

Estimation of Sparsely Observed Signals with an Empirical Bayesian Model

James Matuk, Oksana Chkrebtii, Sebastian Kurtek

Department of Statistics, The Ohio State University

matuk.3@osu.edu, oksana@stat.osu.edu, kurtek.1@stat.osu.edu

Abstract—In many applied settings, the statistical goal is to estimate underlying functions or signals from discrete observations contaminated with noise. The task is often prone to issues during data collection resulting in incomplete measurements (e.g., signals observed on a sparse set of time points, or with entire segments of the signal missing). To address this problem, we develop a Bayesian model that allows joint estimation of the amplitude (y -axis) and phase (x -axis) components of unknown signals. The major advantages of our method are the ability to (1) separately model these two sources of variability, (2) use smooth, complete data to incorporate prior knowledge about the underlying structure in a principled way, and (3) compute and quantify uncertainty of estimated signals. We build our model using a convenient square-root velocity function representation of signals, which allows for metric-based separation and statistical analysis of amplitude and phase. We validate the proposed framework using two simulation studies, and a real data application to estimation of fractional anisotropy profiles based on diffusion tensor imaging measurements on patients with multiple sclerosis.

Index Terms—function estimation, Bayesian model, amplitude, phase, fractional anisotropy

I. INTRODUCTION

Signals or functions arise as primary data objects in a wide variety of scientific fields including biometrics, medical imaging, biology, environmental sciences, and many more [4]. In statistical signal processing, the data generating mechanism that produces the signals is a stochastic processes. When the process is univariate, one can apply statistical methods from the field of functional data analysis. The goals of functional data analysis are the same as in standard settings that involve univariate or multivariate data, and include summarization, modeling of variability, various types of inference, classification, clustering, etc. However, in poor data conditions, the observed signals may be corrupted by noise or simply incomplete, i.e., observed at a sparse set of time points, or with entire segments of the signal missing. In such cases, the primary statistical goal is to estimate the underlying signal and quantify uncertainty in this estimate.

A. Motivation

Function estimation and smoothing address a data quality problem in signal processing wherein only noisy, discrete observations of the signal are available. The problem is often exacerbated by issues during data collection resulting

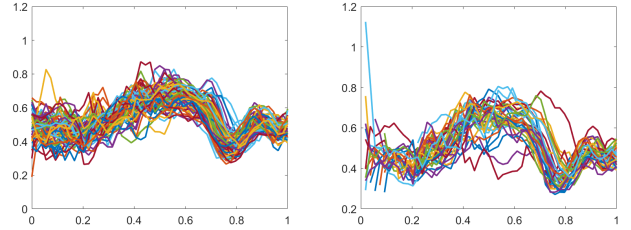


Fig. 1: Fractional anisotropy signals for subjects with complete observations (left) and partial observations (right).

in signals observed only on a sparse set of time points or with entire segments of the signal missing. In this paper, we consider a specific signal estimation problem with the following characteristics. Let f_i represent the signal underlying the i th observation $y_i(\mathbf{t}_i) = f_i(\mathbf{t}_i) + \epsilon_i(\mathbf{t}_i)$, $i = 1, \dots, n$, contaminated with pointwise, iid error $\epsilon_i(\mathbf{t}_i)$. The first n_c functions are observed on dense grids $\mathbf{t}_i = (t_{1i}, \dots, t_{K_i})'$, $t_{1i} < \dots < t_{K_i}$ that represent the input domain well. The remaining $n - n_c$ functions are observed on grids that are sparse with random or systematic missingness such that values for entire regions of the domain are not observed. The overall goal is to use the fully observed functions to build an empirical statistical model that can then be employed to estimate the noisy, sparsely observed functions. We take a Bayesian approach, which allows natural assessment of uncertainty in the estimated functions via the posterior distribution.

An important motivating data example that follows the aforementioned setup occurs in the case of fractional anisotropy (FA) functions along right corticospinal tract locations measured using diffusion tensor imaging (DTI) for patients with multiple sclerosis (MS). This application, and the corresponding dataset, are described in detail in [2]. FA measurements track disease progression in each patient and can be used to predict disability outcomes using a functional regression model with a scalar response. A major advantage of DTI is that it is non-invasive. However, FA measurements based on DTI are often subject to spurious or missing values. In this dataset, of the subset of 100 patients that were diagnosed with MS, 34 patients had missing values within the first five tract locations due to technical issues with the measuring device, while all 55 tract locations were observed for the remaining 66 patients. If one were to conduct a complete

case analysis for these data by discarding the incomplete observations, the resulting inference will result in greater uncertainty and may miss common population characteristics. Figure 1 displays the entire dataset used in our study. The left panel shows the 66 complete FA signals while the right panel displays the 34 incomplete FA signals.

B. Previous Methods

A common model in function estimation represents each signal f_i , which generates the i th observed function, via a basis expansion $f_i(t) = \sum_{b=1}^B c_{i,b} U_b(t)$, where $\{c_{i,b}\}_{b=1,\dots,B}$ are unknown coefficients and $\{U_b(t)\}_{b=1,\dots,B}$ is an application-dependant basis system [4]. A B-spline basis is a common choice for nonparametric function estimation, and has been recently adapted in [9] to enforce shape restrictions by specifying an upper bound on the number of local extrema. In general, estimation is carried out using the least squares or penalized least squares methods, which perform well when the signal-to-noise ratio of the data is relatively high. On the other hand, this approach is less reliable when the data is missing along large portions of the input domain; in those cases, a local basis is often used and can diminish population level features in the pursuit of local fit. Noting these shortcomings, others have used mixed-effects and hierarchical models first discussed in [5]. Through this set up, functions are modeled using population and subject-specific components such that all observations inform population features. In particular, this allows one to use completely observed functions to inform structural features in locations on the input domain where the incomplete functions are not observed. An alternative approach is to consider the plug-in estimator $\hat{f}_i(t) = \hat{\mu} + \sum_{b=1}^B \hat{c}_{i,b} \hat{U}_b(t)$, where $\hat{\mu}$ and $\hat{U}_b$ are the sample mean signal, and principal components, respectively, based on the sample $y_1, \dots, y_n$ and $\hat{c}_{i,b}$ are function-specific coefficients. These estimated quantities are obtained via a two-stage procedure consisting of functional principal component analysis (fPCA) and the Karhunen-Loeve expansion. This method is called principal analysis through conditional expectation (PACE) [10]. Since all complete signals can be used to estimate the mean and principal components, the features of the partially observed signals are informed by these summaries.

C. Contributions and Paper Organization

We develop a new, empirical model for signal estimation based on a natural decomposition of signals into their phase and amplitude variabilities. As demonstrated in [6], failing to account for these two components results in biased inference when considerable phase variability is present. We define a hierarchical Bayesian model over the unknown signals, and naturally introduce information about phase and amplitude of missing observations or segments through prior distributions estimated from complete data. The Bayesian paradigm then allows structured uncertainty quantification for the two components of variability through the posterior distribution.

The rest of this paper is organized as follows. In Section II, we discuss metric-based phase and amplitude separation and

modeling for completely observed signals. We also formalize a statistical model that incorporates phase and amplitude separation and can be used for prediction of partially observed signals, discussing how the model is fitted in practice. In Section III, we demonstrate the effectiveness of our model through two simulated examples, and return to the problem of estimating FA signals from DTI described earlier. Finally, in Section IV, we give a brief summary.

II. PROPOSED METHOD

We consider absolutely continuous signals $f_1, \dots, f_{n_c} : [0, 1] \rightarrow \mathbb{R}$ that belong to a function space $\mathcal{F}$. We model these signals as realizations of a random process on the input domain $[0, 1]$. Since they describe the same phenomenon, one expects them to share common features such as the number of local extrema. However, random observations exhibit variability in the heights of the extreme values and the locations at which they occur. These two notions of variability are referred to as amplitude and phase respectively and are formalized in [4], [6], where it is illustrated that failure to explicitly consider these two components in statistical models can lead to inferential bias. Thus, our model considers amplitude and phase separately. The amplitude component is informed using complete signals generated from the same random process, constraining the estimation problem to a space that better approximates $\mathcal{F}$. While the phase component could also be informed from analysis of the complete signals, this is often difficult in practice. Instead, we choose to incorporate very little prior information about this source of variability in our model to maintain flexibility in the estimation of phase. We begin by describing the necessary statistical tools to analyze the amplitude component of a sample of functions, followed by the elicitation of a statistical model that captures both forms of variability.

A. Phase and Amplitude Separation

The procedure of separating phase and amplitude variability in functional data is referred to as registration. Specifically, it consists of finding “registered” functions $\tilde{f}_1, \dots, \tilde{f}_{n_c} : [0, 1] \rightarrow \mathbb{R}$ and phase functions $\gamma_1, \dots, \gamma_{n_c} : [0, 1] \rightarrow [0, 1]$ that satisfy $\tilde{f}_i = f_i \circ \gamma_i$, $i = 1, \dots, n_c$; the registered functions $\tilde{f}_i$ exhibit no phase variability, i.e., their features are optimally aligned (we formalize this optimality later in this section). Figure 2 shows the separation of phase and amplitude for a simulated dataset. The left panel displays the original sample of signals, which all have two peaks and one valley and mainly differ in the heights of the peaks and valley on $[0, 1]$. The middle panel displays the amplitude component, or registered functions, whose major features are aligned. The right panel shows the phase component, or functions that achieve this registration.

An important choice in phase-amplitude separation and modeling is an appropriate metric. While computationally convenient, [6] show that the standard $\mathbb{L}^2$ metric is not appropriate for this problem due to its lack of invariance to phase variability. Thus, we instead use the Fisher-Rao (FR)

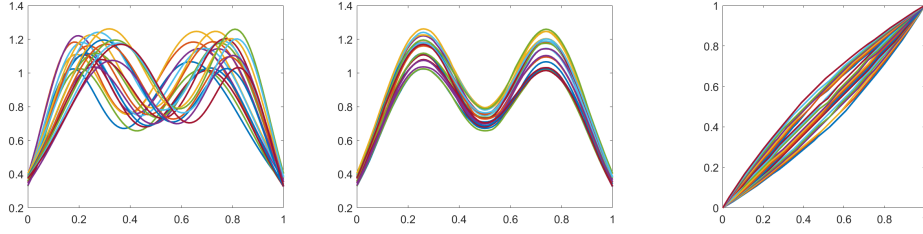


Fig. 2: Registration of signals using the FR metric. Left: observed signals. Middle: Amplitude component (registered functions). Right: Phase component.

Riemmanian metric, as proposed in [7]. An advantage of this metric is that the group of diffeomorphisms, a flexible class of phase functions denoted Γ , from $[0, 1]$ to itself acts on $\mathcal{F}$ by isometries. This ensures that the metric is invariant to phase variability. Computation under this metric is greatly simplified through the square-root velocity function (SRVF) representation of signals given by $Q(f); = q = \text{sign}(\dot{f})\sqrt{|\dot{f}|}$, where $\dot{f}$ is the time derivative of f . The SRVFs are square-integrable, making their representation space $\mathcal{Q}$ equivalent to $\mathbb{L}^2$. Using this transformation, the Fisher-Rao metric on $\mathcal{F}$ simplifies to the $\mathbb{L}^2$ metric on the space of SRVFs. The FR distance can then be easily computed using $d_{FR}(f_1, f_2) = \|q_1 - q_2\|$. Because of this simplification, and the fact that the SRVF representation is invertible up to a translation, $f(t) = Q^{-1}(q, f(0))(t) = f(0) + \int_0^t q(s)|q(s)|ds$, the space of SRVFs is advantageous for defining statistics of interest. Under the FR framework, [7] formally defines the amplitude of a function as follows: for any function $f \in \mathcal{F}$, its amplitude is the orbit $[q] = \{(q, \gamma) | \gamma \in \Gamma\}$, where $(q, \gamma) = (q \circ \gamma)\sqrt{|\dot{\gamma}|}$ is the group action of Γ on $\mathcal{Q}$. The collection of all distinct orbits constitutes the amplitude space $\mathcal{Q}/\Gamma$.

For $f_1, \dots, f_{n_c}$, and corresponding SRVFs $q_1, \dots, q_{n_c}$, the registration procedure involves finding a mean orbit

$$[\hat{\mu}_q] = \underset{[q] \in \mathcal{Q}/\Gamma}{\text{argmin}} \sum_{i=1}^n d([q], [q_i]) = \underset{[q]}{\text{argmin}} \sum_{i=1}^n \min_{\gamma \in \Gamma} \|q - (q_i, \gamma)\|^2.$$

Since one cannot model the entire mean orbit, a representative element, $\hat{\mu}_q \in [\hat{\mu}_q]$, is chosen such that the average phase in the sample is the identity function $\gamma_{id}(t) = t$. The solution to the above optimization problem can be found via a gradient-descent algorithm, and produces the phase components, $\hat{\gamma}_1, \dots, \hat{\gamma}_{n_c}$, and amplitude components, $(q_1, \hat{\gamma}_1), \dots, (q_{n_c}, \hat{\gamma}_{n_c})$ (or equivalently $f_1 \circ \hat{\gamma}_1, \dots, f_{n_c} \circ \hat{\gamma}_{n_c}$), for this sample of functions. This procedure also provides an estimate of the mean amplitude of the process, $\hat{\mu}_q$, that we use to impose structure in our model.

In [8], generative models for phase and amplitude within this framework are defined via fPCA. This principal component model will be adopted to define a prior distribution over the amplitude component. The sample covariance function over the registered SRVFs is given by $\widehat{K}(s, t) = \frac{1}{n_c - 1} \sum_{i=1}^{n_c} ((q_i, \hat{\gamma}_i)(s) - \hat{\mu}_q(s))((q_i, \hat{\gamma}_i)(t) - \hat{\mu}_q(t))$. The fPCA

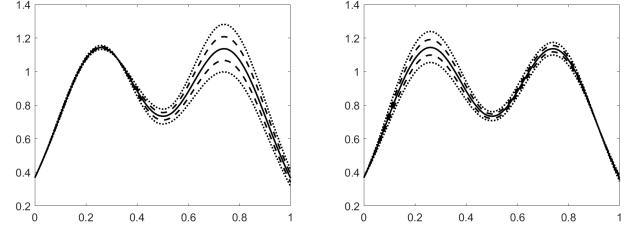


Fig. 3: The mean (solid), and ± 1 (dashed) and ± 2 (dotted) standard deviations in the directions $\hat{U}_1$ (left) and $\hat{U}_2$ (right).

basis is obtained through its eigendecomposition $\widehat{K}(s, t) = \sum_{b=1}^{\infty} \hat{\lambda}_b \hat{U}_b(s) \hat{U}_b(t)$, where each eigenfunction $\hat{U}_b$ is a functional principal component and each eigenvalue $\hat{\lambda}_b$ specifies the amount of variance in the direction $\hat{U}_b$. The set of all eigenfunctions forms an efficient orthonormal basis for the amplitude space. One can visualize the b th dominant mode of variability using the functions $r_{b,k} = \hat{\mu}_q + k\sqrt{\hat{\lambda}_b} \hat{U}_b$ for $k \in \{-2, -1, 0, 1, 2\}$. Figure 3 displays the first two dominant directions of variability for the simulated dataset in Figure 2. The first mode of variability, shown in the left panel, captures the different heights of the valley and the right peak. The second mode of variability, shown in the right panel, captures the different heights of the left peak. Together, these two modes explain most of the amplitude variability in the dataset, achieving significant dimension reduction, i.e., we can efficiently represent functions of this type using only two basis elements. This choice of basis can induce a parsimonious model on the amplitude component of functions generated from the same underlying random process. In fact, [8] demonstrate that this procedure induces a significantly more parsimonious model than fPCA of the unregistered functions.

By truncating the fPCA basis to B elements, we obtain a finite (low-dimensional) representation of the amplitude component: $(q_i, \hat{\gamma}_i) \approx \hat{\mu}_q + \sum_{b=1}^B \hat{c}_{i,b} \hat{U}_b$, where $\hat{c}_{i,b} = \int_0^1 ((q_i, \hat{\gamma}_i)(s) - \hat{\mu}_q(s)) \hat{U}_b(s) ds$. The prior in our Bayesian model will be elicited based on the SRVF mean $\hat{\mu}$, the principal component basis $\{\hat{U}_b\}_{b=1, \dots, B}$, and the principal coefficients $\hat{c}_1, \dots, \hat{c}_{n_c}$ for each SRVF.

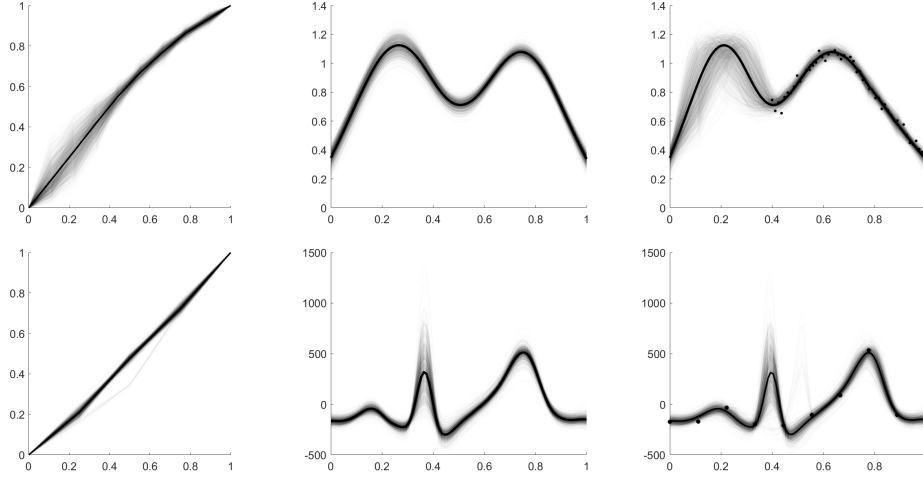


Fig. 4: Posterior draws (transparent gray) and pointwise mean (solid black) for the phase component (left), amplitude component (middle) and overall signal (right). Top: Left peak missing. Bottom: Sparse ECG PQRST signal.

B. Statistical Model

We re-express the additive noise model to incorporate terms for phase, amplitude and translation as,

$$\mathbf{y} = f(\mathbf{t}) + \epsilon(\mathbf{t}) = T + (Q^{-1}(q, 0) \circ \gamma(\mathbf{t})) + \epsilon(\mathbf{t}),$$

where $T \in \mathbb{R}$ is a translation term, γ is the phase component, and the amplitude $q = \hat{\mu}_q + \sum_{b=1}^B c_b \hat{U}_b$ is represented by a basis expansion and an additive population mean $\hat{\mu}_q$. We choose the basis system to be the empirical PCA basis, the amplitude of the underlying function is informed by all of the completely observed data. In addition to providing a parsimonious representation of amplitude, the fPCA basis ensures identifiability of all model parameters, since for different coefficient vectors $\mathbf{c}'_1, \mathbf{c}'_2$, the corresponding SRVFs $\hat{\mu}_q + \sum_b^B c'_{1,b} \hat{U}_b$ and $\hat{\mu}_q + \sum_b^B c'_{2,b} \hat{U}_b$ are guaranteed to belong to different orbits in $\mathcal{Q}/\Gamma$. We assume that ϵ is a conditionally independent zero-mean Gaussian process $\epsilon(\mathbf{t}) | \Sigma \sim \text{MVN}_K(0_K, \Sigma_K)$ with $\Sigma = \sigma^2 I_K$. Then, the likelihood of the data is,

$$\mathbf{y} | \mathbf{c}, \gamma, T, \sigma^2 \sim \text{MVN}_K(T + (Q^{-1}(q, 0) \circ \gamma(\mathbf{t})), \sigma^2 I_K).$$

Based on the mean and fPCA basis, estimated from completely observed signals, we form an empirical prior distribution for the basis coefficients for the amplitude component: $\mathbf{c} \sim \text{MVN}_B(0, \hat{\Sigma}_c)$, where $\hat{\Sigma}_c$ is the estimated covariance using the principal coefficients, $\hat{\mathbf{c}}_1, \dots, \hat{\mathbf{c}}_{n_c}$ (the mean of the principal coefficients is always zero). For translation, we choose a simple conjugate prior distribution: $T \sim N(\hat{\mu}_T, \hat{\sigma}_T^2)$, where $\hat{\mu}_T$ and $\hat{\sigma}_T^2$ denote the empirical mean and variance of the initial points $f_1(0), \dots, f_{n_c}(0)$. We also specify a conjugate, diffuse Inverse-Gamma($\alpha_\sigma, \beta_\sigma$) prior for the noise level σ^2 . Specifying a prior distribution for the phase component is more involved as Γ is a complicated geometric space [3]. Bharath and Kurtek [1] propose a computationally convenient prior model on the space of phase functions with desirable statistical properties. The prior is based on a Dirichlet model

on the increments of a random partition sampled from the uniform distribution on $[0, 1]$. This ensures that the distribution is centered around the identity phase function. The resulting distribution contains a hyperparameter θ , which controls spread. We use a small value of θ , which results in a diffuse prior. Since the posterior distribution is not tractable, we base posterior inference on Markov chain Monte Carlo (MCMC) samples obtained via a Metropolis-within-Gibbs algorithm. We omit implementation details for brevity.

III. EXPERIMENTAL DATA AND RESULTS

To illustrate the performance of the model, we present two simulation scenarios: (1) with a major portion of the signal is missing, and (2) with the signal observed on a very sparse set of time points. Our approach is then used to estimate FA signals from incomplete observations.

Simulation 1: We consider the case where a large portion, i.e., the entire left peak, of the observed signal is missing. The entire noisy dataset is shown in the left panel of Figure 5. A smoothed version of the same data, obtained using smoothing splines through generalized cross validation, is shown in the middle panel; this is the data that we use to estimate the amplitude prior for our model. The missing data scenario is shown in the right panel. The top panel of Figure 4 shows posterior samples of the phase (left) and amplitude (middle) components, as well as their composition (right). The posterior samples are visualized via transparent gray lines, while the pointwise posterior mean is shown in black, and the observed data is shown as black points. Due to the prior amplitude structure based on the fPCA model, all posterior draws have two peaks and one valley. As expected, the uncertainty is much greater in locations where no data is observed. The model appears to fit the observed data points well.

Simulation 2: For the next simulation, we sparsify an ECG PQRST complex as shown in the right panel of Figure 6; we will estimate the full signal from only ten evenly spaced points.

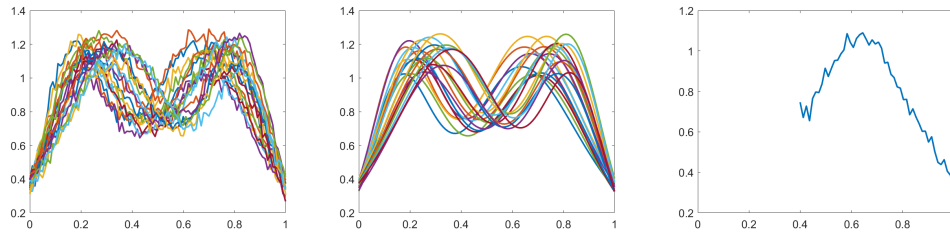


Fig. 5: Noisy (left) and smoothed (middle) observed data. Signal with missing peak (left).

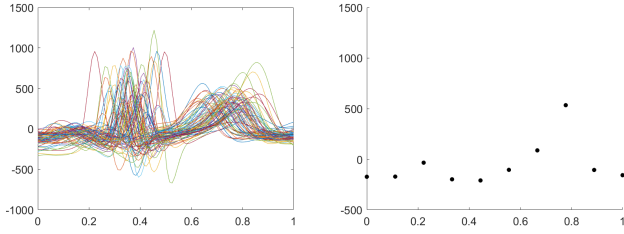


Fig. 6: Smoothed ECG PQRST signals (left), and a sparsely observed signal (right).

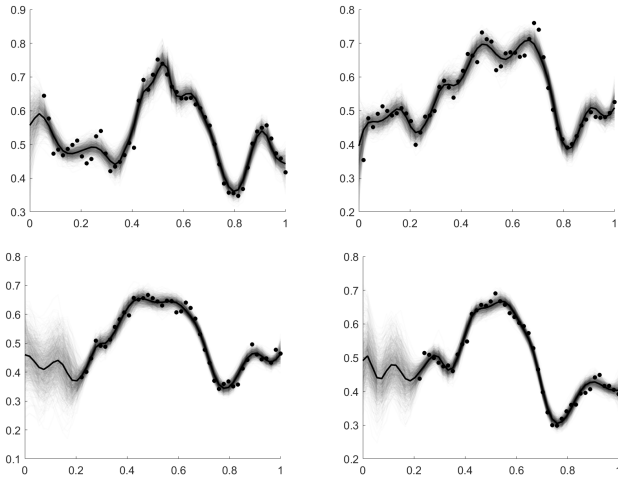


Fig. 7: Two examples of posterior draws and posterior mean where only one or two time points are missing (top), and where 10% of the initial time points are missing (bottom).

We build our amplitude prior using a smoothed dataset shown in the left panel of Figure 6. The results are provided in the bottom row of Figure 4. Although the large middle peak is not observed at all, our model is able to fit a valid PQRST signal to the data. As expected, the uncertainty in the height of the middle peak is large relative to the rest of the function.

Estimation of FA Signals: Finally, we return to the DTI dataset that helped motivate our model. First, we estimate the prior amplitude model using fPCA of smoothed, complete FA signals. While we do not display the modes of variability here, the first mode seems to capture the different slope of the FA signals around the initial point, and the subsequent

modes capture differences in the heights of the various peaks and valleys of the signals. We selected $B = 19$ in the fPCA model, which accounts for about 95% of the overall variability in the complete data. In the incomplete data, there were two common scenarios: (1) only the first one or two observations are missing or (2) first 10% of the observations are missing. In Figure 7, we visualize posterior draws and mean signals for these two common scenarios. The first scenario (top) demonstrates that the model fits the two examples very well, regardless of the difference in phase and amplitude of the two functions. The second scenario (bottom) demonstrates that the model provides structured uncertainty for the posterior mean function; the amount of uncertainty greatly depends on the number of available data points.

IV. CONCLUSIONS

In this paper, we defined a model-based function estimation approach that explicitly accounts for the two forms of variability inherent in functional data. Our model uses empirical prior information to put relevant structure on the amplitude space, and allows uncertainty assessment and visualization for the estimated amplitude and phase functions.

REFERENCES

- [1] Bharath, K., Kurtek, S. (2019). Distribution on warp maps for alignment of open and closed curves. *Journal of the American Statistical Association*, Forthcoming, arXiv:1708.04891v2.
- [2] Goldsmith, J., Greven, S., Crainiceanu, C. (2012). Corrected confidence bands for functional data using principal components. *Biometrics*, 69(1), 41-51.
- [3] Lu, Y., Herbei, R., Kurtek, S. (2017). Registration of functions with a Gaussian process prior. *Journal of Computational and Graphical Statistics*, 26(4), 894-904.
- [4] Ramsay, J.O., Silverman, B.W. (2005). *Functional data analysis*. New York: Springer.
- [5] Shi, M., Weiss, R., 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(2), 151-163.
- [6] Srivastava, A., Klassen, E.P. (2016). *Functional and shape data analysis*. New York: Springer.
- [7] Srivastava, A., Wu, W., Kurtek, S., Klassen, E., Marron, J.S. (2011). Registration of functional data using Fisher-Rao metric. arXiv:1103.3817v2.
- [8] Tucker, J.D., Wu, W., Srivastava, A. (2013). Generative models for functional data using phase and amplitude separation. *Computational Statistics and Data Analysis*, 61, 50-66.
- [9] Wheeler, M., Dunson, D., Herring, A. (2016). Bayesian local extrema splines. *Biometrika*, 104(4), 939-952.
- [10] Yao, F., Miller, H.G., Wang, J.L. (2005). Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association*, 100(470), 577-590.