component. We consider the same three posterior estimates as in the previous section and compare them to the PACE and WPACE estimates. Boxplots of the $\mathbb{L}^2$ distance from the true function to the five different estimates are shown in Figure 7(a). It appears that all methods show comparable performance.

Consensus amongst the different estimators occurs mainly due to the fact that the $\mathbb{L}^2$ distance criterion greatly penalizes misalignment between the true and estimated functions. Consider the example visualized in Figures 7(c) and (d). The three estimates in (c) are based on the proposed model, while the two estimates in (d) correspond to PACE and WPACE. The estimated functions in panel (c) are clearly better at capturing the shape of the PQRST complex. Unfortunately, the estimated phase results in a slight misalignment of the very pronounced R peak, which carries a significant penalty based on the $\mathbb{L}^2$ distance. This misalignment in the estimate is due to a lack of observations along this important feature of the PQRST complex. On the other hand, the PACE and WPACE estimates in panel (d) are not at all successful at capturing the true shape of the PQRST complex. To confirm this behavior, we additionally display the $\mathbb{L}^2$ distances between the amplitude components of the estimated PQRST complexes and the true amplitude component. This is done by first optimally aligning the estimates to the true PQRST complex using the eFR metric. The corresponding boxplots are displayed in Figure 7(b). It is clear that, with respect to this measure, the proposed model recovers amplitude features much better than PACE or WPACE. In fact, the plug-in estimate performs better than the PACE and WPACE estimates in 90% and 96% of the 100 iterations, respectively. Similarly, the MAP (pointwise) estimates perform better than the PACE and WPACE estimates in 84% (84%) and 90% (93%) of the 100 iterations, respectively.

### 4.2. Bayesian Model With Shape-Restricted Amplitude Prior

Next, we focus on examples where the proposed shape-restricted amplitude prior is most appropriate to estimate functions under considerable noise and sparsity.

#### 4.2.1. Simulated Example 1: Functions With Low Signal-to-Noise Ratio

We first consider simulated functional data that not only contains phase and amplitude variability, but also low signal-to-noise ratio. The data is shown in Figure 8(a). A shape-restricted amplitude prior is appropriate in this setting since we know that the underlying functions should have two peaks and a single valley. We use the settings $\theta_\gamma = 10$, $m_\gamma = 3$, $B = 10$, $H = 3$, $\alpha = (0.25, 0.5, 0.75)$ and $M = -1$ in the model. Marginal posterior means of the amplitude and phase components of the unknown functions are shown in the first and second rows of panel (c), respectively. In comparison, the amplitude and phase components estimated by WPACE are given in (d). The WPACE result appears unsatisfactory in two ways: (i) there are obvious artifacts in the estimated amplitude functions, and (ii) the phase component is underestimated, resulting in a fair amount of phase variability that remains in the amplitude estimates. On the other hand, the proposed model is able to appropriately account for the pointwise noise. It results in estimated amplitude functions that have correctly aligned peaks and valleys. Furthermore, there is a common degree of smoothness provided by the shape-restricted amplitude prior. In fact, the proposed model is able to estimate the true error variance (red line) in the likelihood fairly well (panel (b)).

#### 4.2.2. Real Data Example 1: Berkeley Growth Velocity Functions

To further illustrate the structure imposed by modeling amplitude with the shape-restricted amplitude prior, we applied our

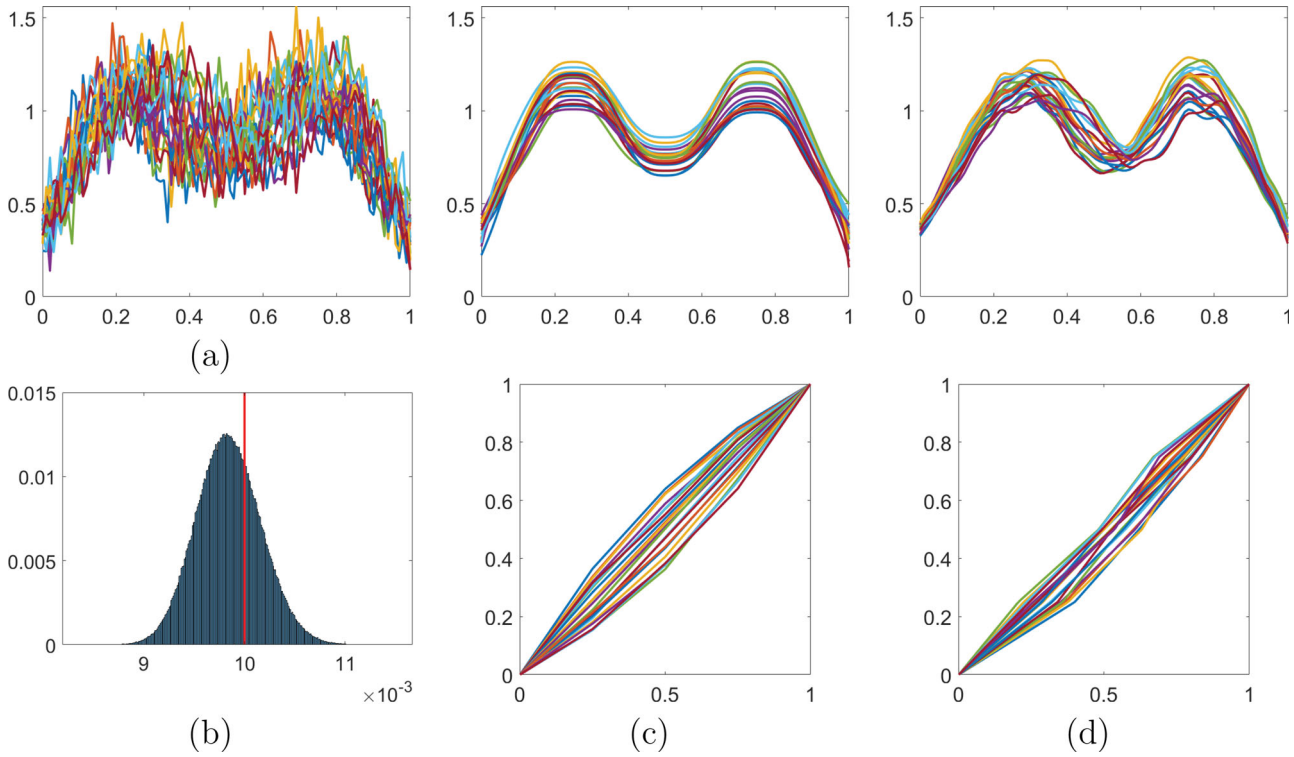


**Figure 8.** (a) Simulated observations with high noise level. (b) Normalized histogram of posterior draws of the error variance $\sigma^2$, with the value used to generate the data in red. (c) Posterior means of the amplitude (top) and phase (bottom) components estimated using the proposed Bayesian model with the shape-restricted amplitude prior. (d) Estimated amplitude (top) and phase (bottom) using WPACE.

methods to the well-known Berkeley growth dataset (Ramsay and Silverman 2005), in which the heights of children were tracked over the course of their lives from one to 18 years of age. We only use a subset of the data corresponding to 39 boys from the original study, shown in the introduction in Figure 1(c). In many cases, it is more natural to study the rate or velocity of growth, that is, the derivative of height with respect to time, than growth itself. This allows for better understanding of growth patterns of the subjects wherein peaks in the velocity functions correspond to growth spurts, with the last, largest peak being the pubertal growth spurt. Consequently, we consider two different settings for our model; the first one allows for a single pubertal growth spurt in the amplitude functions, $H = 2$, $\alpha = (0.57, 0.72)$, while the second one allows for an additional earlier growth spurt, $H = 4$, $\alpha = (0.23, 0.37, 0.57, 0.72)$. We also set $\theta_\gamma = 10$ and $B = 20$ in both models. In both cases, $\alpha$ was chosen based on visual inspection of the cross-sectional average of the data. Alternatively, one can specify $\alpha$ according to a uniform spacing on the domain $[0, 1]$.

The estimated amplitude and phase functions under the two amplitude prior settings are visualized in Figures 9(a) and (b). A comparison to the result generated by WPACE is given in panel (c). As expected, the registration and level of smoothness for the pubertal growth spurt is similar across panels (a) and (b). The main difference between the two sets of amplitude estimates is in the additional growth spurt estimated for some of the subjects in (b). Contrasting these results to the WPACE registration and estimation results shown in (c), the prepubertal growth spurt is often smoothed-out, with the model failing to properly align many of the subjects' pubertal growth spurts. This is especially surprising since the phase functions estimated by WPACE are much less regular than those estimated using the proposed model. Figure 10 shows detailed inferential results for three individuals with and without the prepubertal growth spurt. In panel (a), it appears that the subject has a fairly significant prepubertal growth spurt. When the amplitude prior in our model is restricted to allow for a single peak, the pattern of observations around this growth spurt is treated as noise and consequently it is missed in the resulting estimated function. Additionally, posterior uncertainty in this region is relatively large. In contrast, when the prior is relaxed to allow for an additional growth spurt, we are able to nicely estimate both the prepubertal and the pubertal growth spurts. In panel (b), there appears to be a single large growth spurt. Both estimates provided by our model appear to fit the data well. Finally, in (c), it is unclear whether there is a small prepubertal growth spurt. Again, both estimates provided by our model are reasonable. This example suggests that fixing the amplitude hyperparameter $H$ a-priori can be limiting in practice and motivates future work to jointly estimate $H$ for different subjects. The WPACE estimates are shown in each panel in red; the WPACE estimate in (a) appears to severely oversmooth and underestimate the pubertal growth spurt.

#### 4.2.3. *Real Data Example 2: Noisy FA Functions From DT-MRI*

We consider estimation and registration of complete, but noisy, fractional anisotropy functions obtained from diffusion tensor-magnetic resonance imaging (DT-MRI). DT-MRI is a neuroimaging modality that traces the diffusion of water molecules in the brain. A scan of a subject's brain provides a $3 \times 3$ matrix at each voxel in the image that describes the constraints of local motion of water molecules. This information is essential

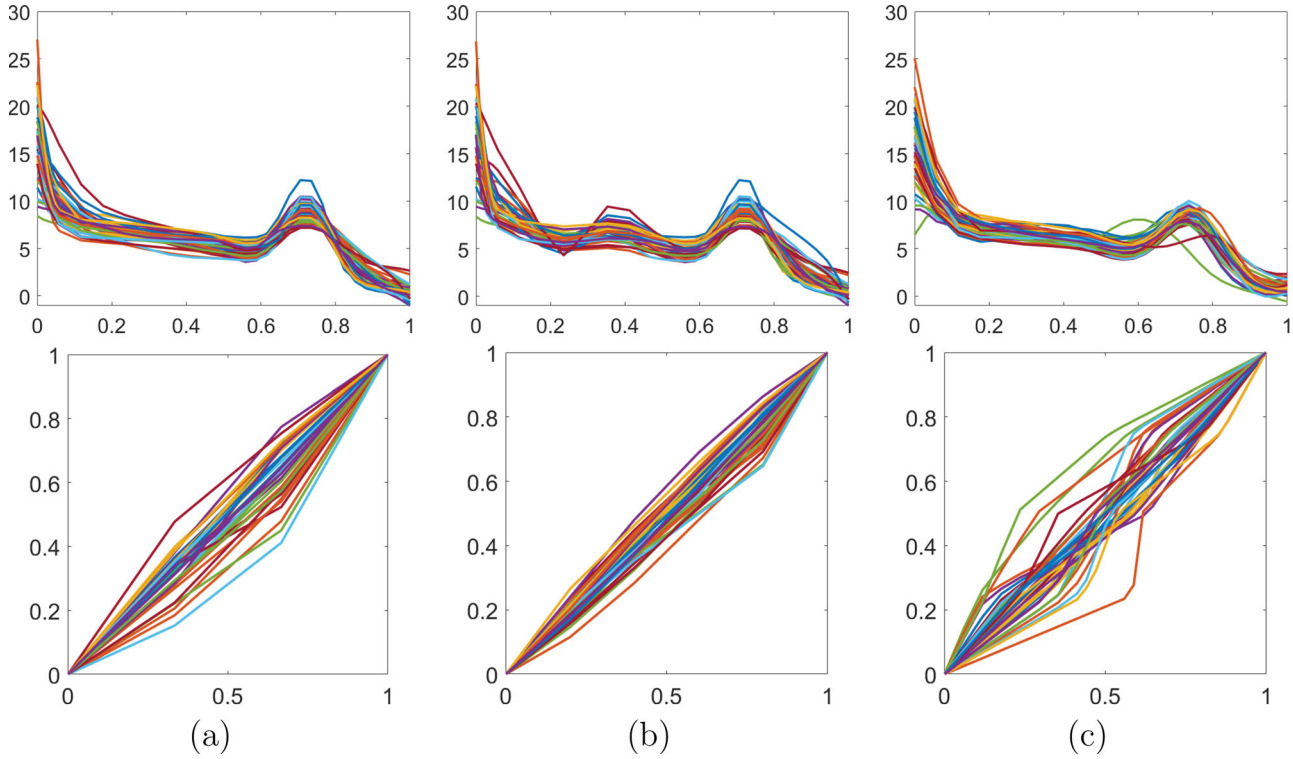


**Figure 9.** Posterior means of the amplitude (top) and phase (bottom) components based on the proposed model with the shape-restricted amplitude prior with (a) $H = 2$ and (b) $H = 4$. (c) Estimated amplitude (top) and phase (bottom) functions using WPACE.

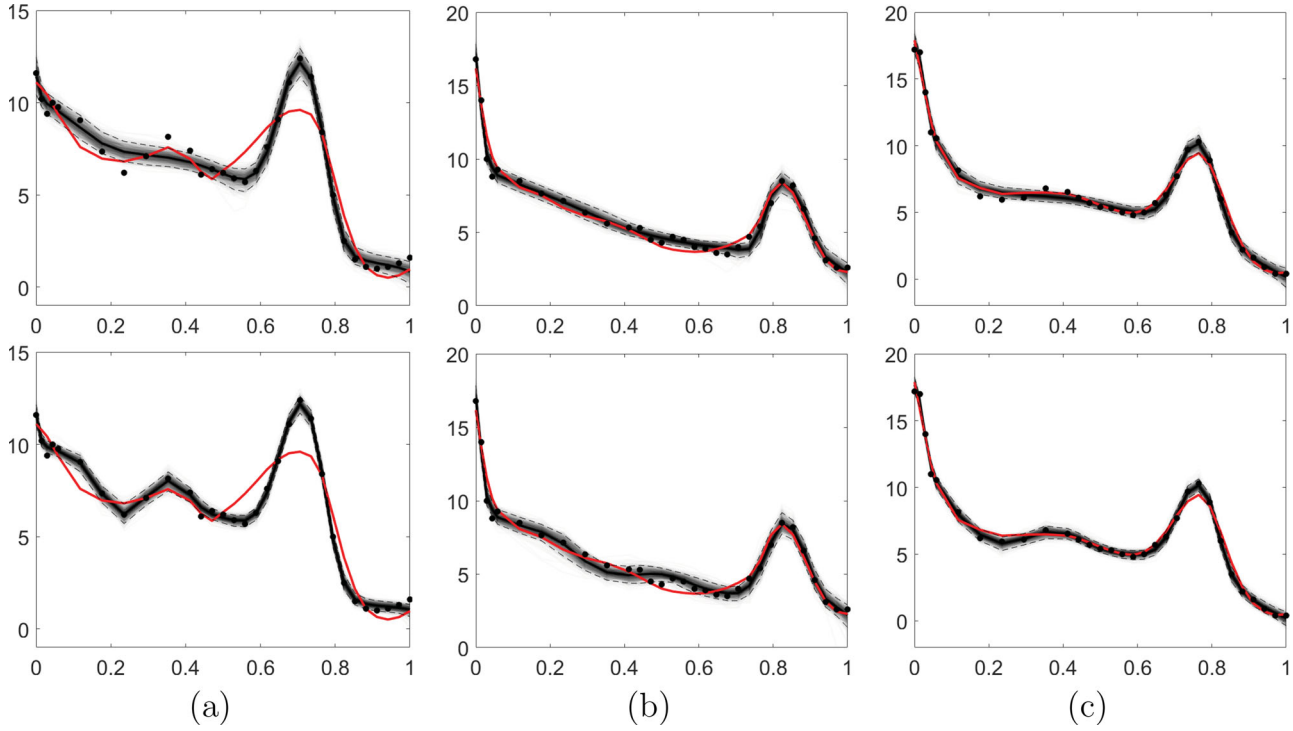


**Figure 10.** Posterior draws (transparent), mean (solid black), and credible interval (dashed) when $H = 2$ and (top) $H = 4$ (bottom), with a comparison to the WPACE fitted function (red) for three different subjects.

to understanding white matter in the brain, which constitutes areas made up of axons or tracts. Tracts connect neurons and allow for the transmittance of electric signals from one area of the brain to another, affecting overall brain function. Due to anisotropic diffusion of water along tracts, they can be extracted from the information contained in a DT-MRI, along with other quantities of interest that describe the properties of a tract by summarizing its degree of anisotropy.

Functional anisotropy (FA) measurements along tracts provide a voxelwise summary of the eigenvalues, denoted

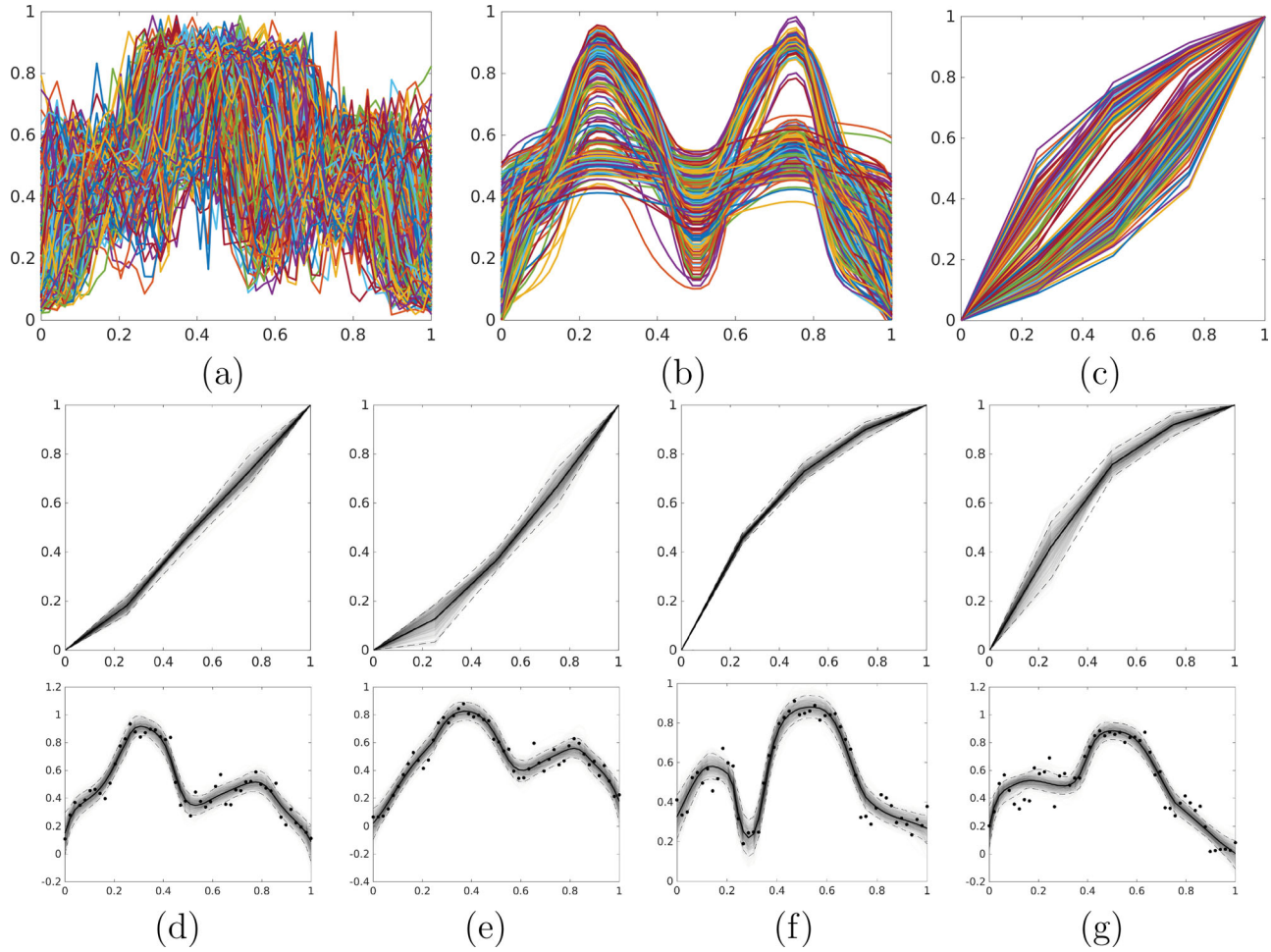


**Figure 11.** (a) FA functions estimated based on fiber tracts corresponding to the callosum forceps major, which connects the left and right occipital lobes. (b) Posterior mean amplitude functions. (c) Posterior mean phase functions. (d–g) Posterior draws (transparent), pointwise mean (solid black) and 95% credible interval (dashed) for the phase component (top) and the composition of amplitude and phase (bottom). FA functions in (d) and (e) have a larger peak followed by a smaller peak, while FA functions in (f) and (g) have a smaller peak followed by a larger peak.

by $\nu_1$, $\nu_2$, $\nu_3$, of the diffusion matrices. At each voxel in the image, FA is given by the scalar quantity $FA = \sqrt{\frac{3}{2}}\sqrt{\frac{(\nu_1-\bar{\nu})^2+(\nu_2-\bar{\nu})^2+(\nu_3-\bar{\nu})^2}{\nu_1^2+\nu_2^2+\nu_3^2}}$, where $\bar{\nu} = \frac{\nu_1+\nu_2+\nu_3}{3}$. A large FA value indicates a large degree of anisotropy. For practitioners, this summary of a DT-MRI provides a measurement of the quality of neuronal connections between particular regions of interest.

A major advantage of DT-MRI as a diagnostic tool is that it is noninvasive. However, FA measurements based on DT-MRI are subject to spurious values. The data that we consider here are a subset of the combination antiretroviral therapy (cART) dataset of HIV patients, which was collected in a local research study at the University of Rochester. The focus of this study was on assessing the neurodegeneration of HIV-infected subjects during cART treatment. The diffusion MR sequence that was used to acquire this data was 60 direction diffusion weighted imaging with field map correction. The MR scanner used was a 3T Siemens Tim Trio. Our data analysis involves FA functions estimated from a single cluster of fiber tracts corresponding to the callosum forceps major, which connects the left and right occipital lobes. This bundle crosses the splenium of the corpus callosum. For more details on this dataset and data processing involved, see Zhu et al. (2013). The initial clustering was performed using the methods described in Zhang et al. (2018).

In this case, we use the shape-restricted prior amplitude model for estimation. Using visual inspection, we selected $H = 3$, $M = -1$ and $\alpha = (0.25, 0.5, 0.75)$. We also set $\theta_\gamma = 10$. The observed data is shown in Figure 11(a). It is clear that there is considerable noise and misregistration in this data. The estimated posterior mean amplitude functions (i.e., registered FA functions) are shown in Figure 11(b). We discover an interesting pattern; it appears that not all FA functions in this tract cluster have the same pattern of extrema: evidently, from panel (b), we note that some amplitude functions have a large left peak and a small right peak, while the right peak is larger than the left peak in other functions. The estimated posterior mean phase functions are given in panel (c), where again we observe two clusters.

Four detailed estimation results for amplitude and phase are shown in panels (d)–(g). The top portion of each panel shows the posterior samples of phase functions only (again, the posterior mean is given by a solid black phase function), while the bottom portion shows the composition of the posterior samples of amplitude and phase functions (the solid black function corresponds to the posterior mean). Based on these results, it

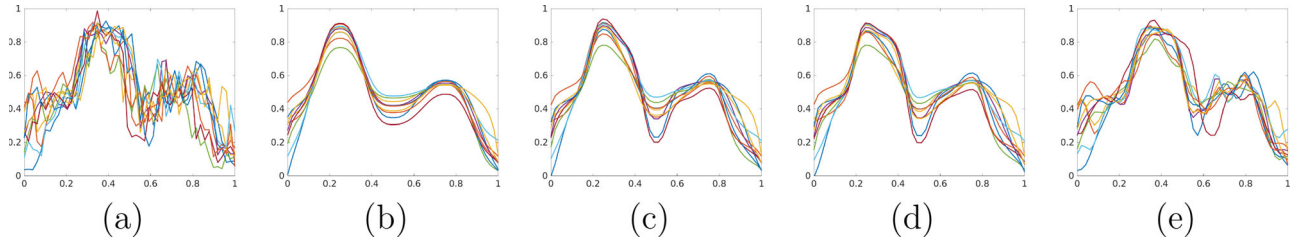


**Figure 12.** (a) FA functions estimated based on a subset of fiber tracts corresponding to the callosum forceps major, which connects the left and right occipital lobes. Posterior mean amplitude estimates based on the shape-restricted amplitude prior model with $H = 3, \alpha = (0.25, 0.5, 0.75)$, and (b) $B = 10$, (c) $B = 15$ and (d) $B = 20$. (e) WPACE registration and estimation result.

appears that the two amplitude and phase clusters correspond to each other, that is, the FA functions with a higher first peak correspond to phase functions that fall below identity phase while the FA functions with a higher second peak correspond to phase functions that fall above identity phase. Thus, we are able to simultaneously register the FA functions within this fiber tract cluster and identify smaller subclusters. These subclusters correspond to different geometry of the fiber tracts, encoded in the different patterns of extrema in the FA functions. Overall, we have found that roughly 76% of the FA functions in this fiber tract have a larger left peak, while the remaining functions have a larger second peak.

Next, we analyze an even smaller subset of the data, where we identified only the FA functions with a larger first peak. This dataset is displayed in Figure 12(a). In panels (b)–(d), we display the posterior mean estimates of the amplitude component when $B = 10$, $B = 15$, and $B = 20$, respectively. We again use $H = 3$, $\alpha = (0.25, 0.5, 0.75)$, and $\theta_\gamma = 10$. It appears that when $B = 10$, the estimates are very smooth. When $B$ is increased to 15 or 20, the posterior mean amplitude functions look very similar. We compare our result to WPACE in panel (f). It is clear that the functions are not registered well by the WPACE approach.

#### 4.2.4. Real Data Example 3: Bone Mineral Density

We consider indirect X-ray measurements of BMD, associated with skeletal health and diseases such as osteoporosis (Bachrach et al. 1999). While osteoporosis affects individuals later in life, bone development during adolescence through early adulthood can be used to assess an individual's risk for the disease. We focus on a subset of the entire dataset corresponding to females aged 9–25 years that had their BMD measured during two, three or four doctor appointments over the course of several years. Of primary interest in the study was the mean difference in BMD for four different ethnicities. Figure 13(a) shows the data where the different ethnic groups are highlighted using different colors. While this subset of the data has been used to classify individuals in previous work (James and Hastie 2001; Delaigle and Hall 2013), our focus is on estimation of a mean BMD trajectory for each group that accommodates phase variability. We restrict our model by forcing all individuals within the same ethnic group to have a common translation and a common strictly increasing amplitude function (prior hyperparameters are set to $B = 5$, $H = 0$, and $M = 1$). We do allow for individual phase variability with $\theta_\gamma = 100$ and $m_\gamma = 1$, regularizing the phase functions toward simple, small deviations from the identity element. The composition of the amplitude and phase

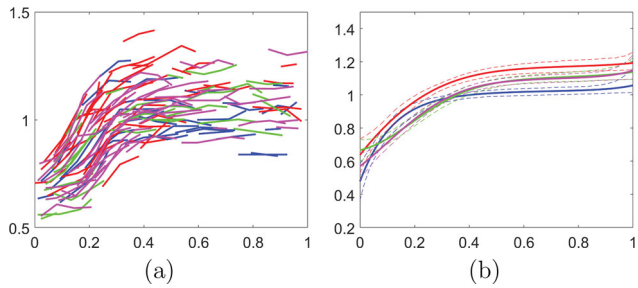


**Figure 13.** (a) BMD measurements colored by ethnicity: Asian (blue), Black (red), Hispanic (green) and Caucasian (magenta). (b) Posterior estimates of mean BMD functions (solid lines), with 95% credible intervals (dashed lines), for the four ethnic groups.

samples then corresponds to individual BMD trajectories, and our interest is in their mean estimate and the associated posterior uncertainty. Our assumption that the mean BMD function is strictly increasing stems from the age range of the individuals in the study, and enforces the principle that, on average, BMD increases during this period.

The resulting groupwise mean BMD trajectory estimates, with 95% credible intervals, are shown in Figure 13(b). The results obtained via the proposed model visually corroborate the major finding of the original study that the Black ethnic group (red) has a significantly higher BMD than the other groups. Structurally, the estimated mean function for the Asian group, shown in blue, is also quite different from the others: there is a lot of fast BMD growth early followed by minimal growth later on in life. The estimated mean BMD functions for the Caucasian and Hispanic groups appear extremely similar. The original study concluded that the Asian, Caucasian and Hispanic groups have BMD patterns that are difficult to distinguish from one another. Please see Section 3 of the supplementary materials for additional discussion of this example.

## 5. Discussion and Future Work

We have presented a Bayesian framework for simultaneous registration and estimation of functional observations that can handle challenging observational regimes such as sparsity, fragmentation and low signal-to-noise ratio. The framework explicitly accounts for phase and amplitude variability, and imposes amplitude restrictions in situations where different types of prior information are available. Indeed, we show that the key to inference under general observational regimes is the ability to inform the prior on the underlying structure of amplitude. We

demonstrate the performance of the proposed model on several different simulated and real data examples to show the diverse range of scenarios that can be analyzed. We also compare the proposed method to a state-of-the-art competitor.

We have identified several directions for future work. First, we acknowledge that the prior information required for the use of this framework may not always be available. Consequently, for both the empirical and shape-restricted amplitude prior scenarios, we suggest additional modeling strategies that will broaden the proposed framework. In the first case, we will incorporate an additional hierarchical layer corresponding to subspace estimation (via fPCA) for the amplitude subspace. This will allow for direct propagation of uncertainty from the training stage to the estimation stage. In the second case, we will treat the number of local extrema allowed in the amplitude estimates $H$ as well as their ordering pattern $M$, as unknown quantities to be estimated. This challenging extension of our model will allow us to apply it in scenarios when knowledge about the patterns of extrema is unavailable a-priori and enough data is available to inform all these model components.

Second, we will extend the proposed framework to common modeling tasks in FDA including clustering, classification and functional regression. Importantly, the estimation, registration and further modeling will be achieved via a single unified framework.

Finally, we will extend the proposed model to other functional data scenarios including shapes of higher-dimensional curves, images and shapes of surfaces. We note that the proposed model will have to undergo major reformulation for the case of images and shapes of surfaces due to higher dimensionality of the domain for registration (phase) of these functional data objects (Jermyn, Kurtek, and Laga 2017).

## Supplementary Materials

The Supplemental Materials include additional simulated and real data examples, implementation details and diagnostics for the MCMC algorithm used to sample from the posterior distribution, and detailed sensitivity analyses for various modeling choices.

## Acknowledgments

We thank Dr. Zhengwu Zhang and Mr. Yuchuan Zhuang from the University of Rochester for sharing the cART dataset. We also thank the associate editor and two reviewers for their thoughtful comments. We acknowledge the Isaac Newton Institute for Mathematical Sciences for hosting the Variational Methods and Effective Algorithms for Imaging and Vision in Fall 2017 where many of the research ideas presented in this article originated.

## Funding

This research was partially funded by NSF DMS-1613054, NIH R37-CA214955 (to SK and KB), NSF DMS-2015374 (to KB), and NSF CCF-1740761, NSF DMS-2015226, and NSF CCF-1839252 (to SK).

## References

Bachrach, L. K., Hastie, T., Wang, M.-C., Narasimhan, B., and Marcus, R. (1999), "Bone Mineral Acquisition in Healthy Asian, Hispanic, Black, and Caucasian Youth: A Longitudinal Study," *Journal of Clinical Endocrinology & Metabolism*, 84, 4702–4712. [15]

Bharath, K., and Kurtek, S. (2020), "Distribution on Warp Maps for Alignment of Open and Closed Curves," *Journal of the American Statistical Association*, 115, 1378–1392. [3,6]

Claeskens, G., Silverman, B. W., and Slaets, L. (2010), "A Multiresolution Approach to Time Warping Achieved by a Bayesian Prior-Posterior Transfer Fitting Strategy," *Journal of the Royal Statistical Society*, Series B, 72, 673–694. [3]

Dasgupta, S., Pati, D., and Srivastava, A. (2019), "Bayesian Shape-Constrained Density Estimation," *Quarterly of Applied Mathematics*, 77, 399–422. [2]

——— (2020), "A Two-Step Geometric Framework for Density Modeling," *Statistica Sinica*, 30, 2155–2177. [2]

Delaigle, A., and Hall, P. (2013), "Classification Using Censored Functional Data," *Journal of the American Statistical Association*, 108, 1269–1283. [15]

——— (2016), "Approximating Fragmented Functional Data by Segments of Markov Chains," *Biometrika*, 103, 779–799. [3]

Descary, M.-H., and Panaretos, V. M. (2019), "Functional Data Analysis by Matrix Completion," *The Annals of Statistics*, 47, 1–38. [2]

Ferraty, F., and Vieu, P. (2006), *Nonparametric Functional Data Analysis: Theory and Practice*, New York: Springer-Verlag. [1]

Fu, E., and Heckman, N. (2019), "Model-Based Curve Registration via Stochastic Approximation EM Algorithm," *Computational Statistics & Data Analysis*, 131, 159–175. [3,6]

James, G. M., and Hastie, T. J. (2001), "Functional Linear Discriminant Analysis for Irregularly Sampled Curves," *Journal of the Royal Statistical Society*, Series B, 63, 533–550. [15]

Jermyn, I. H., Kurtek, S., Laga, H., and Srivastava, A. (2017), *Elastic Shape Analysis of Three-Dimensional Objects*, San Rafael: Morgan & Claypool Publishers. [16]

Kneip, A., and Gasser, T. (1992), "Statistical Tools to Analyze Data Representing a Sample of Curves," *The Annals of Statistics*, 20, 1266–1305. [3,7,8]

Kneip, A., and Ramsay, J. O. (2008), "Combining Registration and Fitting for Functional Models," *Journal of the American Statistical Association*, 103, 1155–1165. [3,5]

Kraus, D. (2015), "Components and Completion of Partially Observed Functional Data," *Journal of the Royal Statistical Society*, Series B, 777, 777–801. [3]

Kurtek, S., Wu, W., Christensen, G. E., and Srivastava, A. (2013), "Segmentation, Alignment and Statistical Analysis of Biosignals With Application to Disease Classification," *Journal of Applied Statistics*, 40, 1270–1288. [10]

Lawton, W. H., Sylvestre, E. A., and Maggio, M. S. (1972), "Self Modeling Nonlinear Regression," *Technometrics*, 14, 513–532. [3]

Liebl, D. (2019), "Partially Observed Functional Data: The Case of Systematically Missing Parts," *Computational Statistics and Data Analysis*, 131, 104–115. [3]

Lu, Y., Herbei, R., and Kurtek, S. (2017), "Bayesian Registration of Functions With a Gaussian Process Prior," *Journal of Computational and Graphical Statistics*, 26, 894–904. [3,6]

Marron, J. S., Ramsay, J. O., Sangalli, L. M., and Srivastava, A. (2015), "Functional Data Analysis of Amplitude and Phase Variation," *Statistical Science*, 30, 468–484. [1,3,4,10]

Rakêt, L. L., Sommer, S., and Markussen, B. (2014), "A Nonlinear Mixed-Effects Model for Simultaneous Smoothing and Registration of Functional Data," *Pattern Recognition Letters*, 38, 1–7. [3]

Ramsay, J. O., and Li, X. (1998), "Curve Registration," *Journal of the Royal Statistical Society*, Series B, 60, 351–363. [3]

Ramsay, J., and Silverman, B. (2005), *Functional Data Analysis*, New York: Springer. [1,2,3,4,12]

Rice, J., and Wu, C. (2001), "Nonparametric Mixed Effects Models for Unequally Sampled Noisy Curves," *Biometrics*, 57, 253–259. [2,3]

Shi, M., Weiss, R. E., and Taylor, J. M. G. (1996), "An Analysis of Paediatric CD4 Counts for Acquired Immune Deficiency Syndrome Using Flexible Random Curves," *Journal of the Royal Statistical Society*, Series C, 45, 151–163. [2]

Srivastava, A., and Klassen, E. (2016), *Functional and Shape Data Analysis*, New York: Springer. [1,2,3,4]

Srivastava, A., Wu, W., Kurtek, S., Klassen, E., and Marron, J. S. (2011), "Registration of Functional Data Using Fisher-Rao Metric," arXiv no. 1103.3817. [3,4,5]

Strait, J., Chkrebtii, O., and Kurtek, S. (2019), "Automatic Detection and Uncertainty Quantification of Landmarks on Elastic Curves," *Journal of the American Statistical Association*, 114, 1002–1017. [3,8]

Tang, R., and Müller, H.-G. (2008), "Pairwise Curve Synchronization for Functional Data," *Biometrika*, 95, 875–889. [3,8]

Telesca, D., and Inoue, L. (2008), "Bayesian Hierarchical Curve Registration," *Journal of the American Statistical Association*, 103, 328–339. [3,6]

Tucker, J. D., Wu, W., and Srivastava, A. (2013), "Generative Models for Functional Data Using Phase and Amplitude Separation," *Computational Statistics & Data Analysis*, 61, 50–66. [4,5]

Wahba, G. (1990), *Spline Models for Observational Data*, Philadelphia, PA: SIAM. [2]

Wang, J.-L., Chiou, J.-M., and Müller, H.-G. (2016), "Functional Data Analysis," *Annual Review of Statistics and Its Application*, 3, 257–295. [3]

Wang, K., and Gasser, T. (1997), "Alignment of Curves by Dynamic Time Warping," *The Annals of Statistics*, 25, 1251–1276. [3]

Wheeler, M., Dunson, D., and Herring, A. (2017), "Bayesian Local Extrema Splines," *Biometrika*, 104, 939–952. [2,3,7]

Yang, J., Zhu, H., Choi, T., and Cox, D. D. (2016), "Smoothing and Mean-Covariance Estimation of Functional Data With a Bayesian Hierarchical Model," *Bayesian Analysis*, 11, 649–670. [2]

Yao, F., Müller, H.-G., and Wang, J.-L. (2005), "Functional Data Analysis for Sparse Longitudinal Data," *Journal of the American Statistical Association*, 100, 577–590. [3,8]

Zhang, F., Wu, Y., Norton, I., Rigolo, L., Rathi, Y., Makris, N., and O'Donnell, L. J. (2018), "An Anatomically Curated Fiber Clustering White Matter Atlas for Consistent White Matter Tract Parcellation Across the Lifespan," *NeuroImage*, 179, 429–447. [14]

Zhang, Z., Descoteaux, M., and Dunson, D. B. (2019), "Nonparametric Bayes Models of Fiber Curves Connecting Brain Regions," *Journal of the American Statistical Association*, 114, 1505–1517. [7]

Zhou, R., Serban, M., Gebraeel, N., and Müller, H.-G. (2014), "A Functional Time Warping Approach to Modeling and Monitoring Truncated Degradation Signals," *Technometrics*, 56, 67–77. [3,6]

Zhu, T., Zhong, J., Hu, R., Tivarus, M., Ekholm, S., Harezlak, J., Ombao, H., Navia, B., Cohen, R., and Schifitto, G. (2013), "Patterns of White Matter Injury in HIV Infection After Partial Immune Reconstitution: A DTI Tract-Based Spatial Statistics Study," *Journal of NeuroVirology*, 19, 10–23. [14]