

RESEARCH ARTICLE

Elastic Functional Principal Component Regression[†]

J. Derek Tucker¹ | John R. Lewis¹ | Anuj Srivastava²

¹Statistical Sciences, Sandia National Laboratories, Albuquerque, New Mexico, USA

²Department of Statistics, Florida State University, Tallahassee, Florida, USA

Correspondence

*J. Derek Tucker PO Box 5800 MS 1202, Albuquerque, NM 87185 Email:
jdtuck@sandia.gov

Summary

We study regression using functional predictors in situations where these functions contain both phase and amplitude variability. In other words, the functions are misaligned due to errors in time measurements, and these errors can significantly degrade both model estimation and prediction performance. The current techniques either ignore the phase variability, or handle it via preprocessing, i.e., use an off-the-shelf technique for functional alignment and phase removal. We develop a functional principal component regression model which has a comprehensive approach in handling phase and amplitude variability. The model utilizes a mathematical representation of the data known as the square-root slope function. These functions preserve the \mathbb{L}^2 norm under warping and are ideally suited for simultaneous estimation of regression and warping parameters. Using both simulated and real world data sets, we demonstrate our approach and evaluate its prediction performance relative to current models. In addition, we propose an extension to functional logistic and multinomial logistic regression.

KEYWORDS:

Compositional noise, functional data analysis, functional Principal Component Analysis, functional regression

1 | INTRODUCTION

The statistical analysis of functional data is fast gaining prominence in the statistics community because this kind of “big data” is central to many applications. For instance, the functional data can be found in a broad swath of application areas

[†]This paper describes objective technical results and analysis. Any subjective views or opinions that might be expressed in the paper do not necessarily represent the views of the U.S. Department of Energy or the United States Government. Supported by the Laboratory Directed Research and Development program at Sandia National Laboratories, a multi-mission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy’s National Nuclear Security Administration under contract DE-NA0003525. This research was in part supported by the National Technical Nuclear Forensics Center (NTNFC) of the U.S. Department of Homeland Security (DHS). The authors would like to thank Dr. Marc Welliver at SNL for his technical support during this work.

ranging from biology, medicine, and chemistry to geology, sports, and financial analysis. In this problem, some of the random quantities of interest are functions of independent variables, (e.g., time, frequency), and are studied as elements of an appropriate function space, often a Hilbert space. The analysis can include common statistical procedures such as computing statistical summaries, estimating parametric and nonparametric distributions, and generating inferences under noisy observations. One common problem in functional data analysis is regression modeling where the function variables are used as predictors to estimate a scalar response variable.

More precisely, let the predictor functions be given by $\{f_i : [0, T] \rightarrow \mathbb{R}, i = 1, 2, \dots, n\}$ and the corresponding observations of a single variable be y_i . The standard functional linear regression model for this set of observations is

$$y_i = \alpha + \int_0^T f_i(t)\beta(t) dt + \epsilon_i, \quad i = 1, \dots, n \quad (1.1)$$

where $\alpha \in \mathbb{R}$ is the intercept, $\beta(t)$ is the regression-coefficient function and $\epsilon_i \in \mathbb{R}$ are random errors. This model was first studied by Ramsay and Dalzell [28] and Cardot et al. [5]. The model parameters are usually estimated by minimizing the sum of squared errors (SSE),

$$\{\alpha^*, \beta^*(t)\} = \arg \min_{\alpha, \beta(t)} \sum_{i=1}^n |y_i - \alpha - \int_0^T f_i(t)\beta(t) dt|^2.$$

These values form maximum-likelihood estimators of parameters under the additive white-Gaussian noise model. One problem with this approach, is that for any finite n , since β is a full function there are infinitely many solutions for β without imposing any further restrictions. In other words, it is an element of an infinite-dimensional space while its specification for any n is finite dimensional. Ramsay and Silverman [29] proposed two approaches to handle this issue: (1) Represent $\beta(t)$ using p basis functions in which p is kept large to allow desired variations of $\beta(t)$ ($p \leq n$), and (2) add a roughness penalty term to the objective function (SSE) which selects a smooth solution by finding an optimal balance between the SSE and the roughness penalty. The basis can come from Fourier analysis, splines, or fPCA (Reiss and Ogden [30]).

Current literature in functional linear regression is focused primarily on the estimation of the coefficient of $\beta(t)$ under a basis representation. For example, [8, 6, 13, 15] discuss estimation and/or inference of $\beta(t)$ for different cases for the standard functional linear model and the interpretation of $\beta(t)$. Focusing on prediction of the scalar response, Cai and Hall [4] studied the estimation of $\int f_i(t)\beta(t) dt$. In some situations the response variable y_i is categorical and the standard linear model will not suffice. James [14] extended the standard functional linear model to functional logistic regression to be able to handle such situations. Müller and Stadtmüller [27] extend the generalized model to contain a dimension reduction by using a truncated Karhunen-Loève expansion. Recently, Gertheiss et al. [10] included variable selection to reduce the number of parameters in the generalized model.

In practice the predictor functions are observed at discrete points and not the full interval $[0, T]$. Furthermore, in some situations, these observations are corrupted by noise along the time axis. That is, one observes $\{(t + \eta(t), f(t))\}$ instead of $\{(t, f(t))\}$ where the random variables $\eta(t)$ are constrained so that the observation times do not cross each other. While some papers have assumed parametric models for $\eta(t)$ (Carroll et al. [7]) and incorporated them in the estimation process, the others have ignored them completely. It is more natural to treat these measurement variables in a nonparametric form as follows: We assume that observation times are given by $\gamma(t)$ where γ is a monotonic function with appropriate boundary conditions ($\gamma(0) = 0$, $\gamma(T) = T$). Consequently, the observations are modeled as $\{\gamma(t), f(t)\}$ where γ captures a random noise component that needs accounting for in the estimation process. The effect of γ is a warping of f with a nonlinear shift in the *locations* of peaks and valleys but no changes in the heights of those peaks and valleys; this warping differs across realizations (observations) and, hence, is termed as *warping or compositional noise*. Some authors have also called it the phase variability in functional data Srivastava and Klassen [34], Marron et al. [25]. If the phase variability is ignored, the resulting model may fail to capture patterns present in the data and will lead to inefficient data models. One way to handle this noise is to capture both the phase and amplitude variability properly in the regression model. It is more natural to include handling of warping noise, or alignment, in the regression model estimation itself and perform a joint inference on all model variables under the same objective function. Recently, Gervini [11] has proposed a functional linear regression model that includes phase variability in the model. It uses a random-effect simultaneous linear model on the warping parameters and the principal component scores (of aligned predictor functions). However, this method involves fPCA on the original functional space and has shown to be inferior for unaligned data (Tucker et al. [37];Lee and Jung [23]). Tucker et al. [37] showed that if this variability is not accounted for properly when performing fPCA, the results will be misleading due to incorrect shape of the calculated mean function.

In this paper we focus on problems where the functional data contains random phase variability. To handle that variability, we propose a regression model that incorporates the phase variability through the use of functional principal component regression (fPCR) where this variability is handled in a parsimonious way. The basic idea is to use a fPCA method as the basis that is able to capture the amplitude variability, phase variability, or both, in the regression problem. This allows the model to capture the variability that is important in predicting the outcome from the data. Using this representation and the geometry of the warping function γ , we construct the model and outline the resulting prediction procedures. The fPCR method was first proposed by Reiss and Ogden [30], but they fail to account for the phase variability found in functional data. Additionally, one could use a robust fPCA method such as Bali et al. [1], but this method also fails to account for phase and amplitude variability explicitly in the data. We then extend this framework to the logistic regression case where the response can take on categorical data. We will illustrate this application using both simulated

and real data sets, which includes sonar, gait, and electrocardiogram data. The physiological data is studied in the context of classification of disease types or the separation of individuals.

The rest of this paper is organized as follows: In Section 2, we review the relevant material from functional regression and in Section 3 we develop the elastic fPCR model. In Section 4 we extend the elastic fPCR to the logistic and multinomial logistic case. In Sections 5 and 6, we report the results of applying the proposed approach to a simulated data set and seven real data sets from various application domains. Finally, we close with a brief summary and some ideas for future work in Section 7.

2 | FUNCTIONAL PRINCIPAL COMPONENT REGRESSION

We start with a more common functional regression model, and then develop an “elastic” principal component version that accounts for phase variability of functional data. Without loss of generality we assume the time interval of interest to be $[0, 1]$. Let f be a real-valued function on $[0, 1]$; from a theoretical perspective we restrict to functions that are absolutely continuous on $[0, 1]$ and we let \mathcal{F} denote the set of all such functions. In practice, since observed data are discrete, this assumption is not a restriction.

2.1 | Functional Principal Component Regression Model

Let $\{f_i\}$ denote observations of a predictor function variable and let $y_i \in \mathbb{R}$, be the corresponding response variable. The standard functional linear regression model for this set of observations is

$$y_i = \alpha + \langle f_i, \beta \rangle + \epsilon_i, \quad i = 1, \dots, n \quad (2.1)$$

where $\langle f_i, \beta \rangle = \int f_i(t) \beta(t) dt$ is the \mathbb{L}^2 functional inner product, α is the bias and $\beta(t)$ is the regression coefficient function. The model is usually determined by minimizing the sum of squared errors (SSE).

$$\{\alpha^*, \beta^*(t)\} = \arg \min_{\alpha, \beta(t)} \sum_{i=1}^n |y_i - \alpha - \int f_i(t) \beta(t) dt|^2. \quad (2.2)$$

One problem with this approach, is that for any finite n , it is possible to perfectly interpolate the responses if no restrictions were placed on $\beta(t)$. Specifically, since $\beta(t)$ is infinite dimensional, we have infinite degrees of freedom to form $\beta(t)$ in which we can make the SSE equal zero. Ramsay and Silverman [29] proposed two approaches with the first representing $\beta(t)$ using a p -dimensional basis in which p is hopefully large enough to capture all variations of $\beta(t)$. The second approach is adding a penalty term which shrinks the variability of $\beta(t)$ or smooths its response.

Functional principal component regression uses the principal components as the basis functions where the model is determined by minimizing

$$\{\alpha^*, \mathbf{b}^*\} = \arg \min_{\alpha, \mathbf{b}} \sum_{i=1}^n |y_i - \alpha - \sum_{j=1}^{n_o} \langle f_i, \xi_j \rangle b_j|^2, \quad (2.3)$$

where $p = n_o$ principal components are used, $\xi(t)$ is the corresponding eigenfunction, and $\mathbf{b} = [b_1, \dots, b_j]$. It should be noted that $f_i(t)$ here is mean centered.

3 | ELASTIC FPCR MODEL

In order to properly account for the variability we can use the vertical fPCA and horizontal fPCA presented in Tucker et al. [37]. These fPCA methods account for the variability, by first separating the phase and amplitude and then performing the fPCA on the spaces separately. Using these methods one can construct a regression on the amplitude space using the square-root slope function (SRSF), q or specifically the aligned SRSF \tilde{q} , and the phase space using the warping functions, γ , the motivation for the using of SRSF will be explained later. A third option is to use the method developed by Lee and Jung [23] which is an extension of the method developed by Tucker et al. Lee and Jung proposes a joint fPCA which generates a function g^C which concatenates the function (f) and the warping function. We propose a slight modification of this work to use the SRSF, due to its better theoretical properties. The concatenated function actually works on a simplified geometry because the warping function is first transformed to the unit Hilbert Sphere and then to its tangent space. This simplification and SRSF modification allows the use of a metric that is a proper distance both in the vertical and horizontal case. By using the joint fPCA the regression model can be performed on the amplitude and phase simultaneously. Table 1 presents the three domains and where the regression is performed.

TABLE 1 Functional Principal Component Regression Domains.

| | Vertical fPCA | Horizontal fPCA | Joint fPCA |
|----------------|---------------|-----------------|-----------------------------|
| Representation | \tilde{q} | γ | $g^C = [\tilde{q} \ Cv(t)]$ |
| Variability | Amplitude | Phase | Amplitude + Phase |
| Metric | Fisher-Rao | Fisher-Rao | Fisher-Rao |

3.1 | Elastic fPCA

We begin by giving a short review of the vertical and horizontal fPCA of Tucker et al. [37] and the joint phase-amplitude fPCA method of Lee and Jung [23], with a slight modification which will be described clearly in later sections. These

methods are based on the functional data analysis approach outlined in Srivastava et al. [36], Kurtek et al. [21], and Tucker et al. [37]; see those references for more details on this background material.

Let Γ be the set of orientation-preserving diffeomorphisms of the unit interval $[0, 1]$: $\Gamma = \{\gamma : [0, 1] \rightarrow [0, 1] \mid \gamma(0) = 0, \gamma(1) = 1, \gamma \text{ is a diffeomorphism}\}$. Elements of Γ play the role of warping functions. For any $f \in \mathcal{F}$ and $\gamma \in \Gamma$, the composition $f \circ \gamma$ denotes the time-warping of f by γ . With the composition operation, the set Γ is a Lie group with the identity element $\gamma_{id}(t) = t$. This is an important observation since the group structure of Γ is seldom utilized in past papers on functional data analysis.

As described in Tucker et al. [37], there are two metrics to measure the amplitude and phase variability of functions. These metrics are proper distances, one on the quotient space \mathcal{F}/Γ (i.e., amplitude) and the other on the group Γ (i.e., phase). The amplitude or y -distance for any two functions $f_1, f_2 \in \mathcal{F}$ is defined to be

$$d_a(f_1, f_2) = \inf_{\gamma \in \Gamma} \|q_1 - (q_2 \circ \gamma)\sqrt{\dot{\gamma}}\|, \quad (3.1)$$

where $q(t) = \text{sign}(\dot{f}(t))\sqrt{|\dot{f}(t)|}$ is known as the square-root slope function (SRSF) (\dot{f} represents the derivative of f). The optimization problem in Equation 3.1 is most commonly solved using a Dynamic Programming algorithm; see Robinson [31] for a detailed description. If f is absolutely continuous, then $q \in \mathbb{L}^2([0, 1], \mathbb{R})$ (Robinson [31]), henceforth denoted by \mathbb{L}^2 . For the properties of the SRSF and the reason for its use in this setting, we refer the reader to Srivastava et al. [35], Marron et al. [25] and Lahiri et al. [22]. Moreover, it can be shown that for any $\gamma_1, \gamma_2 \in \Gamma$, we have $d_a(f_1 \circ \gamma_1, f_2 \circ \gamma_2) = d_a(f_1, f_2)$, i.e., the amplitude distance is invariant to function warping.

3.2 | Simplifying Geometry of Γ

The space of warping functions, Γ , is an infinite-dimensional nonlinear manifold, and therefore cannot be treated as a standard Hilbert space. To overcome this problem, we will use tools from differential geometry to perform statistical analyses and to model the warping functions. The following framework was previously used in various settings including; (1) modeling re-parameterizations of curves (Srivastava and Jermyn [32]), (2) putting prior distributions on warping functions (Kurtek [19] and Lu et al. [24]), (3) studying execution rates of human activities in videos (Veeraraghavan et al. [39]), and many others. It is also very closely related to the square-root representation of probability density functions introduced by Bhattacharya [2], and later used for various statistical tasks (see e.g., Kurtek and Bharath [20]).

We represent an element $\gamma \in \Gamma$ by the square-root of its derivative $\psi = \sqrt{\dot{\gamma}}$. Note that this is the same as the SRSF defined earlier, and takes this form since $\dot{\gamma} > 0$. The identity γ_{id} maps to a constant function with value $\psi_{id}(t) = 1$. Since $\gamma(0) = 0$, the mapping from γ to ψ is invertible and one can reconstruct γ from ψ using $\gamma(t) = \int_0^t \psi(s)^2 ds$. An important advantage of this transformation is that since $\|\psi\|^2 = \int_0^1 \psi(t)^2 dt = \int_0^1 \dot{\gamma}(t) dt = \gamma(1) - \gamma(0) = 1$, the set of all such ψ s is

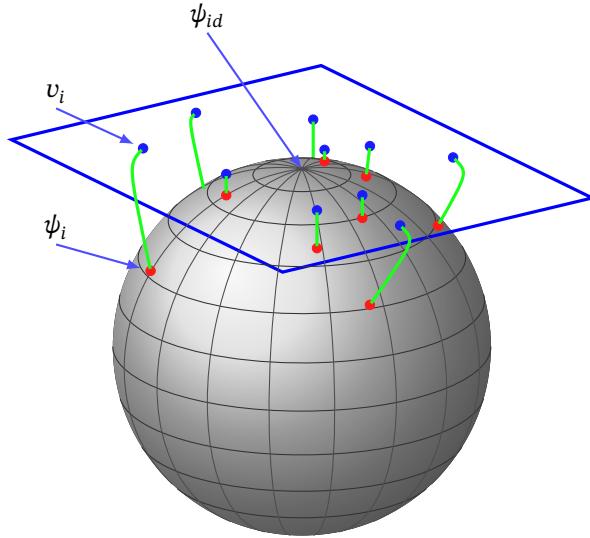


FIGURE 1 Depiction of the SRSF space of warping functions as a sphere and a tangent space at the identity element ψ_{id} .

the positive orthant of the Hilbert sphere $\Psi = \mathbb{S}_\infty^+$ (i.e., a unit sphere in the Hilbert space \mathbb{L}^2). In other words, the square-root representation simplifies the complicated geometry of Γ to a unit sphere. The distance between any two warping functions, i.e., the phase distance, is exactly the arc-length between their corresponding SRSFs on the unit sphere \mathbb{S}_∞ :

$$d_p(\gamma_1, \gamma_2) = d_\psi(\psi_1, \psi_2) \equiv \cos^{-1} \left(\int_0^1 \psi_1(t) \psi_2(t) dt \right).$$

Figure 1 shows an illustration of the SRSF space of warping functions as a unit sphere.

3.3 | Mapping to the Tangent Space at Identity Element

While the geometry of $\Psi \subset \mathbb{S}_\infty$ is more tractable, it is still a nonlinear manifold and computing standard statistics remains difficult. Instead, we use a tangent (vector) space at a certain fixed point for further analysis. The tangent space at any point $\psi \in \Psi$ is given by: $T_\psi(\Psi) = \{v \in \mathbb{L}^2 \mid \int_0^1 v(t) \psi(t) dt = 0\}$. To map between the representation space Ψ and tangent spaces, one requires the exponential and inverse-exponential mappings. The exponential map at a point $\psi \in \Psi$ denoted by $\exp_\psi : T_\psi(\Psi) \mapsto \Psi$, is defined as

$$\exp_\psi(v) = \cos(\|v\|)\psi + \sin(\|v\|) \frac{v}{\|v\|}, \quad (3.2)$$

where $v \in T_\psi(\Psi)$. Thus, $\exp_\psi(v)$ maps points from the tangent space at ψ to the representation space Ψ . Similarly, the inverse-exponential map, denoted by $\exp_\psi^{-1} : \Psi \mapsto T_\psi(\Psi)$, is defined as

$$\exp_\psi^{-1}(\psi_1) = \frac{\theta}{\sin(\theta)}(\psi_1 - \cos(\theta)\psi), \quad (3.3)$$

where $\theta = d_p(\gamma_1, \gamma)$. This mapping takes points from the representation space to the tangent space at ψ .

The tangent space representation v is sometimes referred to as a *shooting vector*, as depicted in Figure 1. The remaining question is which tangent space should be used to represent the warping functions. A sensible point on Ψ to define the tangent space is at the sample Karcher mean $\hat{\mu}_\psi$ (corresponding to $\hat{\mu}_\gamma$) of the given warping functions or the identity element ψ_{id} . For details on the definition of the sample Karcher mean and how to compute it, please refer to Tucker et al. [37].

3.4 | Vertical Functional Principal Components

Let f_1, \dots, f_n be a given set of functions, and q_1, \dots, q_n be the corresponding SRSFs, μ_q be their Karcher Mean, and let \tilde{q}_i s be the corresponding aligned SRSFs using Algorithm 1 from Tucker et al. [37]. In performing vertical fPCA, one also needs to include the variability associated with the initial values, i.e., $\{f_i(0)\}$, of the given functions. Since representing functions by their SRSFs ignores vertical translation, this variable is treated separately. That is, a functional variable f is analyzed using the pair $(q, f(0))$ rather than just q . This way, the mapping from the function space \mathcal{F} to $\mathbb{L}^2 \times \mathbb{R}$ is a bijection. In practice, where q is represented using a finite partition of $[0, 1]$, say with cardinality T , the combined vector $h_i = [q_i \ f_i(0)]$ simply has dimension $(T + 1)$ for fPCA. We can define a sample covariance operator for the aligned combined vector $\tilde{h} = [\tilde{q}_1 \ f_i(0)]$ as

$$K_h = \frac{1}{n-1} \sum_{i=1}^n (\tilde{h}_i - \mu_h)(\tilde{h}_i - \mu_h)^\top \in \mathbb{R}^{(T+1) \times (T+1)},$$

where $\mu_h = [\mu_q \ \bar{f}(0)]$. Taking the SVD, $K_h = U_h \Sigma_h V_h^\top$, we can calculate the directions of principle variability in the given SRSFs using the first $p \leq n$ columns of U_h and can then converted back to the function space \mathcal{F} , via integration, for finding the principal components of the original functional data. Moreover, we can calculate the observed principal coefficients as $\langle \tilde{h}_i, U_{h,j} \rangle$.

One can then use this framework to visualize the vertical principal-geodesic paths. The basic idea is to compute a few points along geodesic path $\tau \mapsto \mu_h + \tau\sqrt{\Sigma_{h,jj}}U_{h,j}$ for $\tau \in \mathbb{R}$ in \mathbb{L}^2 , where $\Sigma_{h,jj}$ and $U_{h,j}$ are the j^{th} singular value and column, respectively. Then, obtain principle paths in the function space \mathcal{F} by integration.

3.5 | Horizontal Functional Principal Components

To perform horizontal fPCA we will use the tangent space at μ_ψ Srivastava and Klassen [34] to perform analysis, where μ_ψ is the mean of the transformed warping functions. Algorithm 2 from Tucker et al. [37] can be used to calculate this mean. In this tangent space we can define a sample covariance function:

$$K_\psi = \frac{1}{n-1} \sum_{i=1}^n v_i v_i^\top \in \mathbb{R}^{T \times T}.$$

The singular value decomposition (SVD) of $K_\psi = U_\psi \Sigma_\psi V_\psi^\top$ provides the estimated principal components of $\{\psi_i\}$: the principal directions $U_{\psi,j}$ and the observed principal coefficients $\langle v_i, U_{\psi,j} \rangle$. This analysis on \mathbb{S}_∞ is similar to the ideas presented in [33] although one can also use the idea of principal nested sphere for this analysis [16]. The columns of U_ψ can then be used to visualize the principal geodesic paths.

3.6 | Joint Functional Principal Components

To model the association between the amplitude of a function and its phase, Lee and Jung [23] use a concatenated function g^C on the extended domain $[0, 2]$ (for some $C > 0$). The domain is extended as a concatenated function $g(t)$ is created as a combination of the original function and the warping function. Since the domain is $[0, 1]$ for both the function and warping function, we treat $g(t)$ on the extended domain as is defined below:

$$g^C(t) = \begin{cases} f^*(t), & t \in [0, 1) \\ Cv(t-1), & t \in [1, 2] \end{cases} \quad (3.4)$$

where f^* only contains the function's amplitude (i.e., after alignment via SRSFs). The superscript C is used to denote the dependence of the principal coefficients on the scaling factor. Furthermore, Lee and Jung [23] assume that $g^C \in \mathbb{L}^2([0, 2], \mathbb{R})$. The parameter C is introduced to adjust for the scaling imbalance between f^* and v . In the current work, we make a modification to the method of Lee and Jung [23]. In particular, it seems more appropriate to construct the function g^C using the SRSF q^* of the aligned function f^* , since q^* is guaranteed to be an element of \mathbb{L}^2 . Explicitly, we will use \tilde{h} as it includes the vertical translation value $f_i(0)$. Thus, with a slight abuse in notation, we proceed with the following joint representation of amplitude and phase:

$$g^C(t) = \begin{cases} h^*(t), & t \in [0, 1) \\ Cv(t-1), & t \in [1, 2] \end{cases} \quad (3.5)$$

where C is again used to adjust for the scaling imbalance between h^* and v .

Henceforth, we assume that h^* is sampled using $T+1$ points and v is sampled using T points, making the dimensionality of $g^C \in \mathbb{R}^{2T+1}$. Then, given a sample of amplitude-phase functions $\{g_1^C, \dots, g_n^C\}$, and their sample mean $\hat{\mu}_g^C = [\hat{\mu}_{q^*}, \hat{\mu}_v^C]$, we

can compute the sample covariance matrix as

$$K_g^C = \frac{1}{n-1} \sum_{i=1}^n (g_i^C - \hat{\mu}_g^C)(g_i^C - \hat{\mu}_g^C)^\top \in \mathbb{R}^{(2T+1) \times (2T+1)}. \quad (3.6)$$

It should be noted that the mean $\hat{\mu}_g^C$ is always zero. Taking the Singular Value Decomposition, $K_g^C = U_g^C \Sigma_g^C (V_g^C)^\top$, we calculate the joint principal directions of variability in the given amplitude-phase functions using the first $p \leq n$ columns of U_g^C . These can be converted back to the original representation spaces (\mathcal{F} and γ) using the mappings defined earlier. Moreover, one can calculate the observed principal coefficients as $\langle g_i^C, U_{g,j}^C \rangle$, for the i^{th} function with the j^{th} principal component.

This framework can be used to visualize the principal modes of variability. First, the matrix U_g^C is partitioned into the pair $(U_{q^*}^C, U_v^C)$. Then, the amplitude and phase paths within one standard deviation of the mean are computed as

$$q_{\tau,j}^{*C} = \hat{\mu}_{q^*} + \tau \sqrt{\Sigma_{g,jj}^C} U_{q^*,j}^C \quad (3.7)$$

$$v_{\tau,j}^C = \tau \frac{\sqrt{\Sigma_{g,jj}^C}}{C} U_{v,j}^C, \quad (3.8)$$

where $\tau \in \mathbb{R}$, $\Sigma_{g,jj}$ and U_j^C are the j^{th} principal component variance and direction of variability, respectively. Then, one can obtain a joint amplitude-phase principal path by composing $f_{\tau,j}^{*C}$ (this is the function corresponding to SRSF $q_{\tau,j}^{*C}$) with $\gamma_{\tau,j}^C$ (this is the warping function corresponding to $v_{\tau,j}^C$).

The results of the above procedure clearly differ for variations of C . For example, using small values of C , the first few principal directions of variability will capture more amplitude variation, while for large values of C , the leading directions reflect more phase variation. Lee and Jung [23] present a data-driven method for estimating C for a given sample of functions which minimizes the reconstruction error. We use this approach in the current work to determine an appropriate value of C and a sensitivity study will be provided later on in the paper determining its impact on regression performance. Other metrics in cross-validation can be used such as prediction performance or selecting C based on which variability the user wants to emphasize in the data.

3.7 | Elastic Functional Principal Component Regression Model

The regression model then is

$$y = \alpha + \sum_{j=1}^{n_o} \langle x_i, \xi_j \rangle b_j \quad (3.9)$$

and can be found by solving

$$\{\alpha^*, \mathbf{b}^*\} = \arg \min_{\alpha, \mathbf{b}} \left[\sum_{i=1}^n |y_i - \alpha - \sum_{j=1}^{n_o} \langle x_i, \xi_j \rangle b_j|^2 \right], \quad (3.10)$$

where the appropriate function is substituted in for x_i and appropriate eigenfunction for ξ_j from Table 1 depending on which fPCA is used for the regression.

The solution for the optimal α^* and \mathbf{b}^* is found using ordinary least squares. Define $Z = [\mathbf{1} \ \Theta]$, where $\mathbf{1}$ is a vector of ones and $\Theta \in \mathbb{R}^{N \times n_o}$ is the matrix containing the principal coefficients for the N samples for n_o principal components and $\mathbf{y} = [y_1, \dots, y_n]^\top$. Then the solution for α^* and \mathbf{b}^* is

$$[\alpha^*, \mathbf{b}^*]^\top = (Z^\top Z)^{-1} Z^\top \mathbf{y}.$$

We perform fPCA in the SRSF space under the \mathbb{L}^2 norm for the amplitude and performing fPCA in the tangent space of warping functions after the square-root transform. Since these fPCA's are identical to the fPCA one does in a Hilbert space under the \mathbb{L}^2 norm, the standard asymptotic analysis directly applies (Dauxois et al. [9], Cardot et al. [5], Reiss and Ogden [30]).

4 | ELASTIC GENERALIZED FPCR MODEL

We now develop the generalized version of the Elastic fPCR model and then apply it to the logistic and multinomial logistic case. This model is an extension of the linear regression model with a given appropriate link function. The elastic generalized regression model is defined as

$$y = h\left(\sum_{j=1}^{n_o} \langle x_i, \xi_j \rangle b_j\right),$$

where $h(\cdot)$ is the link function. As with the linear model one substitutes the appropriate eigenfunction for ξ_j from Table 1 depending on which fPCA is used for the regression. The choice of the link function determines the type of model. In this paper we study the logistic functional regression model and will discuss these more fully next.

4.1 | Logistic Functional Regression Model

Let $\{f_i\}$ denote observations of a predictor function variable and let $y_i \in \{-1, 1\}$, for $i = 1, \dots, n$ be the corresponding binary response variable. We define the probability of the function f_i being in class 1 ($y_i = 1$) as

$$P(y_i = 1|f_i) = \frac{1}{1 + \exp(-[\alpha + \int_0^1 f_i(t)\beta(t) dt])}.$$

This is nothing but the logistic link function $\phi(t) = 1/(1 + \exp(-t))$ applied to the conditional mean in a linear regression model: $\alpha + \int_0^1 f_i(t)\beta(t)dt$ (James [14]). Using this relation, and the fact that $P(y = -1|f_i) = 1 - P(y = 1|f_i)$, we can

express the data likelihood as:

$$\pi(\{y_i\}|\{f_i\}, \alpha, \beta) = \prod_{i=1}^n \frac{1}{1 + \exp(-y_i [\alpha + \int_0^1 f_i(t)\beta(t) dt])}. \quad (4.1)$$

Assuming we observe a sequence of i.i.d. pairs $\{f_i(t), y_i\}, i = 1, \dots, n$, the model is identified by maximizing the log-likelihood according to,

$$\{\alpha^*, \beta^*\} = \arg \max_{\alpha, \beta(t)} (\log \pi(\{y_i\}|\{f_i\}, \alpha, \beta)).$$

This optimization has been the main focus of the current literature (see e.g., Ramsay and Silverman [29], Cardot et al. [6], Hall and Horowitz [13]).

4.2 | Elastic Logistic fPCR

Now consider the situation where functional predictors can include phase variability as well as the amplitude variability.

We will use the Elastic fPCR method with the logistic link function

$$\pi(\{y_i\}|\{f_i\}, \alpha, \mathbf{b}) = \prod_{i=1}^n \frac{1}{1 + \exp(-y_i [\alpha + \sum_{j=1}^{n_o} \langle x_i, \xi_j \rangle b_j])}$$

where the appropriate fPCA model is used for the proper variability.

The search for α and \mathbf{b} is performed by maximizing the log-likelihood. We can combine all the parameters – intercept α and coefficients b_i s – in a vector form $\boldsymbol{\theta} = [\alpha, b_1, \dots, b_{n_o}]^\top$ and let $\mathbf{z}_i = [1, \langle x_i, \xi_1 \rangle, \dots, \langle x_i, \xi_{n_o} \rangle]^\top$. The optimal parameter vector is then defined according to:

$$\boldsymbol{\theta}^* = \arg \max_{\boldsymbol{\theta} \in \mathbb{R}^{p+1}} \sum_{i=1}^n \log(\phi(y_i \boldsymbol{\theta}^\top \mathbf{z}_i)), \quad (4.2)$$

where $\phi(t) = 1/(1+\exp(-t))$. There is no known analytical solution to this optimization problem. Since the objective function is concave, we can use a numerical method such as Conjugate Gradient or the Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm (Mordecai [26]) for finding $\boldsymbol{\theta}^*$. To use these algorithms we need the gradient of the log-likelihood ($\log \pi(\{y_i\}|\{f_i\}, \alpha, \beta)$), which is given by:

$$\nabla L(\boldsymbol{\theta}) = \sum_{i=1}^n -y_i \mathbf{z}_i (\phi(y_i \boldsymbol{\theta}^\top \mathbf{z}_i) - 1).$$

In this paper we will use the Limited Memory BFGS (L-BFGS) algorithm due to its storage efficiency in dealing with a large number of predictors (Byrd et al. [3]). Similar to ideas discussed in [10], one can also seek a sparse representation by including a \mathbb{L}_1 or \mathbb{L}_2 penalty on \mathbf{b} in Eqn. 4.2.

4.3 | Extension to Elastic Multinomial Logistic fPCR

We can extend the elastic logistic fPCR to the case of multinomial response, i.e. y_i has more than two classes. In this case, we have observations $\{(f_i(t), y_i)\}$ and the response variable can take on m categories, $y_i \in \{1, \dots, m\}$, for $i = 1, \dots, n$. For simplification, we abuse the notation by coding the response variable y as a m -dimensional vector with a 1 in the k th component when $y = k$ and zero, otherwise. Next, we define the probability of the function f being in class k as

$$P(y^{(k)} = 1 | \{\alpha^{(j)}\}, \{\mathbf{b}^{(j)}\}, f) = \frac{\exp(\alpha^{(k)} + \sum_{j=1}^{n_o} \langle x, \xi_j \rangle b_j^{(k)})}{1 + \sum_{l=1}^{m-1} \exp(\alpha^{(l)} + \sum_{j=1}^{n_o} \langle x, \xi_j \rangle b_j^{(l)})}.$$

We assume $\alpha^{(m)} = 0$ and $\mathbf{b}^{(m)} = \mathbf{0}$ without loss of generality.

Using the above probability and the multinomial definition of the problem, we can express the log-likelihood of observations $\{(x_i(t), y_i)\}$ as

$$L_m(\{\alpha^{(i)}\}, \{\mathbf{b}^{(i)}\}) = \sum_{i=1}^n \left[\sum_{k=1}^{m-1} y_i^{(k)} \left[\alpha^{(k)} + \sum_{j=1}^{n_o} \langle x_i, \xi_j \rangle b_j^{(k)} \right] - \log \left(1 + \sum_{l=1}^{m-1} \exp \left(\alpha^{(l)} + \sum_{j=1}^{n_o} \langle x_i, \xi_j \rangle b_j^{(l)} \right) \right) \right],$$

where once again the appropriate fPCA model is used for dimension reduction and modeling.

The optimal $\{\alpha^{*(i)}\}$ and $\{\mathbf{b}^{*(i)}\}$ can be found as earlier by maximizing the log-likelihood. We can re-express the maximization of the log-likelihood as:

$$\theta^* = \arg \max_{\theta} \sum_{i=1}^n \left[\sum_{j=1}^{m-1} y_i^{(j)} \theta^{(j)\top} \mathbf{z}_i - \log \left(1 + \sum_{j=1}^{m-1} \exp(\theta^{(j)\top} \mathbf{z}_i) \right) \right], \quad (4.3)$$

where $\theta = [\alpha^{(k)}, b_1^{(k)}, \dots, b_{n_o}^{(k)}]^\top$ and $\mathbf{z}_i = [1, \langle x_i, \xi_1 \rangle, \dots, \langle x_i, \xi_{n_o} \rangle]^\top$. There is no direct analytical solution to this optimization and it has to be solved numerically. Since, the function is concave we will use the L-BFGS algorithm to find the solution numerically. To use this algorithm we need the gradient of the log-likelihood, which is given by:

$$\frac{\partial L_m(\theta)}{\partial \theta^{(k)}} = \sum_{i=1}^n \left[y_i^{(k)} \mathbf{z}_i - \frac{1}{1 + \sum_{j=1}^{m-1} \exp(\theta^{(j)\top} \mathbf{z}_i)} \exp(\theta^{(k)\top} \mathbf{z}_i) \mathbf{z}_i \right].$$

We then can find the optimal $\{\alpha^{*(j)}\}$ and $\{\mathbf{b}^{*(j)}\}$ using L-BFGS.

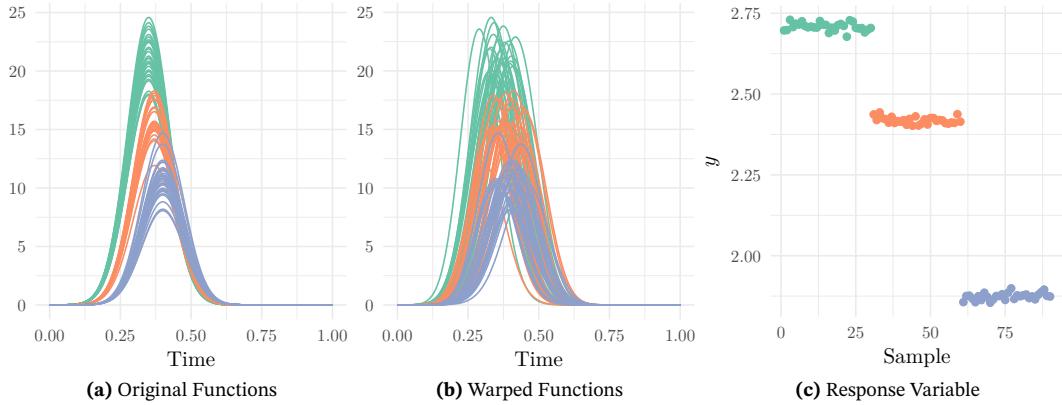


FIGURE 2 Simulated regression data with phase and amplitude variability.

5 | SIMULATION RESULTS

5.1 | Elastic Functional Principal Component Regression

To illustrate the proposed elastic functional regression method we applied the model to a simulated data constructed using

$$f_i(t) = a_i \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(t - \mu_j)^2}{2\sigma^2}\right),$$

where $a_i \sim \mathcal{N}(d_j, 0.05)$. The means were chosen according to three models: 1) Combined Amplitude & Phase Variability ($\mu_j \in [0.35, 0.37, 0.40]$ and $d_j \in [4, 3, 2]$), 2. Amplitude Variability only ($\mu_j \in [0.35, 0.35, 0.35]$ and $d_j \in [4, 3.7, 4]$), and 3) Phase Variability only ($\mu_j \in [0.35, 0.40, 0.50]$ and $d_j \in [4, 4, 4]$). A total of 20 functions were generated for each case and $\sigma = 0.075$. The generated functions are shown in Fig. 2 (a), 3 (a), and 4 (a), for cases 1, 2, and 3, respectively. The functions are then randomly warped to generate the warped data, $\{f_i\}$ as shown in Fig. 2 (b), 3 (b), and 4 (b), respectively. The response variable y_i was generated according to Eqn. 2.1 with $\alpha = 0$, $\beta(t) = 0.5 \sin(2\pi t) + 0.9 \cos(2\pi t)$, and is shown in Fig. 2 (c), 3 (c), and 4 (c).

TABLE 2 Calculated SSE values using four different fPCR methods for 3 different types of variability.

| | Elastic Joint | Elastic Vertical | Elastic Horizontal | Standard |
|------------|------------------------|------------------|------------------------|-----------------|
| Combined | 0.0875 (0.0237) | 0.3498 (0.1248) | 0.0838 (0.0300) | 0.5075 (0.1483) |
| Vertical | 0.2345 (0.0905) | 0.2390 (0.0869) | 2.0782 (1.9058) | 0.3173 (0.0393) |
| Horizontal | 0.1474 (0.0998) | 9.1292 (5.3023) | 0.2155 (0.1339) | 1.8729 (1.1719) |

Table 2 provides the SSE for each of the three cases with the lowest SSE shown in bold computed using 5-fold cross-validation. In the table we present the mean of the SSE across the folds, along with the standard deviation. For the data with the combined variability, the joint fPCA in the elastic fPCR model is slightly out performed by the horizontal fPCA.

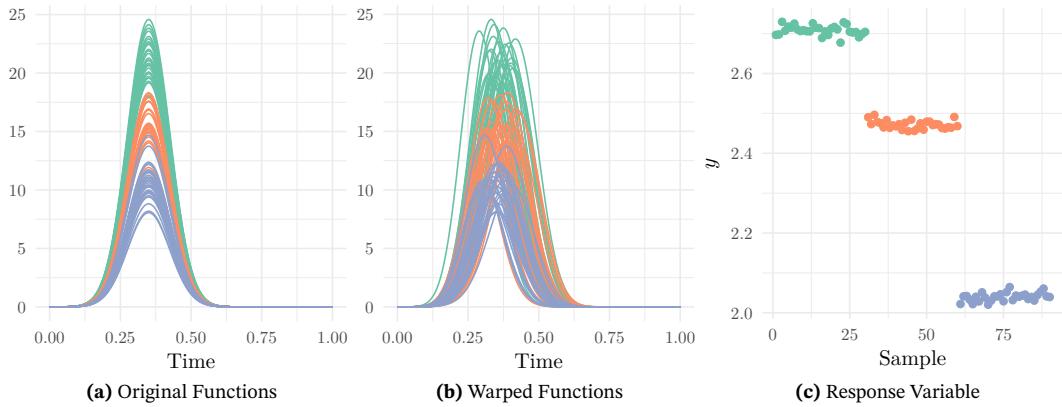


FIGURE 3 Simulated regression data with amplitude variability.

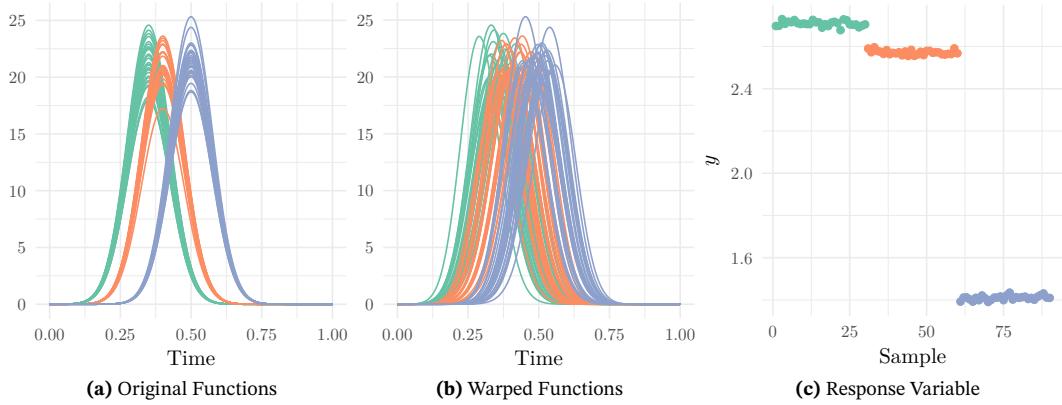


FIGURE 4 Simulated regression data with phase variability.

In the cases with both the vertical and horizontal variability, the joint elastic fPCA method performed the best with the corresponding vertical or horizontal fPCA method being very close. This is somewhat expected as the joint fPCA method is able to capture both types of variability. We compared the results from the elastic method to those using standard fPCR found in the literature on the warped data and those results are shown in the last column. In all cases the elastic method outperforms the standard fPCR method presented by Reiss and Ogden [30].

For the use of the joint fPCA in the elastic fPCR model, it is necessary to choose the parameter C to correct for scaling imbalance and for minimizing the reconstruction error or via cross-validation. The value of C can have a significant effect on the regression model performance and we have looked at its effect on the SSE for the three types of variability presented in Figs. 2 - 4 . For each of the three cases, we varied C from 1 to 120 and collected the results. The lowest SSE for each variability type is different, and for the combined variability being the largest value. This makes sense because when the variability is combined a larger value is required to handle the scaling imbalance. The SSE also begins to increase for the vertical variability when C gets large, as all the weight is on the warping functions and they are not good predictors in this

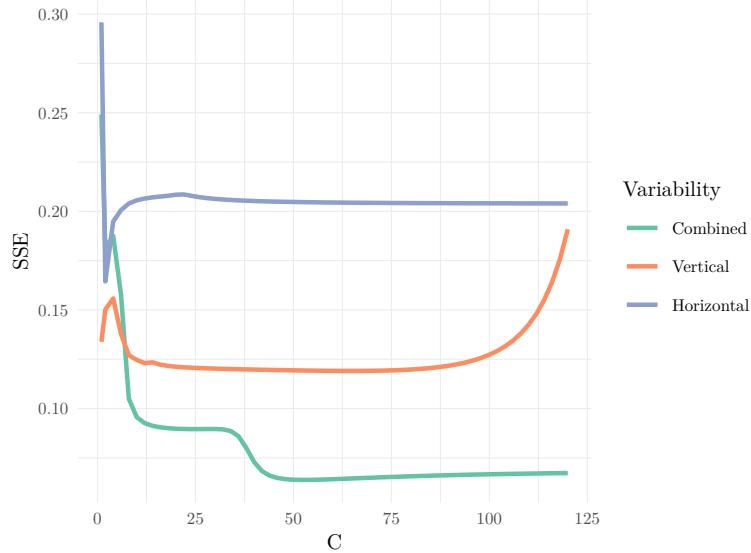


FIGURE 5 Sum of Squared errors versus C for the joint fPCA in the elastic fPCR model

case. Additionally, for the horizontal variability, the lowest SSE is for a low value of C and for the vertical variability the lowest SSE is found for a higher value of C .

5.2 | Elastic Logistic fPCR

To illustrate the elastic functional logistic regression method, we evaluated the model on a similar simulated data used in the previous section. The means were chosen according to three models: 1) Combined Amplitude & Phase Variability ($\mu_j \in [0.35, 0.37]$ and $d_j \in [4, 3]$), 2. Amplitude Variability only ($\mu_j \in [0.35, 0.35]$ and $d_j \in [4, 3.7]$), and 3) Phase Variability only ($\mu_j \in [0.35, 0.40]$ and $d_j \in [4, 4]$). A total of 20 functions were generated for each case using $\sigma = 0.075$. The functions were then randomly warped to generate the warped data, $\{f_i\}$ and the label was 1 for the first case and -1 for the second case.

Table 3 provides the probability of classification for each of the three cases. The analysis was performed using 5-fold cross-validation. In the table we present the mean probability of classification across the folds, along with the standard deviation. For the data with the combined variability the joint fPCA in the elastic logistic fPCR model performed the best. For the cases with both the vertical and horizontal variability, the corresponding elastic fPCA methods performed well with the joint fPCA method having the best performance. We compared the results to the standard logistic fPCR, where the logistic link function is applied to the method of Reiss and Ogden [30]. The results are shown in the last column of Table 3. In all cases the elastic method outperforms the standard logistic fPCR method.

TABLE 3 Calculated probability of classification values using four different fPCR methods for 3 different types of variability for logistic regression.

| | Elastic Joint | Elastic Vertical | Elastic Horizontal | Standard |
|------------|------------------------|------------------|--------------------|-----------------|
| Combined | 0.9750 (0.0342) | 0.9375 (0.0765) | 0.9750 (0.0342) | 0.9625 (0.0342) |
| Vertical | 0.9250 (0.0280) | 0.8500 (0.0948) | 0.6250 (0.0625) | 0.8750 (0.0442) |
| Horizontal | 0.9250 (0.0815) | 0.6000 (0.1630) | 0.8875 (0.1355) | 0.8750 (0.0442) |

5.3 | Elastic Multinomial Logistic fPCR

To illustrate our elastic functional regression method, we evaluated the model on a simulated data constructed in the elastic fPCR case. Each of the functions was randomly warped similar to the previous cases. The response variable y_i in this case was categorical with values $j \in \{1, 2, 3\}$ depending on the corresponding model.

Table 4 provides the probability of classification for each of the three cases using 5-fold cross-validation. For the data with the combined variability the horizontal and joint fPCA in the elastic multinomial logistic fPCR model performed the best. The cases with the vertical and horizontal variability the corresponding elastic fPCA methods performed well with the joint fPCA method having the best performance. We compared the results to using standard multinomial logistic fPCR found in the literature on the warped data and is shown in the last column of Table 4. In all cases the elastic method outperforms the standard multinomial logistic fPCR method.

TABLE 4 Calculated probability of classification values using four different fPCR methods for 3 different types of variability for multinomial logistic regression.

| | Elastic Joint | Elastic Vertical | Elastic Horizontal | Standard |
|------------|-----------------|------------------------|------------------------|-----------------|
| Combined | 0.9420 (0.0633) | 0.9246 (0.0355) | 0.9670 (0.0348) | 0.8663 (0.1597) |
| Vertical | 0.8589 (0.0801) | 0.8916 (0.0229) | 0.3822 (0.0710) | 0.8749 (0.0780) |
| Horizontal | 0.9176 (0.0480) | 0.3822 (0.1036) | 0.9666 (0.0187) | 0.9510 (0.0436) |

6 | APPLICATIONS TO REAL DATA

Here, we present the results on multiple real data sets for the three elastic regression models. For the elastic fPCR we use the Sonar data set presented in (Tucker et al. [38]) where we predict the volumes of two targets. We demonstrate the elastic logistic fPCR model on three sets. The data consists of physiological data, specifically, gait and electrocardiogram (ECG) measurements from various patients. Phase variability is naturally found in this type data, as during collection the signals always start and stop at the different times for each measurement. For example, when measuring a heart beat one cannot assure that the measurement starts on the same part of the heartbeat for each patient measured. For the elastic

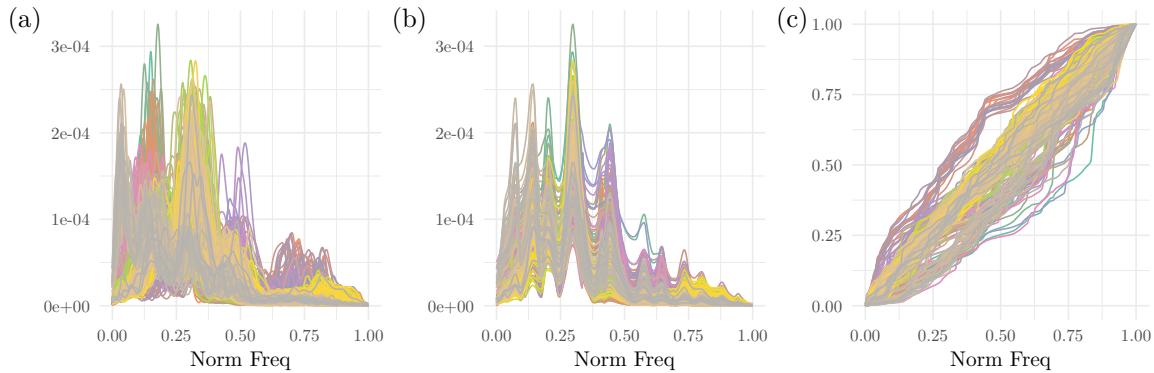


FIGURE 6 Alignment of the sonar dataset. (a) Original functions. (b) Aligned functions (amplitude). (c) Warping functions (phase).

multinomial logistic fPCR model we demonstrate on two sets that consist of physiologic data similar to those use to test the logistic regression method.

6.1 | Sonar Data

The data set used in these experiments was collected at the Naval Surface Warfare Center Panama City Division (NSWC PCD) test pond. For a description of the pond and a similar experimental setup the reader is referred to Kargl et al. [17]. The raw SONAR data was collected using a 1 - 30kHz LFM chirp and data was collected for a solid aluminum cylinder and a aluminum pipe. The aluminum cylinder is 2ft long with a 1ft diameter; while the pipe is 2ft long with an inner diameter of 1ft and 3/8 inch wall thickness. During the experiment the targets were placed with added uncertainty of their orientation. The acoustic signals were generated from the raw SONAR data to construct target strength as a function of frequency and aspect angle.

Figure 6 (a) presents the original functions for the acoustic color measurements at 0° aspect angle. There appears to be significant amplitude and phase variability between functional measurements due to experimental collection uncertainty. Not accounting for the phase variability can greatly affect summary statistics and follow-on statistical models. Figure 6 (b) and (c) show the aligned functions (amplitude) and warping functions (phase), respectively. Overall there is significant difference between the original functions and the aligned functions. With the large amount of phase variability, the frequency structure of the data was lost. As a result, cross-sectional methods without alignment will not capture this important difference in the functions.

TABLE 5 Calculated SSE values using four different fPCR methods for the sonar data set.

| | Elastic Joint | Elastic Vertical | Elastic Horizontal | Standard |
|-----|------------------------|------------------|--------------------|-----------------|
| SSE | 0.1210 (0.0472) | 0.1497 (0.1399) | 0.1597 (0.0485) | 0.1908 (0.1184) |

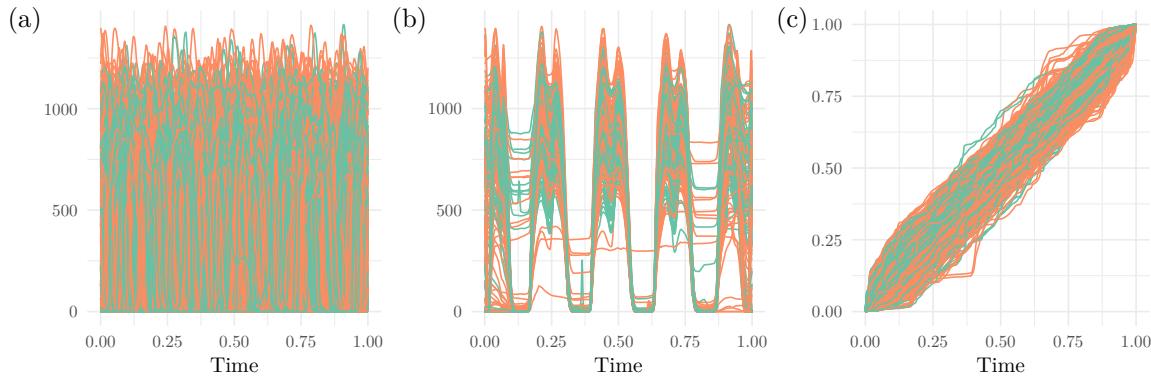


FIGURE 7 Alignment of the gait data dataset. (a) Original functions. (b) Aligned functions (amplitude). (c) Warping functions (phase).

Table 5 presents the sum of squared errors (SSE) calculated using 5-fold cross-validation. For this data set, we use ten principal components resulting in a ten-dimensional model for all four methods. In the table we present the mean of the SSE across the folds, along with the standard deviation. We compare the three elastic versions and standard functional principal component regression and the lowest SSE is shown in bold. The lowest SSE is the joint elastic fPCR method and all three elastic methods have lower SSE than the standard method in predicting the volume from the sonar data. With the high degree of phase and amplitude variability in the data the elastic method has higher prediction accuracy due to its ability to capture the variability properly.

6.2 | Gait Data

The Gait data is a collection of gait measurements for patients having Parkinson's disease and those not having Parkinson's disease and is from the gaitpdb data set on Physionet (Goldberger et al. [12]). This database contains gait measurements from 93 patients with idiopathic Parkinson's disease and 73 healthy patients. The gait was measured using vertical ground reaction force records of subjects as they walked at their usual, self-selected pace for approximately 2 minutes on level ground.

Figure 7 (a) presents the original functions for the gait data and are colored for the two different classes. There appears to be significant amplitude and phase variability between functional measurements due to experimental collection uncertainty and where one subject will start and stop their gate. Figure 7 (b) and (c) show the aligned functions (amplitude) and warping functions (phase), respectively. Overall there is significant difference between the original functions and the aligned functions. With the large amount of phase variability, the temporal structure of the gait will be lost in the analysis. As a result, cross-sectional methods without alignment will not capture this important difference in the functions, which can lead to lower predictive power of any developed models.

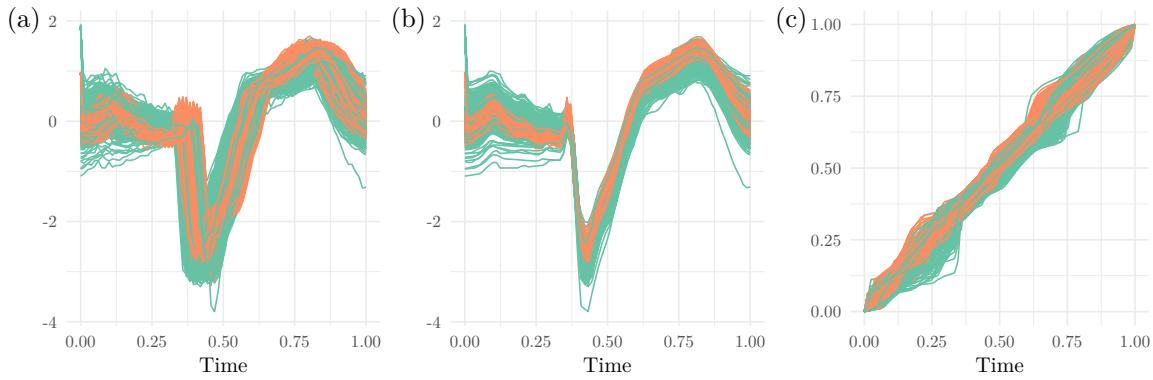


FIGURE 8 Alignment of the TwoLead ECG data dataset. (a) Original functions. (b) Aligned functions (amplitude). (c) Warping functions (phase).

The first row in Table 6 presents the calculated mean probability of classification (PC) using 5-fold cross-validation. For this data set, we use five principal components resulting in a five-dimensional model for all four methods. In the table we present the mean of the PC across the folds, along with the standard deviation. We compare the three elastic versions and standard logistic fPCR and the largest PC is shown in bold. The vertical elastic logistic fPCR method has the largest PC and all three elastic methods have a higher PC than the standard method in predicting if the subject has Parkinson's from the gait measurement. This suggests that a large portion of the information is contained in the amplitude variability.

6.3 | TwoLead ECG Data Set

The TwoLead ECG data set, is a collection of ECG measurements from the MIT-BIH Long-Term ECG Database available as well from Physionet. Which contains long term ECG measurements with beat annotations. Heartbeats were extracted that were annotated normal and abnormal.

Figure 8 (a) presents the original electrocardiogram measurements. Again, there appears to be significant phase variability between functional measurements due to timing uncertainty across collections. Figure 8 (b) and (c) show the aligned functions (amplitude) and warping functions (phase), respectively. Overall there is a noticeable alignment and better definition of the wave structure.

The second row in Table 6 presents the mean PC and standard deviation for this data set from 5-fold cross-validation. As in the previous analysis, we use five principal components resulting in a five-dimensional model for all four methods. The vertical elastic logistic fPCR method produced the largest PC and all three elastic methods have higher PC than the standard method.

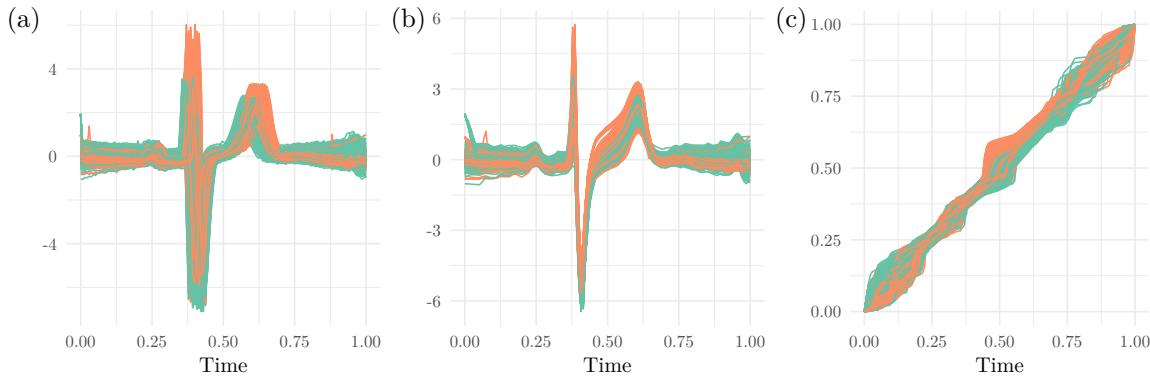


FIGURE 9 Alignment of the ECGFiveDays data dataset. (a) Original functions. (b) Aligned functions (amplitude). (c) Warping functions (phase).

6.4 | ECGFiveDays Data Set

The ECGFiveDays data set, is a collection of ECG measurements from a 67 year old male. There are two classes which are simply the data of the ECG measurements which are 5 days apart. The task is then to distinguish between the two days as the wandering baseline was not removed from the recordings. The data set is the ECGFiveDays from the UCR Time Series Classification Database (Keogh et al. [18]). Moreover, the previous two data sets can also be obtained from the UCR database under the names ECG200 and TwoLeadECG, respectively.

Figure 9 (a) presents the original electrocardiogram measurements from the ECGFiveDays set. Again, there appears to be some phase variability between functional measurements due to timing uncertainty across collections. Figure 9 (b) and (c) show the aligned functions (amplitude) and warping functions (phase), respectively. Overall there is a noticeable alignment and separation of the two classes in both the aligned functions and the warping functions.

TABLE 6 Calculated probability of correct classification using four different logistic fPCR methods for three different data sets.

| | Elastic Joint | Elastic Vertical | Elastic Horizontal | Standard |
|-------------|------------------------|------------------------|--------------------|-----------------|
| Gait | 0.6467 (0.0321) | 0.6900 (0.0465) | 0.6333 (0.0425) | 0.4300 (0.0923) |
| TwoLead ECG | 0.9113 (0.0125) | 0.9845 (0.0163) | 0.9156 (0.0146) | 0.8012 (0.0133) |
| ECGFiveDays | 0.9570 (0.0228) | 0.8902 (0.0462) | 0.8473 (0.0429) | 0.9061 (0.0297) |

The last row in Table 6 presents the calculated mean PC and standard deviation from 5-fold cross-validation. Again for this data set, we use five principal components resulting in a five-dimensional model for all four methods. For this data the joint elastic fPCR method produced the largest PC. This suggests that there is a combination of both phase and amplitude that contribute to correct classification, which is not captured by the standard method.

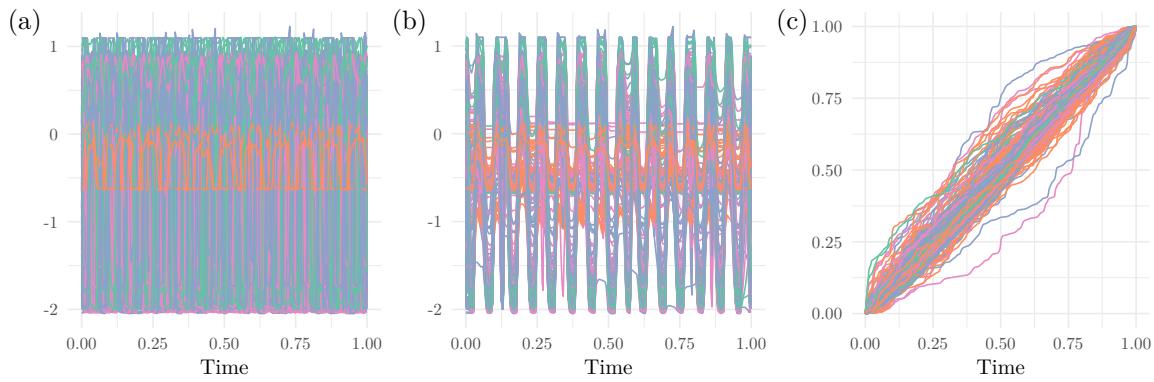


FIGURE 10 Alignment of the Gaitndd data dataset. (a) Original functions. (b) Aligned functions (amplitude). (c) Warping functions (phase).

6.5 | Gaitndd Data set

The Gaitndd data set is a collection of gait measurements for patients having Parkinson’s disease, Amyotrophic lateral sclerosis, Huntington’s disease, and healthy controls and is from the gaitndd data set on Physionet (Goldberger et al. [12]). This database contains gait measurements from 15, 20, 13, and 16 patients for the respective disease classes. The gait was measured using vertical ground reaction force records of subjects as they walked at their usual pace.

Figure 10 (a) presents the original gait measurements and are colored for the different classes. There is a large amount of phase and amplitude variability between the functional measurements. Figure 10 (b) and (c) show the aligned functions (amplitude) and warping functions (phase), respectively. Overall there is a large improvement in the structure after alignment some class definition can be noticed in the functions.

The first row in Table 7 presents the calculated mean PC and standard deviation. For this data set, we use ten principal components resulting in a ten-dimensional model for all four methods. The vertical elastic fPCR method and all three elastic methods have higher PC than the standard method. Given that the vertical has the largest PC, suggests there is a large amplitude component in how each disease affects the gait.

6.6 | CinC ECG Data Set

The last data set is a collection of ECG measurements from multiple torso-surface sites. There are measurements from 4 different people which are the 4 different classes. The data set is from the 2007 Physionet CinC challenge and is also found as the CinC data set from the UCR Time Series Classification Database (Keogh et al. [18]).

Figure 11 (a) presents the original ECG measurements and are colored for the different classes. There is a large phase and amplitude variability between the functional measurements. Figure 11 (b) and (c) show the aligned functions (amplitude) and warping functions (phase), respectively. Overall there is a large improvement in the structure after alignment

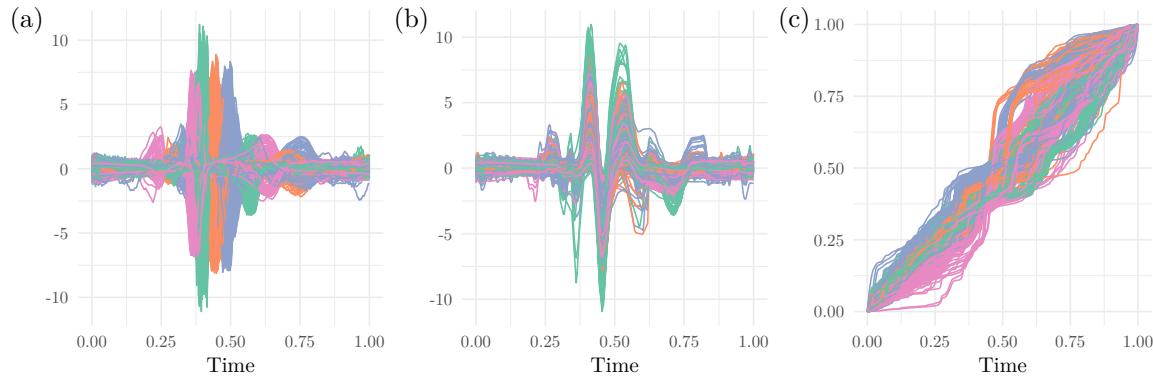


FIGURE 11 Alignment of the CinC ECG data dataset. (a) Original functions. (b) Aligned functions (amplitude). (c) Warping functions (phase).

and noticeable class separation in the warping functions. This suggests that the phase will have a large contribution to the classification performance.

TABLE 7 Calculated probability of correct classification using four different multinomial logistic fPCR methods for two different data sets.

| | Elastic Joint | Elastic Vertical | Elastic Horizontal | Standard |
|----------|-----------------|------------------------|------------------------|-----------------|
| Gaitndd | 0.3949 (0.0648) | 0.4888 (0.0326) | 0.3645 (0.0398) | 0.3123 (0.0413) |
| CinC ECG | 0.6785 (0.0403) | 0.6342 (0.0311) | 0.6954 (0.0452) | 0.3297 (0.0374) |

The last row in Table 7 presents the calculated mean PC and standard deviation using 5-fold cross-validation. For this data set, we use ten principal components resulting in a ten-dimensional model for all four methods. The largest PC is the horizontal elastic fPCR method and all three elastic methods have higher PC than the standard method. As was noted from the alignment, this suggests there is a large phase component which contributes to the separation of the classes. When accounting for this variability properly, the performance of correct classification is dramatically larger than just performing standard multinomial functional principal component regression.

7 | CONCLUSION AND FUTURE WORK

The statistical modeling and classification of functional data with phase variability is a challenging task. We have proposed a new functional principal component regression approach that addresses the problem of registering and modeling functions in one elastic-framework. We demonstrated three fPCA methods: 1) joint, 2) vertical, and 3) horizontal that can be used depending on the type of data encountered. This enabled the implementation of a regression model that is geometrically-motivated. We demonstrated the applicability of these models on a three different simulated examples that

contain different types of variability as well as seven real data examples with significant amplitude and phase variabilities. In all cases, we illustrated improvements in prediction power of the proposed models.

We have identified several directions for future work. First, we will explore the application of robust estimation methods applied to the elastic fPCA methods. Second, in many applications, the functional data of interest may be more complex than the simple univariate functions considered in this work; some examples include shapes of curves, surfaces, and images. These more complicated data objects often exhibit different sources of variability, which must be taken into account when computing regression models.

References

- [1] Bali, J., G. Boente, D. E. Tyler, and J. L. Wang, 2011: Robust functional principal components: A projection-pursuit approach. *Annals of Statistics*, **39**, no. 6, 2852–2882.
- [2] Bhattacharya, A., 1943: On a measure of divergence between two statistical populations defined by their probability distributions. *Bull. Calcutta Math. Soc.*, **35**, 99–109.
- [3] Byrd, R. H., P. Lu, J. Nocedal, and C. Zhu, 1995: A limited memory algorithm for bound constrained optimization. *SIAM J. Sci. Comput.*, **16**, no. 5, 1190–1208.
- [4] Cai, T. and P. Hall, 2006: Prediction in functional linear regression. *Annals of Statistics*, **34**, no. 5, 2159–2179.
- [5] Cardot, H., F. Ferraty, and P. Sarda, 1999: Functional linear model. *Statistics & Probability Letters*, **45**, 11–22.
- [6] —— 2003: Spline estimator for the functional linear model. *Statistica Sinica*, **13**, 571–591.
- [7] Carroll, R. J., D. Ruppert, L. A. Stefanski, and C. M. Crainiceanu, 2006: *Measurement Error in Nonlinear Models: A Modern Perspective*. 2nd ed., Chapman and Hall/CRC.
- [8] Cuevas, A., M. Febrero, and R. Fraiman, 2002: Linear functional regression; the case of fixed design and functional response. *The Canadian Journal of Statistics*, **30**, 285–300.
- [9] Dauxois, J., A. Pousse, and Y. Romain, 1982: Asymptotic theory for the principal component analysis of a vector random function: Some applications to statistical inference. *Journal of Multivariate Analysis*, **12**, 136–154.
- [10] Gertheiss, J., A. Maity, and A.-M. Staicu, 2013: Variable selection in generalized functional linear models. *Stat*, **2**, no. 1, 86–101.
- [11] Gervini, D., 2015: Curve registration. *Biometrika*, **102**, no. 1, 1–14.

- [12] Goldberger, A. L., L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C. Peng, and H. E. Stanley, 2000: Physiobank, physiotoolkit, and physionet: Components of a new research resource for complex physiologic signals. *Circulation*, **101**, no. 23, e215–e220.
URL <http://www.physionet.org>
- [13] Hall, P. and J. L. Horowitz, 2007: Methodology and convergence rates for functional linear regression. *Annals of Statistics*, **35**, no. 1, 70–91.
- [14] James, G. M., 2002: Generalized linear models with functional predictors. *Journal of the Royal Statistical Society: Series B*, **64**, 411–432.
- [15] James, G. M., J. Wang, and J. Zhu, 2009: Functional linear regression that's interpretable. *Annals of Statistics*, **37**, no. 5A, 2083–2108.
- [16] Jung, S., I. L. Dryden, and J. S. Marron, 2012: Analysis of principal nested spheres. *Biometrika*, **99**, no. 3, 551–568.
- [17] Kargl, S., K. Williams, T. Marston, J. Kennedy, and J. Lopes, 2010: Acoustic response of unexploded ordnance (UXO) and cylindrical targets) and cylindrical targets. *Proc. of MTS/IEEE Oceans 2010 Conference*, 1 –5.
- [18] Keogh, E., Q. Zhu, B. Hu, Y. Hao, X. Xi, L. Wei, and C. A. Ratanamahatana, 2001: *The UCR time series classification/clustering homepage*.
URL www.cs.ucr.edu/~eamonn/time_series_data/
- [19] Kurtek, S., 2017: A geometric approach to pairwise Bayesian alignment of functional data using importance sampling. *Electronic Journal of Statistics*, **11**, no. 1, 502–531.
- [20] Kurtek, S. and K. Bharath, 2015: Bayesian sensitivity analysis with Fisher–Rao metric. *Biometrika*, **102**, no. 3, 601–616.
- [21] Kurtek, S., A. Srivastava, and W. Wu, 2011: Signal estimation under random time-warpings and nonlinear signal alignment. *Proceedings of Neural Information Processing Systems (NIPS)*.
- [22] Lahiri, S., D. Robinson, and E. Klassen, 2015: Precise matching of PL curves in \mathbb{R}^n in the Square Root Velocity framework. *Geometry, Imaging and Computing*, **2**, 133–186.
- [23] Lee, S. and S. Jung, 2017: Combined analysis of amplitude and phase variations in functional data. *arXiv:1603.01775 [stat.ME]*, 1–21.
URL <https://arxiv.org/abs/1603.01775>

- [24] Lu, Y., R. Herbei, and S. Kurtek, 2017: Bayesian registration of functions with a Gaussian process prior. *Journal of Computational and Graphical Statistics*, **26**, no. 4, 894–904.
- [25] Marron, J. S., J. Ramsay, L. Sangalli, and A. Srivastava, 2015: Functional data analysis of amplitude and phase variation. *Statistical Science*, **30**, no. 4, 468–484.
- [26] Mordecai, A., 2003: *Nonlinear Programming: Analysis and Methods*. Dover Publishing.
- [27] Müller, H. G. and U. Stadtmüller, 2005: Generalized functional linear models. *Annals of Statistics*, **33**, no. 2, 774–805.
- [28] Ramsay, J. O. and C. J. Dalzell, 1991: Some tools for functional data analysis. *Journal of the Royal Statistical Society, Ser. B*, **53**, no. 3, 539–572.
- [29] Ramsay, J. O. and B. W. Silverman, 2005: *Functional Data Analysis*. Springer.
- [30] Reiss, P. and R. Ogden, 2007: Functional principal component regression and functional partial least squares. *Journal of American Statistical Association*, **102**, no. 479, 984–996.
- [31] Robinson, D., 2012: *Functional analysis and partial matching in the square root velocity framework*. Ph.D. thesis, Florida State University.
- [32] Srivastava, A. and I. H. Jermyn, 2009: Looking for shapes in two-dimensional, cluttered point clouds. *IEEE Trans. Pattern Analysis and Machine Intelligence*, **31**, no. 9, 1616–1629.
- [33] Srivastava, A., S. H. Joshi, W. Mio, and X. Liu, 2005: Statistical shape analysis: Clustering, learning and testing. *IEEE Trans. Pattern Analysis and Machine Intelligence*, **27**, no. 4, 590–602.
- [34] Srivastava, A. and E. Klassen, 2016: *Functional and Shape Data Analysis*. Springer-Verlag New York.
- [35] Srivastava, A., E. Klassen, S. Joshi, and I. Jermyn, 2011: Shape analysis of elastic curves in euclidean spaces. *IEEE Trans. Pattern Analysis and Machine Intelligence*, **33**, no. 7, 1415–1428.
- [36] Srivastava, A., W. Wu, S. Kurtek, E. Klassen, and J. S. Marron, 2011: Registration of functional data using fisher-rao metric. *arXiv:1103.3817v2 [math.ST]*.
URL <http://arxiv.org/abs/1103.3817v2>
- [37] Tucker, J. D., W. Wu, and A. Srivastava, 2013: Generative models for functional data using phase and amplitude separation. *Computational Statistics and Data Analysis*, **61**, 50–66.
- [38] — 2014: Analysis of signals under compositional noise with applications to sonar data. *IEEE Journal of Oceanic Engineering*, **39**, no. 2, 318–330.

- [39] Veeraraghavan, A., A. Srivastava, A. K. Roy-Chowdhury, and R. Chellappa, 2009: Rate-invariant recognition of humans and their activities. *IEEE Transactions on Image Processing*, **8**, no. 6, 1326–1339.

AUTHOR BIOGRAPHY



J. Derek Tucker. received his B.S. in Electrical Engineering Cum Laude and M.S. in Electrical Engineering from Colorado State University in 2007 and 2009, respectively. Upon completion of these degrees, he began working as a Research Scientist at the Naval Surface Warfare Center Panama City Division in Panama City, FL. In 2014 he received the Ph.D. degree in Statistics from Florida State University In Tallahassee, FL. While at NSWC-PCD he led various development efforts in automatic target recognition algorithms for synthetic aperture sonar imagery. In the summer of 2014 he joined Sandia National Laboratories and is currently in the Statistical Sciences department. He currently is leading research projects in the area of satellite image registration and point processes modeling for monitoring applications. His research is focused on pattern theoretic approaches to problems in image analysis, computer vision, signal processing, and functional data analysis. In 2017, he received the Director of National Intelligence Team Award for his contributions to the Signal Location in Complex Environments (SLiCE) team.



John R. Lewis. joined Sandia in 2014 shortly after finishing a Ph.D. in Statistics at The Ohio State University. For his dissertation he developed methods for conditioning on insufficient statistics in Bayesian models for the purposes of robustness. At Sandia, John supports a variety of projects using a wide range of statistical methodologies. These methodologies include functional data analysis, design and analysis of computer experiments, uncertainty quantification (UQ), and design of experiments. Other areas of interest include robust estimation techniques, spatiotemporal modeling, and machine learning.



Anuj Srivastava. received the M.S. and Ph.D. degrees in electrical engineering from Washington University, St. Louis, Missouri. in 1993 and 1996, respectively. After spending the year 1996-1997 at Brown University as a visiting researcher, he joined the Department of Statistics at Florida State University (FSU), Tallahassee, as an assistant professor in 1997. He is currently a professor in that department. He has held visiting appointments at University of Lille, France; INRIA, Sophia Antipolis, France; and, UCL, Louvain-Le-Neuve, Belgium. His research is focused on pattern theoretic approaches to problems in image analysis, computer vision, and signal processing. In particular, he has developed computational tools for performing statistical inferences on certain nonlinear manifolds, including shape manifolds of curves and surfaces. He has published more than 170 peer-reviewed journal and conference articles in these areas. He is a fellow member of the ASA and IEEE.

