Geometric empirical Bayesian model for classification of functional data under diverse sampling regimes

James Matuk
The Ohio State University
Columbus, OH, USA
matuk.3@osu.edu

Karthik Bharath
The University of Nottingham
Nottingham, UK
Karthik.Bharath@nottingham.ac.uk

Oksana Chkrebtii
The Ohio State University
Columbus, OH, USA
oksana@stat.osu.edu

Sebastian Kurtek
The Ohio State University
Columbus, OH, USA
kurtek.1@stat.osu.edu

Abstract

Functional data analysis (FDA) is focused on various statistical tasks, including inference, for observations that vary over a continuum, which are not effectively addressed by multivariate methods. A feature of these functional observations is the presence of two distinct forms of variability: amplitude that describes differences in magnitudes of features, e.g., extrema, and phase that describes differences in timings of amplitude features. One area of focus in FDA is the classification of new observations based on previously observed training data that has been split into predefined classes. Existing methods fail to directly account for both phase and amplitude variability, and work under the restrictive assumption that functional observations are measured on a common, fine grid over the input domain. In this work, we address these issues directly by formulating a Bayesian hierarchical model for irregular, fragmented or sparsely sampled functional observations, where training data from different classes are available. Our approach builds on a recently developed inferential framework for incomplete functional observations and the elastic FDA framework for characterizing amplitude and phase variability. The approach operates by inferring individual parameters that separately track amplitude and phase, which can be combined to infer complete functions underlying each observation, and a class parameter, which can be used to discern the class membership of an observation based on the training data. We validate the proposed framework using simulation studies and real data applications, and showcase the advantages of this perspective when both amplitude and phase variability are present in the data.

1. Introduction

Functional data analysis (FDA) is a branch of statistics where entire functions recorded over an input domain are considered as the units of observation. One particular area of focus in FDA is the analysis of functional observations that are grouped into predefined classes, with the goal of inference on class-specific parameters, testing hypotheses about group differences, and assigning a new observation to a predefined class. The work of [1] offers a review of supervised classification methods used to address the latter problem. The authors describe the need for methods specific to functional data, because multivariate statistical approaches do not effectively account for the infinite-dimensional nature of functional observations. Classification is further complicated when the observations exhibit multiple forms of variability inherent in a functional dataset [22]. Specifically, failing to account for amplitude and phase variability, which reflect differences in vertical features and the locations at which these features occur along the input domain, respectively, can lead to misleading estimates [9].

In practice, many FDA methods require functional data consisting of densely sampled function evaluations that are measured on a common observation grid across the input domain [15, 18]. However, it is not always the case that function evaluations are available at the same input locations for all observations in a dataset. Consider the data in Figure 1. The data in panel (a) consist of functional observations from three different classes, where all functions are recorded on a common observation grid. These observations could be, for the most part, described by smooth unimodal functions that exhibit amplitude and phase variability, which describe the differences in magnitude and location of peaks within and between classes. The points

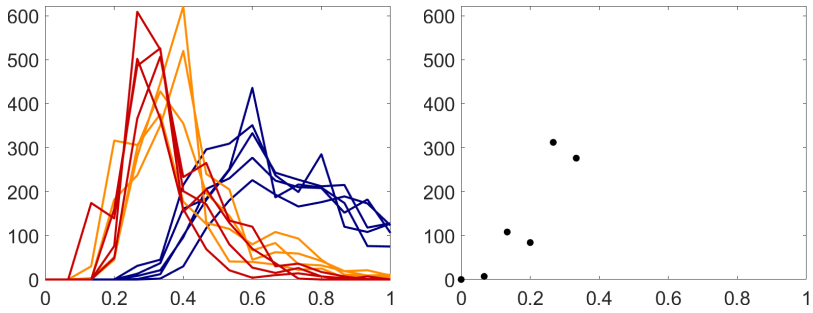

Figure 1. (a) Functional observations recorded over the same observation grid. Colors indicate class membership. (b) A sparse/fragmented observation generated by the same data mechanism as the functions in panel (a).

shown in panel (b) represent a functional observation from the same data-generating process as the functions in panel (a), except this data is sparsely sampled. Methods that assume the same input grid for all observations are not able to answer questions pertaining to class membership, or provide a reasonable estimate of an underlying smooth function that represents each observation in panel (b).

The issues raised by the data in Figure 1 have motivated classification methods tailored to functional data that are recorded on irregular, fragmented or sparse observation grids. These observation regimes are used to describe grids for functional datasets that are not common for all observations, that correspond to subsets of the input domain on which individual observations are recorded, or that correspond to grids with a relatively low number of observation points with respect to the size of the domain. For classification within this context, [8] formulated linear discriminant analysis, [13] formulated a Bayes classifier based on functional mixed effects models, [4] discussed several methods related to quadratic discriminant analysis, and [11] formulated a kernel method. Additionally, for irregular and sparsely observed functional data, the principal analysis through conditional expectation (PACE) framework [25] can be used to perform functional principal component analysis of sparse functions (FPCA), which enables a finite-dimensional representation that has been used in tandem with multivariate classification methods, including quadratic discriminant analysis [20] and support vector machines [23].

None of the aforementioned approaches directly account for amplitude and phase variability, and consequently, the methods may be hindered by confounding of these variabilities [9]. The only related approach that accounts for different sources of variability in observed, noisy data under a Bayesian framework is presented in [17]. There, the authors are interested in a different problem of identifying classes of 2D shapes based on cluttered point clouds extracted from images. In this work, we formulate a Bayesian hierarchical model to address the problem of inference for irregular, fragmented or sparsely sampled functional observations, where training data from different classes are available. The work is based on the inferential framework for functional observations described in [10], and the elastic Functional Data Analysis (EFDA) framework for analyzing amplitude and phase variability [18]. Individual model components that separately track amplitude and phase can be combined to infer a complete function that underlies an observation, and a class parameter can be used to discern the class membership of this observation. The Bayesian modelling paradigm allows for uncertainty quantification for all components of the model.

The rest of this paper is organized as follows. In Section 2, we formulate our model and review the EFDA framework, which is used to study amplitude and phase variability. In Section 3, we validate our model using simulation experiments and present real data applications. In Section 4, we summarize our approach and discuss directions for future work.

2. Model Formulation

In this section, we review the EFDA framework [18] and formulate the proposed Bayesian hierarchical model.

2.1. Amplitude Principal Component Analysis

The proposed Bayesian model relies on FPCA of amplitude variability to efficiently represent amplitude features of functional datasets with empirical bases. This is achieved through the EFDA framework.

The primary goal of EFDA is to separate phase and amplitude variability present in a functional dataset so that they can be analyzed individually [18]. This is an important task, since these variabilities provide complementary information, and failing to take these into account leads to misleading analyses [9]. The separation of amplitude and phase variabilities is achieved through a registration procedure, which decomposes a functional dataset into ampli-

tude components and phase components, so that their composition yields the original dataset; after registration, function features, e.g., extrema, are captured by the amplitude components and occur at the same locations along the input domain. The registration procedure is formalized in [19], which defines an optimality criterion based on the extended Fisher-Rao Riemannian (eFR) metric. This formulation has desirable properties for the registration problem, i.e., invariance to simultaneous warping of functions. Furthermore, the complicated eFR metric can be simplified to the standard $\mathbb{L}^2$ metric through the square-root velocity function (SRVF) representation. Without loss of generality, we restrict our attention to absolutely continuous functions on $[0,1]$. Then, for a function $f \in \mathcal{F} = \{f : [0,1] \to \mathbb{R} \mid f \text{ is absolutely continuous}\}$, its SRVF q is defined through the bijective map $Q : \mathcal{F} \to \mathbb{L}^2([0,1]) =: \mathcal{Q}$,

$$q(t) = Q(f)(t) := \frac{\dot{f}(t)}{\sqrt{|\dot{f}(t)|}}, \tag{1}$$

where $\dot{f}$ denotes the derivative of the function f; if $\dot{f}(t) = 0$, then $q(t) = 0$. Since Q is bijective given a function's initial point, Q^{-1} is given by:

$$f(t) = Q^{-1}(q, f(0))(t) := f(0) + \int_0^t q(s)|q(s)|ds. \tag{2}$$

EFDA defines sample statistics, e.g., averages and covariance functions, under the eFR metric through the SRVF representation as follows. For a functional dataset $f_1, \ldots, f_n$, the mean amplitude SRVF is defined as

$$\hat{\mu}_q := \underset{q \in \mathcal{Q}}{\operatorname{argmin}} \sum_{i=1}^{n} \inf_{\gamma \in \Gamma} d_{\mathbb{L}^2}(q, Q(f_i \circ \gamma))^2, \tag{3}$$

where $d_{\mathbb{L}^2}(\cdot,\cdot)$ denotes the $\mathbb{L}^2$ distance between its arguments and $\gamma \in \Gamma = \{\gamma : [0,1] \to [0,1] \mid \gamma(0) = 0, \gamma(1) = 1 \text{ and } \dot{\gamma} > 0\}$ represents the set of warping (phase) functions of $[0,1]$. The minimizers inside the summation are constrained so that their average is the identity warping function, $\gamma_{id}(t) = t$, to ensure identifiability of the amplitude mean [19]. The set of phase functions, $\hat{\gamma}_i = \underset{\gamma_i \in \Gamma}{\operatorname{arginf}}\, d_{\mathbb{L}^2}(\hat{\mu}_q, Q(f_i \circ \gamma)),\ i = 1 \ldots, n$, and the set of amplitude SRVFs, $\tilde{q}_i = Q(f_i \circ \hat{\gamma}_i),\ i = 1, \ldots, n$, serve as the phase and amplitude components of the original data, respectively.

The work of [22] describes implementations of FPCA for amplitude and phase variabilities based on function registration. For the amplitude component, which is of interest in the current work, the sample amplitude SRVF covariance function is defined as,

$$\hat{K}(s,t) = \frac{1}{n-1} \sum_{i=1}^{n} (\tilde{q}_i(s) - \hat{\mu}_q(s))(\tilde{q}_i(t) - \hat{\mu}_q(t)). \tag{4}$$

Amplitude FPCA is enabled through the Karhuenen-Loève expansion and eigendecomposition of the sample covariance function,

$$\hat{K}(s,t) = \sum_{b=1}^{\infty} \hat{\lambda}_b \hat{\phi}_b(s) \hat{\phi}_b(t), \tag{5}$$

where the principal components, $\hat{\phi}_b,\ b = 1, \ldots$ are orthogonal functions that represent dominant modes of amplitude SRVF variability, and $\hat{\lambda}_b,\ b = 1, \ldots$ are the variances along these modes of variability (arranged in descending order). The principal components serve as an efficient dimension reduction tool, since an amplitude SRVF can be approximated by a B-dimensional vector of coefficients, for some finite B, through

$$\tilde{q}_i(t) \approx \hat{\mu}_q(t) + \sum_{b=1}^{B} \hat{\beta}_{bi} \hat{\phi}_b(t), \tag{6}$$

where $\hat{\beta}_{bi} = \int_0^1 (\tilde{q}_i(s) - \hat{\mu}_q(s)) \hat{\phi}_b(s) ds,\ b = 1, \ldots, B$. We leverage this efficient representation of amplitude SRVFs in the Bayesian model formulation in the next section.

2.2. Proposed Bayesian Model Formulation

We extend the modelling framework specified in [10] for a noisy observation, $\mathbf{y}$, measured on an irregular, fragmented or sparse observation grid, $\mathbf{t} = (t_1, \ldots, t_m)^\top$, to the case where densely sampled training data are available from C classes, $f_i^c,\ i = 1, \ldots, n_c,\ c = 1, \ldots, C$. We use the EFDA framework, Equations 3-5, to summarize the amplitude variability for the functions in each class, and use $\mu_q^c,\ \hat{\phi}_1^c, ..., \hat{\phi}_B^c,\ \hat{\lambda}_1^c, ..., \hat{\lambda}_B^c,\ c = 1, \ldots, C$ to denote class specific amplitude means, principal directions, and variances associated with the directions, respectively. We fix a number of components, B, that explains a large portion of variability across all of the classes. Throughout this section, we use $f(\mathbf{t}) = (f(t_1), \ldots, f(t_m))^\top$ to denote a vector of function evaluations.

In the proposed model, a smooth functional parameter that underlies the noisy and irregular observation is decomposed into amplitude and phase components. A latent class parameter determines the class membership for the observation as well as its amplitude features, through Equation 6. Our model for the phase component is based on a recently developed model for warping functions [2] that is consistent with the goal of modelling phase for irregular, fragmented or sparsely sampled observations. The phase prior is specified by discretizing the phase parameter, γ, on a coarse observation grid on $(0,1)$, $\mathbf{t}_\gamma$, and modelling differences between adjacent components, $p(\gamma(\mathbf{t}_\gamma)) := (\gamma(t_{1,\gamma}), \ldots, \gamma(t_{j+1,\gamma}) - \gamma(t_{j,\gamma}), \ldots, 1 - \gamma(t_{m_\gamma,\gamma})^\top$; here, m_γ specifies the fineness of the discretization for γ.

Using $N_m(\mu, \Sigma)$ to denote an m-dimensional Gaussian density with mean vector μ and covariance Σ, we specify the following empirical amplitude Bayesian hierarchical model for an observation $\mathbf{y}$:

$$\begin{aligned}
\mathbf{y}|c, \beta, \gamma, T, \sigma^2 &\sim N_m\left((Q^{-1}(\tilde{q}^c, T) \circ \gamma)(\mathbf{t}),\, \sigma^2 I_m\right), \\
\tilde{q}^c &= \hat{\mu}^c + \sum_{b=1}^{B} \beta_b \hat{\phi}_b^c, \\
\beta|c &\sim N_B(\mathbf{0}_B, \text{diag}(\hat{\lambda}_1^c, \ldots, \hat{\lambda}_B^c)), \\
T|c &\sim N(\hat{\mu}_T^c, \hat{\tau}_T^c), \\
p(\gamma(\mathbf{t}_\gamma))|c &\sim \text{Dirichlet}(\hat{\theta}_\gamma^c \mathbf{1}_{m_\gamma+1}), \\
c &\sim \text{Discrete-Uniform}(\{1, \ldots, C\}), \\
\sigma^2 &\sim \text{Inverse-Gamma}(\alpha_\sigma, \beta_\sigma). \qquad (7)
\end{aligned}$$

In this formulation, the smooth function $Q^{-1}(\tilde{q}^c, T) \circ \gamma$ that underlies the observation $\mathbf{y}$ is decomposed into components that describe amplitude, $\tilde{q}^c$, phase, γ, and translation, T. The amplitude, phase and translation components depend on the parameter c, which determines the class that most appropriately describes the underlying function; marginal posterior inference on this parameter quantifies the uncertainty in class membership.

The priors for the amplitude, phase and translation parameters are consistent with those described in [10]. The hyperparameters are estimated using the training data within each class. For amplitude, the variance components, $\hat{\lambda}_1^c, \ldots, \hat{\lambda}_B^c$, of the basis coefficient vector, β, correspond to the variability described by the B leading amplitude principal components of a given class. For translation, the mean and variance of the prior correspond to the mean and variance of the initial points, $f_1^c(0), \ldots, f_{n_c}^c(0)$, of training data in a given class. For phase, the concentration parameter, $\hat{\theta}^c$, is determined through maximum likelihood estimation based on the following model for phase functions estimated via registration of training data from a given class, $p(\hat{\gamma}_1^c(\mathbf{t}_\gamma)), \ldots, p(\hat{\gamma}_{n_c}^c(\mathbf{t}_\gamma)) \overset{\text{iid}}{\sim} \text{Dirichlet}(\theta_\gamma^c \mathbf{1}_{m_\gamma+1})$. We do not favor any particular value for the class parameter c a-priori, and specify a discrete-uniform prior on the class labels. As in [10], the prior for σ^2 is chosen to be diffuse and conjugate.

We base posterior inference on Markov chain Monte Carlo samples over the unknown parameters. Because amplitude, phase and translation are all relative with respect to the underlying class parameter, their marginal posterior distributions are potentially multimodal when there is considerable uncertainty in the underlying class of an observation. Consequently, we use an adaptive parallel tempering algorithm, similar to that of [21], where the temperature scheme [6] and the proposal parameters of the Metropolis-within-Gibbs algorithm [16] are automatically tuned for efficiency. We have found that this is a flexible sampling technique that is able circumvent these potential multimodality issues.

3. Applications

In this section, we assess our model on simulated and real datasets. In all examples, we show our inferential results for functions that underlie irregularly sampled observations. Additionally, we base a probabilistic classifier on the marginal posterior distribution over the class parameter. When comparing results to other classification methods, we use the initials EA to denote the classification results based on the proposed empirical amplitude Bayesian hierarchical model.

As a benchmark, we use the approach of [8], where linear discriminant analysis is defined for irregularly sampled functional data. In this formulation, class-specific mean functions and a covariance function are estimated based on training data to define a discriminant function that assigns a new observation to one of the classes. For implementation, we use publicly available MATLAB code under default settings, which is available at `http://faculty.marshall.usc.edu/gareth-james/Research/Research.html`. We refer to results from this method using LD.

Additional methods used for comparison are based on the PACE framework designed to analyze sparse and irregularly sampled functional data [25]. The PACE procedure estimates a mean function and functional principal component basis functions, and is implemented in the PACE MATLAB toolbox [24]. The training data is first projected onto the estimated basis to compute a finite dimensional representation. Then, the resulting training data and associated class labels are used to train different classifiers. In this work, we consider the k-nearest neighbors classifier (KN), random forests (RF), and support vector machines (SV). All of these approaches are implemented in the Python package scikit-learn [14], where corresponding hyperparameters are tuned based on k-fold cross validation. To classify a new observation, it is first projected onto the basis estimated via PACE; the class assignment is determined according to the trained classifier rule.

For all simulated examples, datasets are first split into training and testing sets, and the classification methods are compared using two different measures: classification accuracy (proportion of testing data that was assigned to the correct class) and the multi-category Brier score [3] (squared difference between predicted probability of class membership and the actual class of an observation); we aim for high classification accuracy and low Brier scores.

3.1. Simulated Example

This simulated example was designed to understand the effects of phase variability and severity of fragmentation

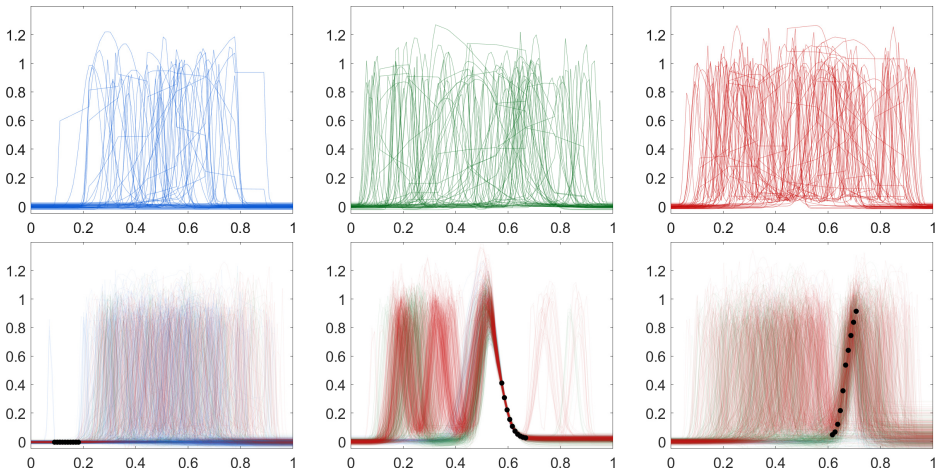

Figure 2. Example with high warping variability and severe fragmentation. Bottom: Model fit for three different observations from the testing data (one from each class). The black points represent the observations, and the lines represent posterior draws of a complete function, i.e., the composition of posterior draws of amplitude and phase, colored by the corresponding posterior draw of the class label.

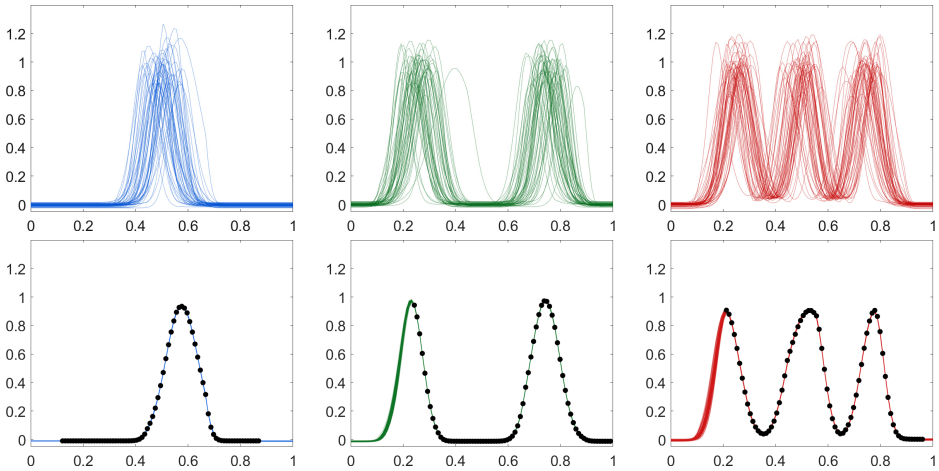

Figure 3. Example with low warping variability and mild fragmentation. Top: Training data from each class. Bottom: Model fit for three different observations from the testing data (one from each class). The black points represent the observations, and the lines represent posterior draws of a complete function, i.e., the composition of posterior draws of amplitude and phase, colored by the corresponding posterior draw of the class label.

on modelling results. Training data were generated from three different classes corresponding to functions that have one, two or three peaks; the amplitude components of functions within the same class only vary according to the height of the peaks. For each simulation setting, the amplitude components are randomly warped by phase functions generated from the probability model $p(\gamma(\mathbf{t}_\gamma)) \sim \text{Dirichlet}(\theta_\gamma \mathbf{1}_{m_\gamma+1})$. We consider three different values of $\theta_\gamma = 5, 25, 125$, corresponding to varying degrees of warping. Testing data are generated in the same way, except that they are fragmented so that only 10%, 25%, 50% or 75% of each function is observed. In total, there are 12 settings in this simulation that correspond to different pairings of phase variability and portion of testing data observed. For these 12 different simulation settings, 50 testing and 50 training observations are simulated in each class. The top row in Figure 2 shows an example training dataset with high warping variability. The black points in the bottom row correspond to testing observations with severe fragmentation. On the other hand, the top row in Figure 3 shows an example training dataset with low warping variability; the black points in the bottom row correspond to testing data with mild fragmentation. In the bottom panel in each of these two figures, we also display the posterior samples of the complete function generated from the proposed model (each function is a posterior amplitude sample composed with the correspond-

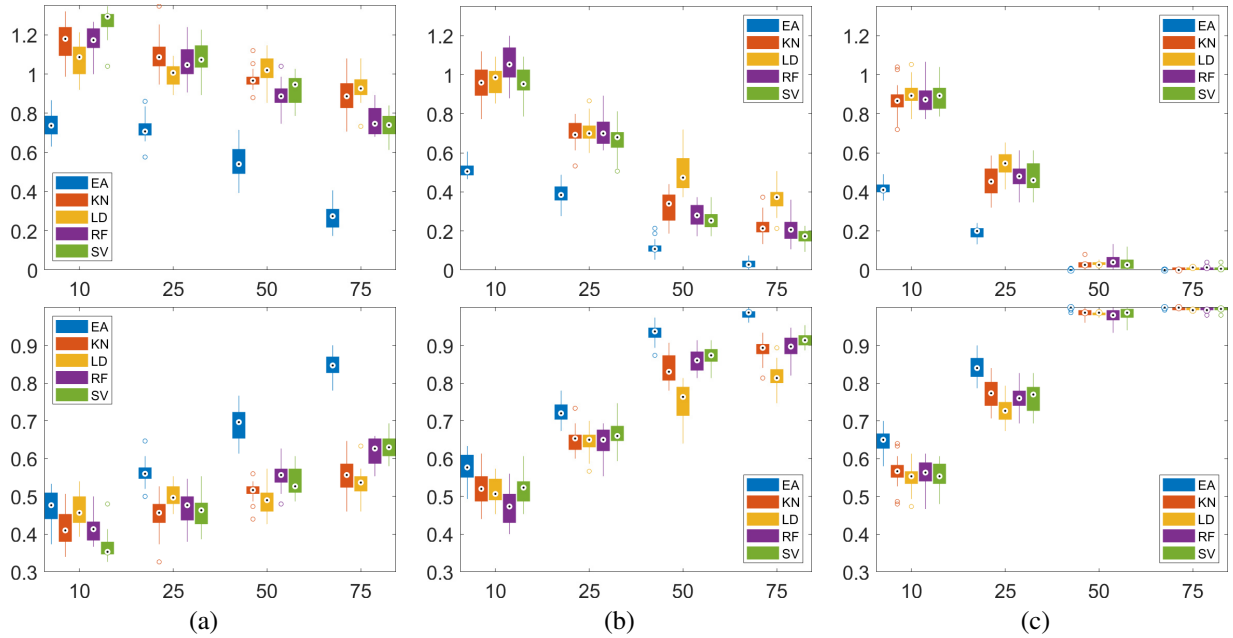

Figure 4. Boxplots of the Brier score (top) and correct classification rate (bottom) based on 20 replicates of the simulations described in Section 3.1. Each plot shows the result for different levels of fragmentation in the testing data (10% corresponds to severe fragmentation and 75% corresponds to mild fragmentation). (a) High phase variability ($\theta_\gamma = 5$). (b) Moderate phase variability ($\theta_\gamma = 25$). (c) Low phase variability ($\theta_\gamma = 125$).

ing posterior phase sample); each posterior sampled is colored according to the corresponding posterior sample of the class label. In the bottom row of Figure 2, we see that there is a lot of uncertainty in the estimated function and its class membership. This is due to the severe fragmentation of the observation. On the other hand, in the bottom row of Figure 3, the uncertainty is extremely small, both in the structure of the function and its class label. Each of the 12 different simulation scenarios was replicated 20 times. The classification results based on the proposed Bayesian model and the other four methods are compared in Figure 4. Each row in this figure corresponds to the two measures of classification performance, with Brier score in the top row and accuracy in the bottom row. In these simulated examples, the proposed Bayesian model is better at accounting for phase variability and different levels of fragmentation as suggested by the low Brier scores and high correct classification rates.

3.2. Phoneme Data with Simulated Fragmentation

To further study the effects of fragmentation on modelling results, we consider the phoneme training and testing datasets available at https://www.math.univ-toulouse.fr/~ferraty/SOFTWARES/NPFDA/index.html. As described in [5], the data correspond to log-periodograms of recordings of phonemes, which are units of sound that combine to form words, from five different classes "sh" as in "she", "dcl" as in "dark", "iy" as in "she", "aa" as the vowel in "dark" and "ao" as the first vowel in "water". Again, we simulated fragmentation so that only 10%, 25%, 50% or 75% of the testing functions were observed in different simulation settings. Figure 5 displays the training phoneme data in the top panel; the thick function is the amplitude mean in each class. In the middle panel, we show five different observations (black points) under severe fragmentation (10% observed); we also show the posterior samples of the complete function (each function is a posterior amplitude sample composed with the corresponding posterior phase sample) colored by the corresponding posterior sample of the class label along with the pointwise posterior mean estimated by conditioning on the class parameter. In the bottom panel, we show similar results when 75% of the testing function is observed. In the middle panel, when only small portions of the testing data are observed, there is considerable uncertainty regarding the structure of the complete function and its class assignment. In the bottom panel, both sources of uncertainty are much smaller. The four fragmentation scenarios were replicated 20 times where we used 250 complete functions as training data (50 per class) and 250 fragmented functions as testing data (50 per class). As in the previous simulated example, we compared the performance of the proposed model to the four competing methods. The results are shown in Figure 6 with Brier scores in the left panel and correct classification rates in the right panel. In terms of the Brier score,

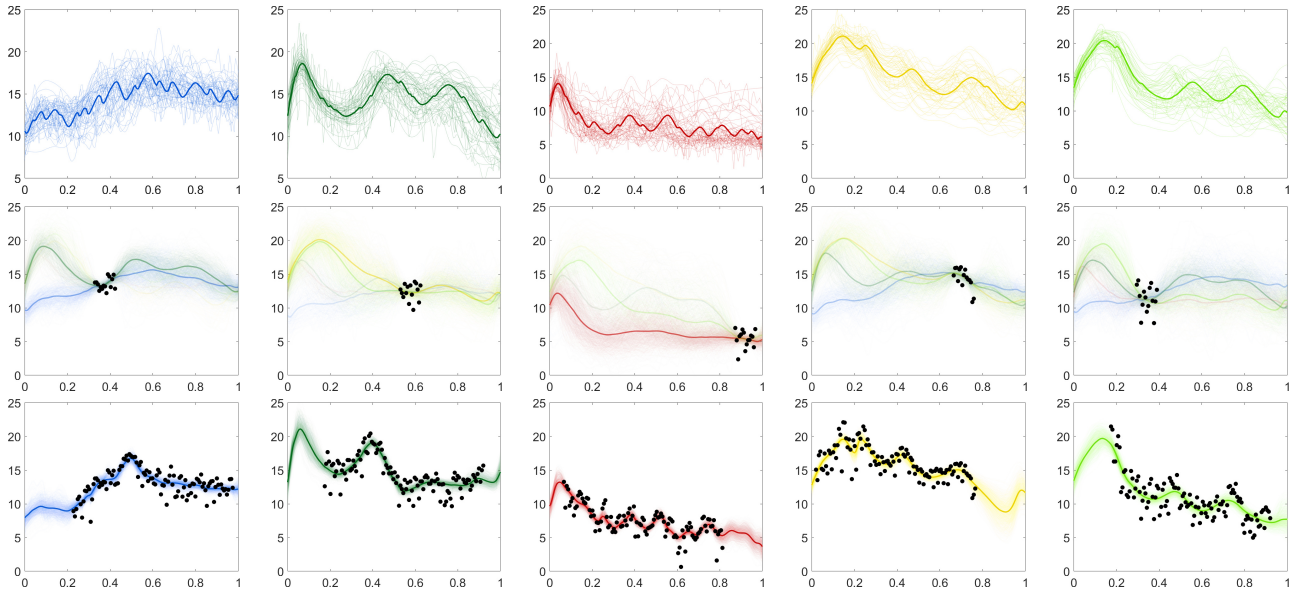

Figure 5. Top: Example phoneme training data. Middle and Bottom: Model fit for five different observations from the testing data (one from each class) under severe (10% observed) and mild (75% observed) fragmentation, respectively. The black points represent the observations, and the lines represent posterior draws of a complete function, i.e., the composition of posterior draws of amplitude and phase, colored by the corresponding posterior draw of the class label.

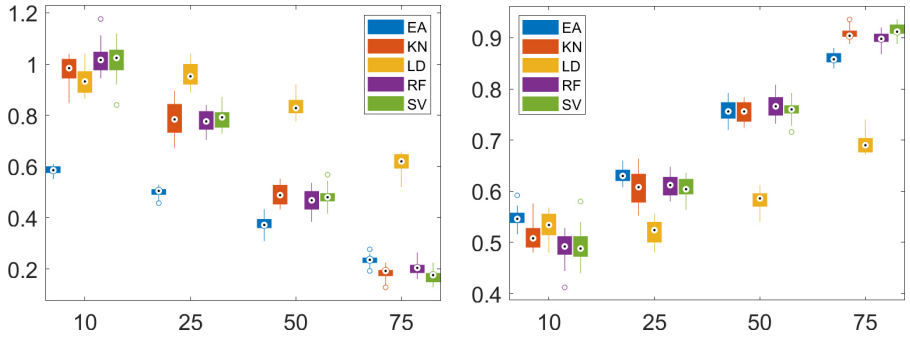

Figure 6. Boxplots of the Brier score (left) and correct classification rate (right) based on 20 replicates of the simulations described in Section 3.2. Each plot shows the result for different levels of fragmentation in the testing data (10% corresponds to severe fragmentation and 75% corresponds to mild fragmentation).

our approach significantly outperforms the other methods under the high and moderate fragmentation scenarios; the performance of our model and the PACE-based approaches is comparable when 75% of each testing function is observed. This suggests that the proposed model appropriately accounts for the uncertainty in the class label and thus leads to lower Brier scores when small portions of the testing data are observed. The overall correct classification rates are similar across all methods. It appears that the LD approach is inferior to the other four methods under most settings.

3.3. Conidial Discharge of Fungi Data

Finally, we return to the motivating example in Figure 1. This data comes from [7], and has been used to study environmental factors related to conidial spore discharge of fungi. In this work, 15 experimental units were incubated at different temperatures: cold (12°C), warm (18.5°C) and hot (25°C). Then, in eight hour intervals, the intensity of conidial discharge was measured in each of the experimental units over the span of 120 hours. In [12], functional data analysis methods were used to discern if there were any significant differences among experimental conditions in the timing of peak discharge and the rate of decay following the peak. There were five, four and four complete experimental units for the cold, warm and hot conditions, respectively. Two of the experimental units, one warm and the other hot, stopped discharging conidia at 48 and 40 hours, and were recorded only partially. These experimental units were dis-

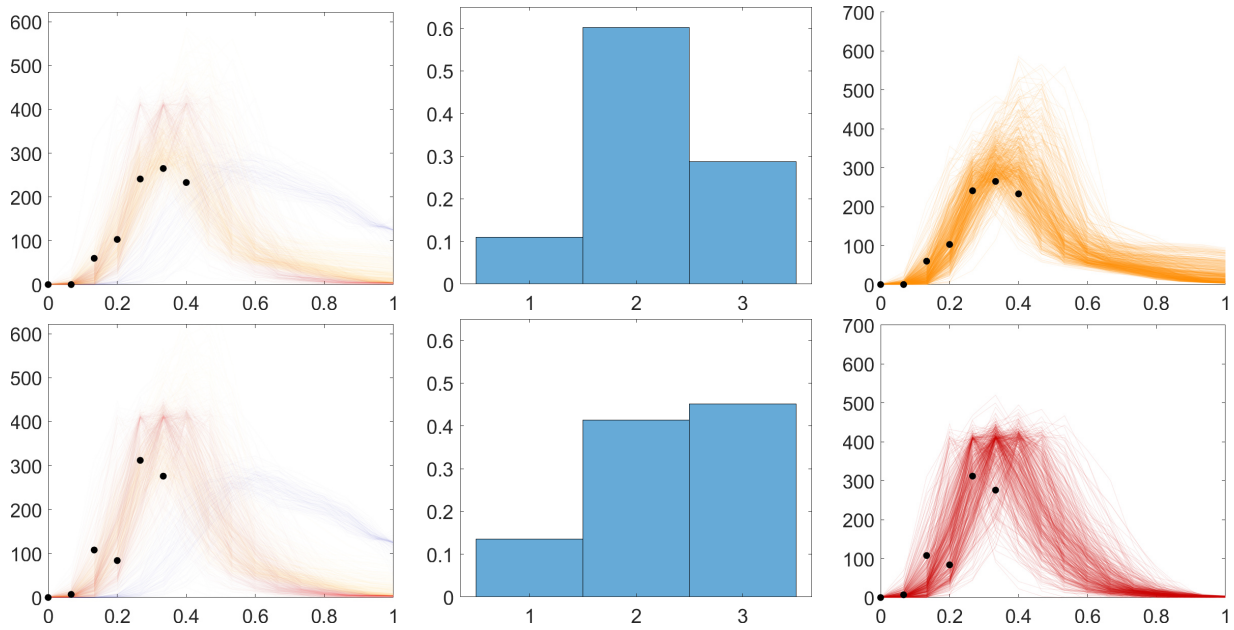

Figure 7. (a) Model fit for the two incomplete observations (black points). The lines represent posterior draws of a complete function, i.e., the composition of posterior amplitude and phase components, colored by the corresponding posterior draw of the class label. (b) Histogram of marginal posterior draws of the class labels (1 = cold, 2 = mild, 3 = hot). (c) Model fit conditioned on the mode of the marginal posterior of the class label.

regarded in the original study.

For this work, we considered complete experiments as training data, and the incomplete experiments as fragmented observations. As such, we wish to infer the experimental condition (i.e., class) and the remaining portion of the function that describes the discharge for the incomplete experiments. Figure 7(a) presents posterior samples from the proposed model for the two incomplete observations (black points). Each posterior sample is a composition of a posterior draw of amplitude and phase components, and is colored according to the corresponding class label. Panel (b) shows a histogram of marginal posterior draws of the class labels for each of the two observations, where 1 corresponds to cold, 2 corresponds to warm, and 3 corresponds to hot. Finally, in panel (c), we show the model fit conditioned on the mode of the marginal posterior distribution for the class label. In both cases, the proposed model favors the experimental condition that generated the incomplete data and offers reasonable estimates of the complete function. We are also able to assess the uncertainty associated with the intensity of conidial discharge for the entire measurement period. It appears that the proposed model has fairly high confidence that the first observation came from the warm experimental condition; for the second observation, the marginal posterior probabilities of the warm and hot conditions are very similar.

4. Discussion and Future Work

In this paper, we propose a Bayesian model for inference on incomplete functional data, when training data from multiple classes are available. In particular, we model the amplitude and phase components of the underlying functions separately. Based on simulated and real data examples, the proposed model outperforms other methods in classification tasks based on irregularly sampled functional data, especially when a relatively small part of the observation is recorded or high phase variability is present.

One shortcoming of the presented model is the need for densely-observed training data. In future work, we plan to incorporate more hierarchical layers into the model, so that the training data can possibly be irregularly sampled as well. This would allow for uncertainty propagation from the training stage of the model to the inferential stage. We also plan to reformulate the model to apply to higher-dimensional functional data, and in particular, shapes.

5. Acknowledgements

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 DMS2015226 and NSF CCF-1839252 (to SK).

References

[1] A. Baìllo, A. Cuevas, and R. Fraiman. Classification methods for functional data. *The Oxford Handbook of Functional Data Analysis*, pages 259–297, 01 2011.

[2] K. Bharath and S. Kurtek. Distribution on warp maps for alignment of open and closed curves. *Journal of the American Statistical Association*, 115(531):1378–1392, 2020.

[3] G. W. Brier. Verification of forecasts expressed in terms of probability. *Monthly Weather Review*, 78(1):1 – 3, 01 Jan. 1950.

[4] A. Delaigle and P. Hall. Classification using censored functional data. *Journal of the American Statistical Association*, 108(504):1269–1283, 2013.

[5] F. Ferraty and P. Vieu. Curves discrimination: a nonparametric functional approach. *Computational Statistics & Data Analysis*, 44(1):161–173, 2003. Special Issue in Honour of Stan Azen: a Birthday Celebration.

[6] C. Geyer. Markov chain Monte Carlo maximum likelihood. In *Computing Science and Statistics, Proceedings of the 23rd Symposium on the Interface, 156*. American Statistical Association, 1991.

[7] P. Herren. Conidial discharge of pandora sp., a potential biocontrol agent against cacopsylla spp. in apple and pear orchards. Master's thesis, University of Copenhagen, Denmark, 2018.

[8] G. M. James and T. J. Hastie. Functional linear discriminant analysis for irregularly sampled curves. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(3):533–550, 2001.

[9] J. S. Marron, J. O. Ramsay, L. M. Sangalli, and A. Srivastava. Functional data analysis of amplitude and phase variation. *Statistical Science*, 30(4):468–484, 2015.

[10] J. Matuk, K. Bharath, O. Chkrebtii, and S. Kurtek. Bayesian framework for simultaneous registration and estimation of noisy, sparse and fragmented functional data. *Journal of the American Statistical Association*, (in press), 2021.

[11] M. Mojirsheibani and C. Shaw. Classification with incomplete functional covariates. *Statistics & Probability Letters*, 139:40–46, 2018.

[12] N. L. Olsen, P. Herren, B. Markussen, A. B. Jensen, and J. Eilenberg. Statistical modelling of conidial discharge of entomophthoralean fungi using a newly discovered pandora species. *PLOS ONE*, 14(5):1–19, 05 2019.

[13] Y. Park and D. G. Simpson. Robust probabilistic classification applicable to irregularly sampled functional data. *Computational Statistics & Data Analysis*, 131:37 – 49, 2019. High-dimensional and functional data analysis.

[14] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[15] J. Ramsay and B. Silverman. *Functional Data Analysis*. Springer, 2005.

[16] G. O. Roberts and J. S. Rosenthal. Examples of adaptive mcmc. *Journal of Computational and Graphical Statistics*, 18(2):349–367, 2009.

[17] A. Srivastava and I. H. Jermyn. Looking for shapes in two-dimensional cluttered point clouds. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(9):1616–1629, 2009.

[18] A. Srivastava and E. Klassen. *Functional and Shape Data Analysis*. Springer, 2016.

[19] A. Srivastava, W. Wu, S. Kurtek, E. Klassen, and J. S. Marron. Registration of functional data using Fisher-Rao metric. *arXiv 1103.3817*, 2011.

[20] M. Stefanucci, L. M. Sangalli, and P. Brutti. Pca-based discrimination of partially observed functional data, with an application to aneurisk65 data set. *Statistica Neerlandica*, 72(3):246–264, 2018.

[21] J. Strait, O. Chkrebtii, and S. Kurtek. Parallel tempering strategies for model-based landmark detection on shapes. *Communications in Statistics - Simulation and Computation*, 0(0):1–21, 2019.

[22] J. Derek Tucker, W. Wu, , and A. Srivastava. Generative models for functional data using phase and amplitude separation. *Computational Statistics & Data Analysis*, 61:50–66, 2013.

[23] Y. Wu and Y. Liu. Functional robust support vector machines for sparse and irregular longitudinal data. *Journal of Computational and Graphical Statistics*, 22(2):379–395, 2013. PMID: 23734071.

[24] F. Yao and J. L. Wang H. G. Müller. *PACE package for Functional Data Analysis and Empirical Dynamics (written in Matlab)*, 2015. MATLAB package version 2.17.

[25] F. Yao, H. G. Müller, and J. L. Wang. Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association*, 100(470):577–590, 2005.