Elastic Functional Data Analysis

J. Derek Tucker

Nov 4, 2022

Department of Statistical Sciences
Sandia National Laboratories

Sandia National Laboratories is a multimission 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-NA-0003525.
SAND NO. 2018-1615 PE

Collaborators

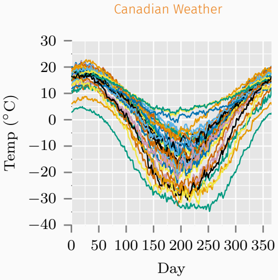
- Anuj Srivastava, Dept of Statistics, FSU
- Sebastian Kurtek, Dept of Statistics, OSU
- Zhengwu Zhang, Dept of Statistics, URMC
- Bo Li, Dept of Statistics, UIUC
- Statistics Sciences, SNL

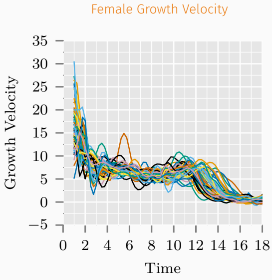
Outline

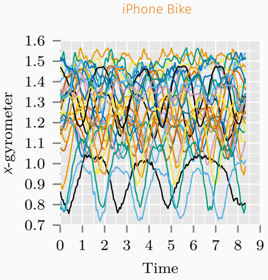
- Definition of Functional Data Analysis
- Mathematical Framework
  - FDA vs Multivariate Statistics
  - Common Metric
- Alignment of Functional Data
  - Elastic Method
- Elastic Methods
  - Elastic Functional Principal Component Analysis
  - Elastic Functional Bayesian Model Calibration

Functional Data Analysis

- Problem of statistical analysis of function data (FDA) is important in a wide variety of applications
- Easily encounter a problem where the observations are real-valued functions on an interval, and the goal is to perform their statistical analysis
- By statistical analysis we mean to compare, align, average, and model a collection of random observations





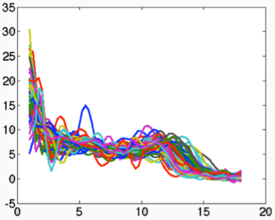


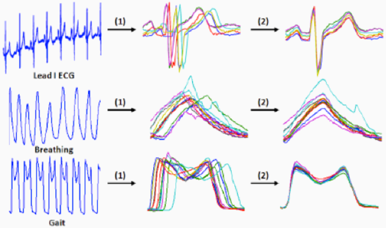
Introduction

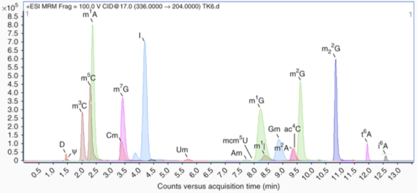
- Questions then arise on how can we model the functions
  - Can we use the functions to classify diseases?
  - Can we use them as predictors in a regression model?
  - It is the same goal (question) of any area of statistical study
- One problem occurs when performing these type of analyses is that functional data can contain variability in time (x-direction) and amplitude (y-direction)
- How do we account for and handle this variability in the models that are constructed from functional data?

Types of Functional Data

- Real-valued functions, with interval domain: $f : [a, b] \to \mathbb{R}$

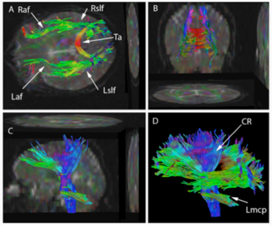


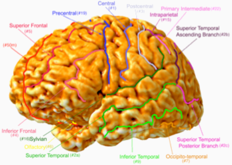


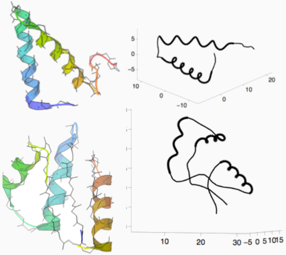


Types of Functional Data

- $\mathbb{R}^n$-valued functions with interval domain, Or Parameterized Curves: $f : [a, b] \to \mathbb{R}^n$
  $f : S^1 \to \mathbb{R}^n$

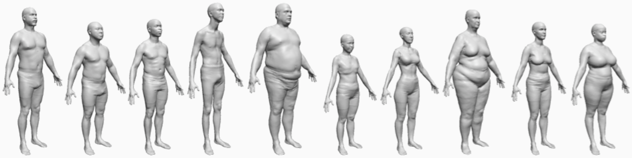


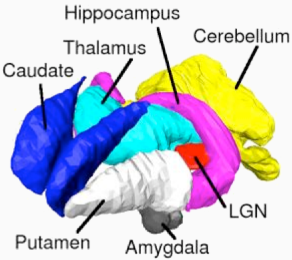


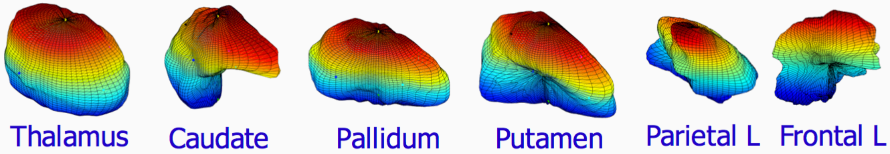


Types of Functional Data

- $\mathbb{R}^3$-R3-valued functions on a spherical domain, Or Parameterized Surfaces: $f : S^2 \to \mathbb{R}^3$

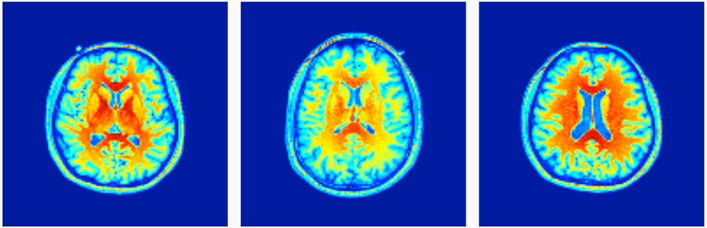






Types of Functional Data

- $\mathbb{R}^n$-valued functions with square or cube domains, Or Images: $f : [0,1]^2 \to \mathbb{R}^n$

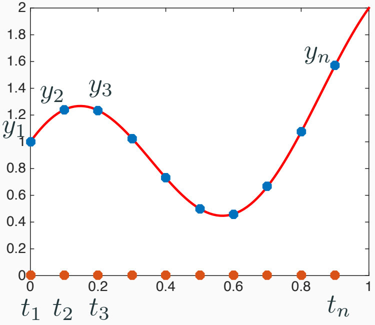


FDA vs Multivariate Statistics

Why FDA? Why Not Multivariate Statistics

- In any computer implementation, one has to discretize functions anyway
- Does this mean FDA is essentially the same as multivariate statistics?
- A closer look...

FDA Versus Multivariate Statistics



Multivariate data

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{bmatrix}$$

Functional data

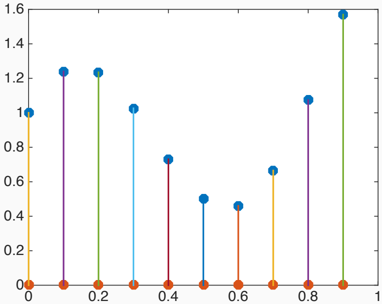
$$\begin{bmatrix} (t_1, y_1) \\ (t_2, y_2) \\ (t_3, y_3) \\ \vdots \\ (t_n, y_n) \end{bmatrix}$$

- Not all observations will have the same time indices
- Even if they do, we want the ability to change time indices

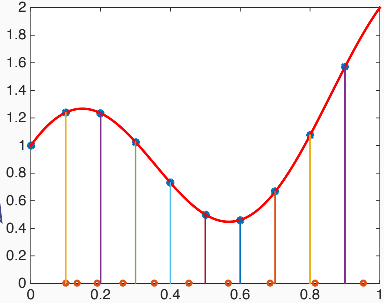
FDA Versus Multivariate Statistics

Functional data

$$\begin{bmatrix} (t_1, y_1) \\ (t_2, y_2) \\ (t_3, y_3) \\ \vdots \\ (t_n, y_n) \end{bmatrix}$$



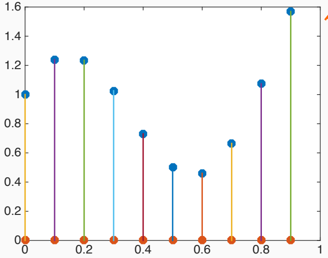
Interpolation

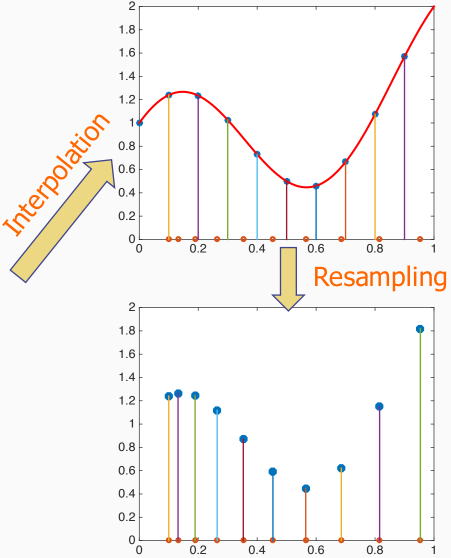


FDA Versus Multivariate Statistics

Functional data

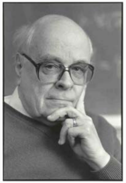
$$\begin{bmatrix} (t_1, y_1) \\ (t_2, y_2) \\ (t_3, y_3) \\ \vdots \\ (t_n, y_n) \end{bmatrix}$$





FDA Versus Multivariate Statistics

- In FDA, one develops the theory on function spaces and not finite vectors, and discretizes the functions only at the final step – computer implementation



- Ulf Grenander: "Discretize as late as possible" (1924-2016)
- Even after discretization, we retain the ability to interpolate resample as needed!

Mathematical Framework

Common Metric Structure for FDA

- Let f be a real-valued function with the domain $[0, 1]$, can be extended to any domain
  - Only functions that are absolutely continuous on $[0, 1]$ will be considered
- The $\mathbb{L}^2$ inner-product:

$$\langle f_1, f_2 \rangle = \int_0^1 f_1(t) f_2(t)\, dt$$

- $\mathbb{L}^2$ distance between functions:

$$||f_1 - f_2|| = \sqrt{\int_0^1 (f_1(t) - f_2(t))^2\, dt}$$

- From these we will build summary statistics, but how good are they?

Summary Statistics under $\mathbb{L}^2$

- Assume that we have a collection of functions, $f_i(t)$, $i = 1, \ldots, N$ and we wish to calculate statistics on this set
- Mean Function

$$\bar{f}(t) = \arg\min_{f \in \mathbb{L}^2} \left(\sum_{i=1}^{N} ||f - f_i||^2 \right)$$

$$\bar{f}(t) = \frac{1}{N} \sum_{i=1}^{n} f_i(t)$$

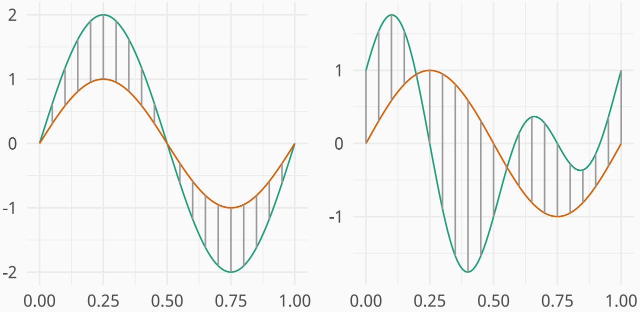
- Variance Function

$$\mathrm{var}(f(t)) = \frac{1}{N-1} \sum_{i=1}^{N} \left(f_i(t) - \bar{f}(t)\right)^2$$

- How good is this choice in FDA?

Common Metric Structure for FDA

- Here, the focus is on measuring/modeling the vertical variability in the data

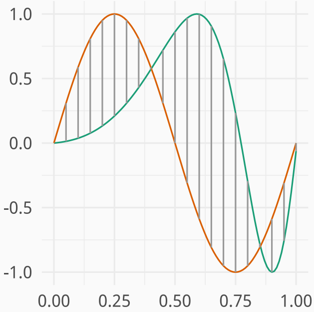


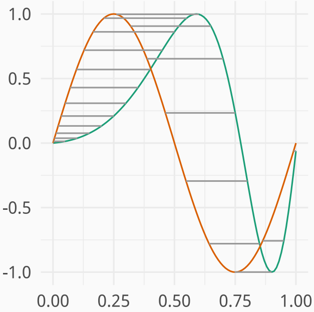
- Measures the norm of the difference $(f_1(t) - f_2(t))$

$$||f_1 - f_2|| = \sqrt{\int_0^1 (f_1(t) - f_2(t))^2 \, dt}$$

Horizontal Variability

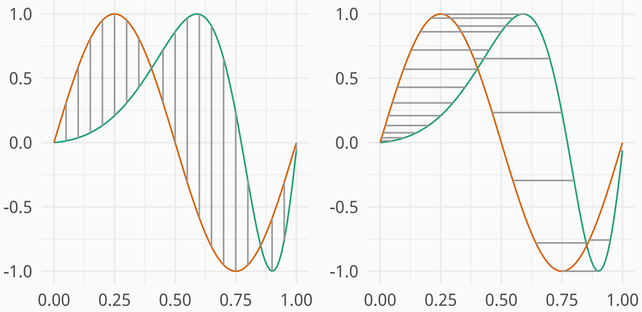
- Horizontal variability: Compares points at same heights but across times





Horizontal Variability

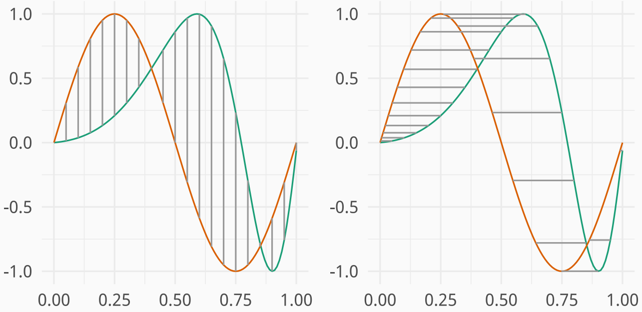
- Horizontal variability: Compares points at same heights but across times



- Is this vertical or horizontal variability?

Horizontal Variability

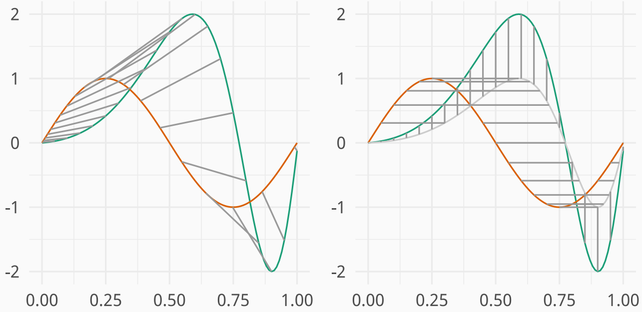
- Horizontal variability: Compares points at same heights but across times



- Is this vertical or horizontal variability?
- In some cases it may be more natural to treat it as horizontal variability

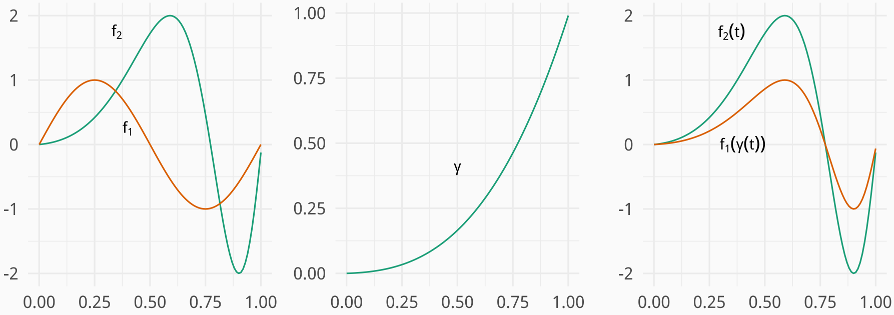
Both Vertical and Horizontal Variability

- In general, functional data has both types of variability,
- How to decompose it into vertical and horizontal components



Separate Vertical and Horizontal

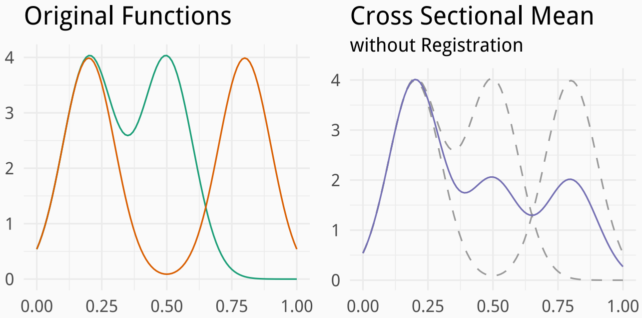
Time warp functions to align their peaks and valleys



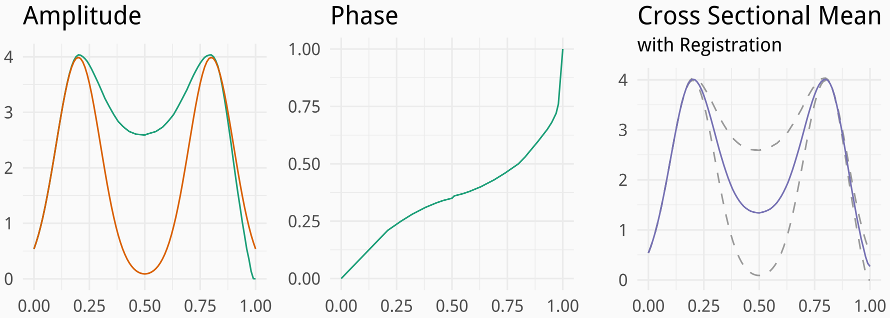
Elastic FDA: Ability to separate and analyze these components, and to draw inferences using both these components of functional data

Classical FDA: Loss of Structure

Cross-sectional statistics ignores horizontal component, often destroys structures

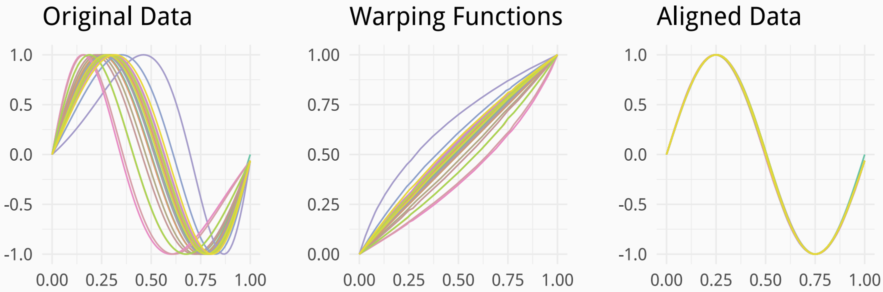


Classical FDA: Loss of Structure



Classical FDA: Inflated Variance

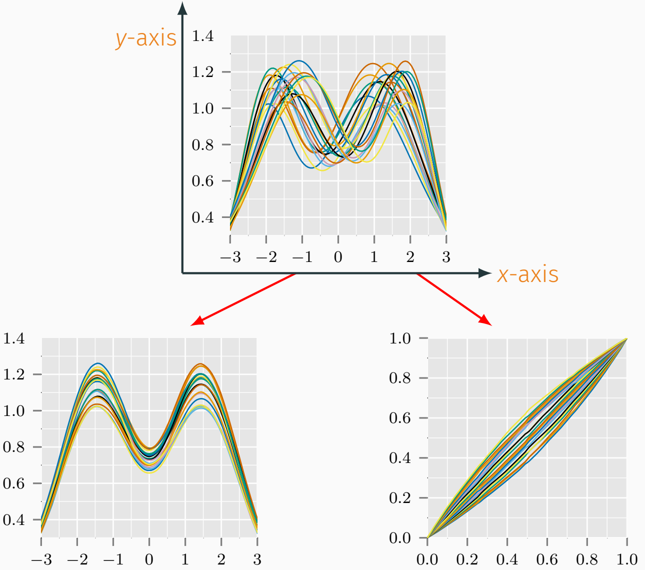
Phase variability artificially inflates variance



High horizontal variance and low vertical variance after alignment

Functional Data Alignment

Functional Data Alignment Improves Model Parsimony



Aligned Functions: "Amplitude" Warping Functions: "Phase"

Two Types of Problems

Problem 1: Alignment of given functional data

Goal: Choose some objective function and optimize alignment. Provide the best alignment algorithm in the community. Decrease amplitude variability as much as possible.

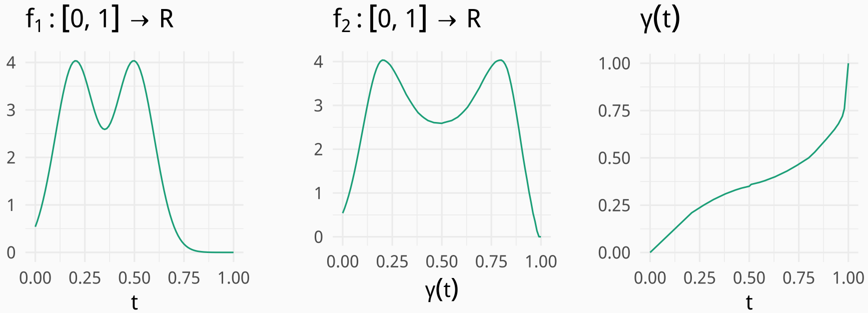
Problem 2: Joint alignment and statistical analysis, i.e. Elastic FDA

Goal: Align the data in the context of a statistical inference problem. For example: Perform joint PCA and alignment $\rightarrow$ Elastic FPCA

Our framework provides solutions in both contexts...

We start with the first problem..

Formulating Registration Problem



The point $f_1(t)$ gets matched with the point $f_2(\gamma(t))$

Define a set of registration functions

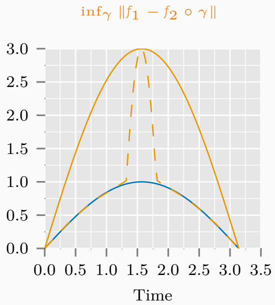
$$\Gamma = \{\gamma : [0,1] \mapsto [0,1] | \gamma(0) = 0, \gamma(1) = 1, \ \text{diffeo}\}$$

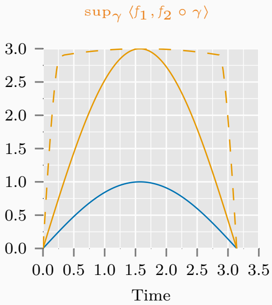
$\mathbb{L}^2$-based Objective Function

- Given f_1, f_2, we have to search for an optimal γ
- What is a good choice of objective function for registration? $\mathbb{L}^2$ norm.

$$\arg\min_{\gamma}(\|f_1 - f_2 \circ \gamma\|^2)$$

- This can result in "pinching effect"





Obejctive Function

- Given f_1, f_2, we have to search for an optimal γ
- What is a good choice of objective function for registration? $\mathbb{L}^2$ norm.

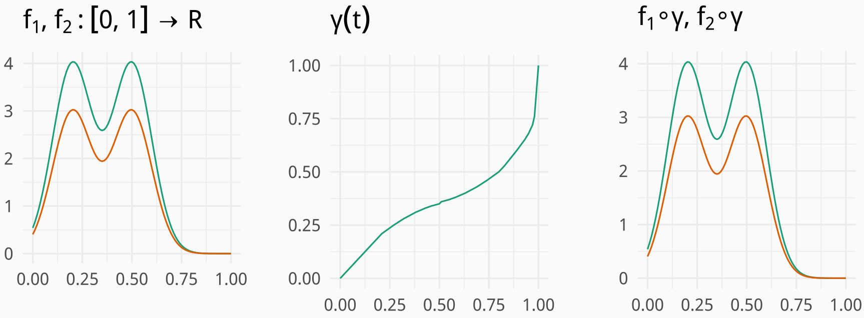
$$\arg\min_{\gamma}(\|f_1 - f_2 \circ \gamma\|^2)$$

- This can result in "pinching effect"
- Common solution: Regularize

$$\arg\min_{\gamma}(\|f_1 - f_2 \circ \gamma\|^2 + \lambda R(\gamma)), \quad R(\gamma) = \int_0^1 \ddot{\gamma}\, dt$$

- Problems: solution is not inverse consistent, lack of invariance

Requirements for Elastic FDA



The two functions have the same correspondences before and after warping $\rightarrow$ objective function should not change!!

However, the $\mathbb{L}^2$ norm changes. Thus, it is not a good metric for Elastic FDA

One Solution: SRVF Representations

$$f : [0,1] \to \mathbb{R}^1$$

- Define Square Root Velocity Function SRVF

$$q : [0,1] \to \mathbb{R}^1, \quad q(t) = \mathrm{sgn}(\dot{f}(t))\sqrt{|\dot{f}(t)|}$$

One Solution: SRVF Representations

$$f : [0,1] \to \mathbb{R}^1$$

- Define Square Root Velocity Function SRVF

$$q : [0,1] \to \mathbb{R}^1, \quad q(t) = \mathrm{sgn}(\dot{f}(t))\sqrt{|\dot{f}(t)|}$$

- if f is absolutely continuous, then q is square-integrable

One Solution: SRVF Representations

$$f : [0, 1] \to \mathbb{R}^1$$

- Define Square Root Velocity Function SRVF

$$q : [0, 1] \to \mathbb{R}^1, \quad q(t) = \mathrm{sgn}(\dot{f}(t))\sqrt{|\dot{f}(t)|}$$

- if f is absolutely continuous, then q is square-integrable
- If a function f is warped to $f \circ \gamma$ then its SRVF changes form q to $(q, \gamma)(t) = (q \circ \gamma)(t)\sqrt{\dot{\gamma}(t)}$

One Solution: SRVF Representations

$$f : [0,1] \to \mathbb{R}^1$$

- Define Square Root Velocity Function SRVF

$$q : [0,1] \to \mathbb{R}^1, \quad q(t) = \operatorname{sgn}(\dot{f}(t))\sqrt{|\dot{f}(t)|}$$

- if f is absolutely continuous, then q is square-integrable
- If a function f is warped to $f \circ \gamma$ then its SRVF changes form q to $(q, \gamma)(t) = (q \circ \gamma)(t)\sqrt{\dot{\gamma}(t)}$
- One can go back from SRVF to f; up to an additive constant

$$f(t) = \int_0^t q(s)|q(s)|\, ds$$

One Solution: SRVF Representations

$$f : [0, 1] \to \mathbb{R}^1$$

- Define Square Root Velocity Function SRVF

$$q : [0, 1] \to \mathbb{R}^1, \quad q(t) = \mathrm{sgn}(\dot{f}(t))\sqrt{|\dot{f}(t)|}$$

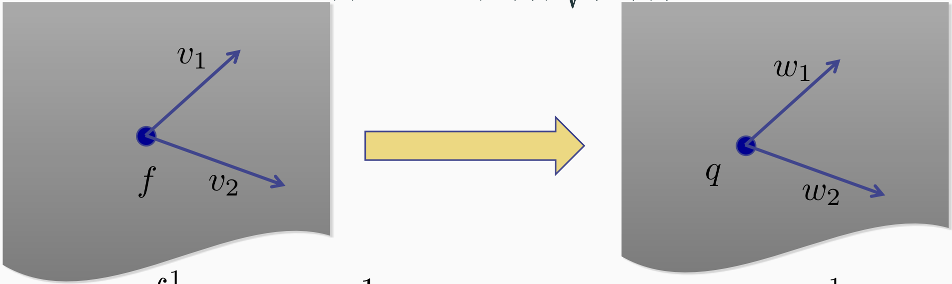
- if f is absolutely continuous, then q is square-integrable
- If a function f is warped to $f \circ \gamma$ then its SRVF changes form q to $(q, \gamma)(t) = (q \circ \gamma)(t)\sqrt{\dot{\gamma}(t)}$
- One can go back from SRVF to f; up to an additive constant

$$f(t) = \int_0^t q(s)|q(s)|\, ds$$

- The space of all SRVFs is $\mathbb{L}^2([0, 1], \mathbb{R})$

Elastic (Fisher-Rao) Metric

$$q(t) = \mathrm{sign}(\dot{f}(t))\sqrt{|\dot{f}(t)|}$$



$$\langle\langle v_1, v_2 \rangle\rangle_f = \int_0^1 \dot{v}_1(t)\dot{v}_2(t)\frac{1}{\dot{f}(t)}dt$$

$$\langle w_1, w_2 \rangle = \int_0^1 w_1(t)w_2(t)dt$$

- The elastic metric has the right invariant properties (stated later)
- However, it is complicated to use
- The mapping $f \mapsto q$ that simplifies this metric into the standard $\mathbb{L}^2$ metric

Different Versions of Fisher-Rao Riemannian Metric

1. Function or CDF Space

$$\langle\langle v_1, v_2\rangle\rangle_f = \int_0^1 \dot{v}_1(t)\dot{v}_2(t)\frac{1}{\dot{f}(t)}\,dt$$

2. PDF Space (Non parametric Fisher-Rao)

$$\langle\langle u_1, u_2\rangle\rangle_g = \int_0^1 u_1(t)u_2(t)\frac{1}{g(t)}\,dt$$

3. PDF Space (Parametric Fisher-Rao)

$$\int_0^1 \left(\frac{\partial}{\partial\theta_i}g(t|\theta)\right)\left(\frac{\partial}{\partial\theta_j}g(t|\theta)\right)\frac{1}{g(t|\theta)}\,dt$$

4. SRVF Space

$$\langle w_1, w_2\rangle = \int_0^1 w_1(t)w_2(t)\,dt$$

SRVF Representation Space

$$f : [0,1] \to \mathbb{R}^1$$

- Why Square Root Velocity Function SRVF? $q(t) = \operatorname{sgn}(\dot{f}(t))\sqrt{|\dot{f}(t)|}$
  - Invariance: for any q_1, q_2, and γ

$$||q_1 - q_2|| = ||(q_1, \gamma) - (q_2, \gamma)||$$

  - In particular, $||q|| = ||(q, \gamma)||$ and hence pinching is not possible
  - Resulting registration problem: Given f_1 and f_2, find their SRVFs, and solve

$$\inf_{\gamma} ||q_1 - (q_2, \gamma)||$$

  - Solve using Dynamic Programming
  - Inverse consistency

if $\gamma_{12} \in \arg\inf_{\gamma} ||q_1 - (q_2, \gamma)||$ then $\gamma_{12}^{-1} \in \arg\inf_{\gamma} ||q_2 - (q_1, \gamma)||$

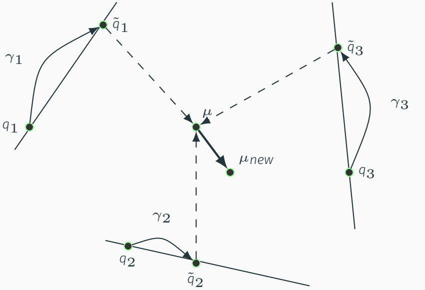
Multiple Registration

- Using the distance function we can compute the Karcher Mean

$$\mu_q = \underset{q \in \mathbb{L}^2}{\arg\min} \sum_{i=1}^{n} \left(\inf_{\gamma_i \in \Gamma} \| q - (q_i, \gamma_i) \|^2 \right)$$

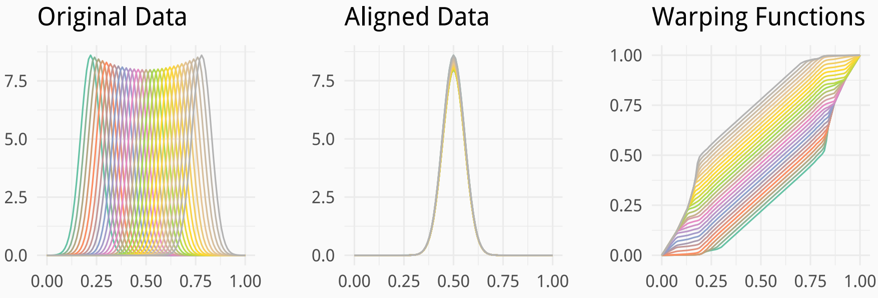
  without a metric, we cannot define the mean

- Algorithm for computing the Karcher mean:



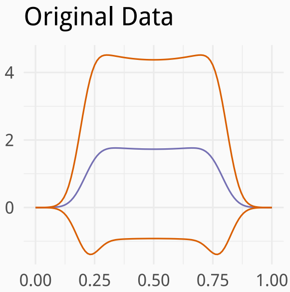
Example: Simulated Data

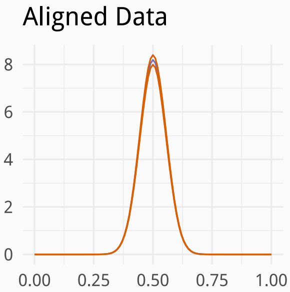
Ensemble Alignment using Karcher Mean



Example: Cross-Sectional Statistics

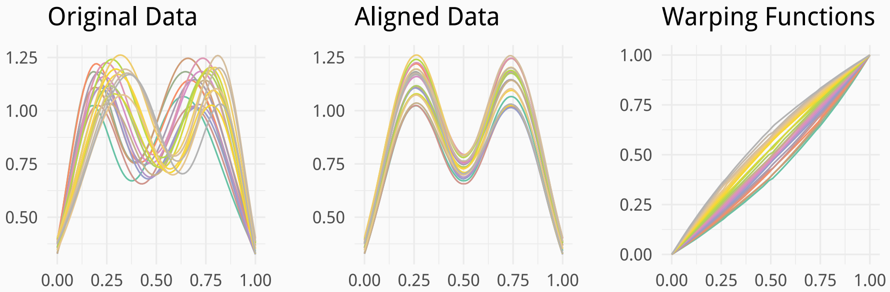
Cross-sectional mean and standard deviation improves after registration





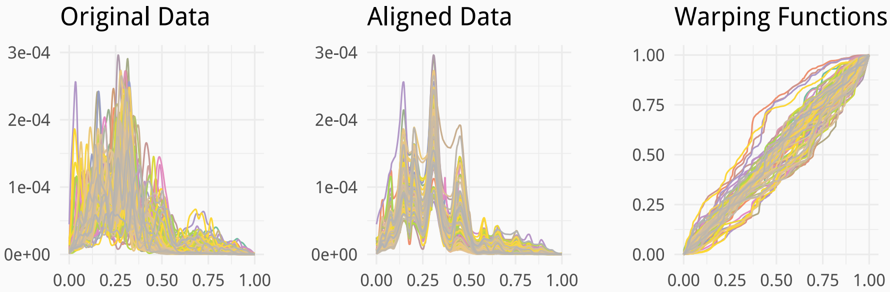
Example: Simulated Data

Ensemble Alignment using Karcher Mean



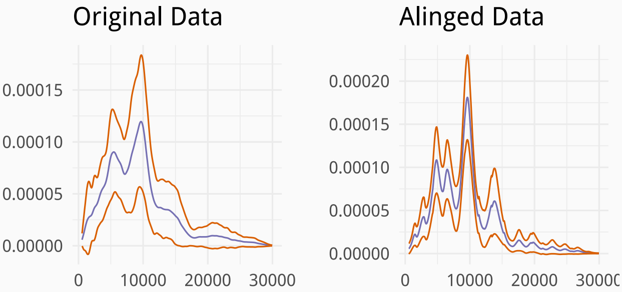
Example: Sonar Data

Sonar data from Naval Surface Warfare Center: Aspect versus Frequency



Sonar Data: Cross-Sectional Statistics

Cross-sectional mean and standard deviation improves after registration



Elastic Functional PCA

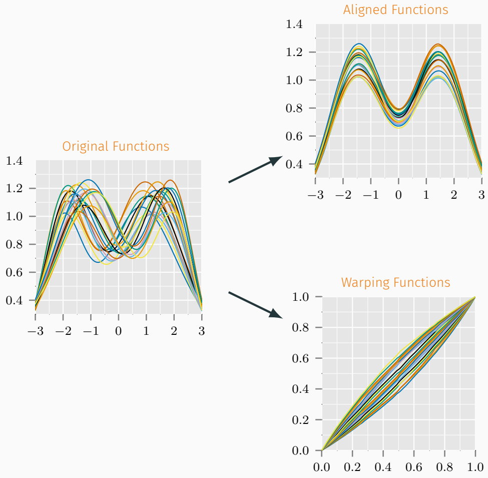
Functional Principal Component Analysis

- The motivation for functional principal component analysis (fPCA) is that the directions of high variance will contain more information than direction of low variance
- The optimization problem for fPCA

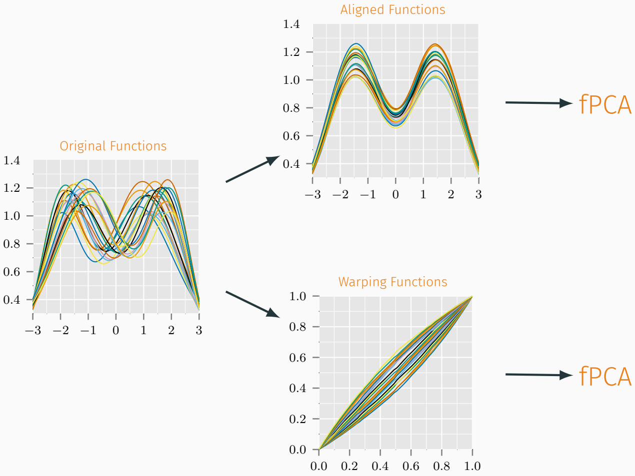
$$\min_{w_i} E\|f - \hat{f}\|^2$$

- where $\hat{f} = \mu_f + \sum_{i=1}^{n} \beta_i w_i(t)$ is the fPCA approximation of f
- We then use the sample covariance function $\mathrm{cov}(t_1, t_2)$ to form a sample covariance matrix K
- Taking the SVD, $K = U\Sigma V^{\mathsf{T}}$ we can calculate the directions of principle variability in the given functions using the first $p \leq n$ columns of U

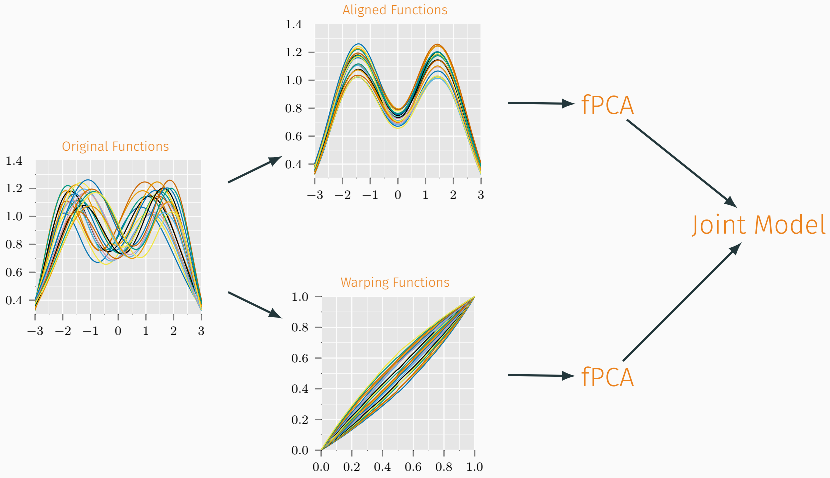
Modeling using Phase & Amplitude Separation



Modeling using Phase & Amplitude Separation

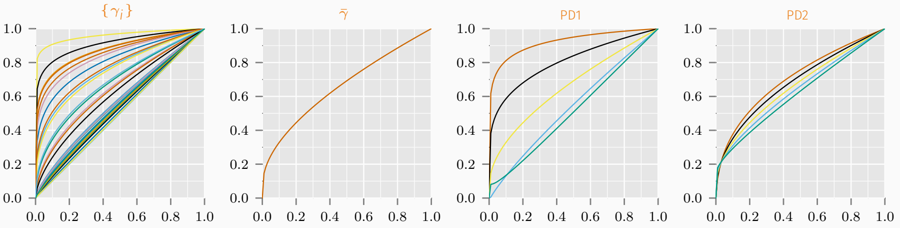


Modeling using Phase & Amplitude Separation

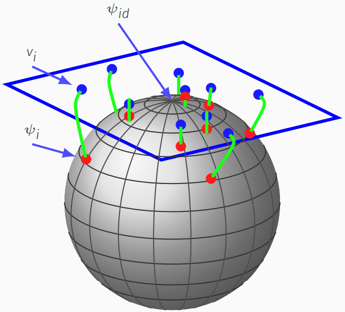


Analysis of Warping Functions (Phase)

- Horizontal fPCA: Analysis of Warping Functions
  - Use SRVF of warping functions, $\psi = \sqrt{\dot{\gamma}}$
  - Karcher Mean: $\gamma \mapsto \sum_{i=1}^{n} d_p(\gamma, \gamma_i)^2$
  - Tangent Space:$T_\psi(\mathbb{S}_\infty) = \{v \in \mathbb{L}^2 | \int_0^1 v(t)\psi(t)dt = 0\}$
  - Sample Covariance Function: $(t_1, t_2) \mapsto \frac{1}{n-1} \sum_{i=1}^{n} v_i(t_1)v_i(t_2)$
  - Take SVD of $K_\psi = U_\psi \Sigma_\psi V_\psi^T$ provides the estimated principal components



Why SRVF of γ_i



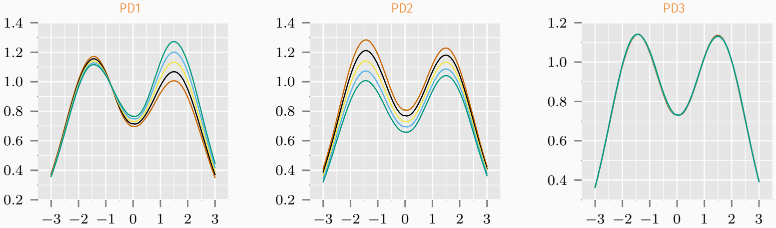
- Γ is a nonlinear manifold and it is infinite dimensional
- Represent an element $\gamma \in \Gamma$ by the square-root of its derivative $\psi = \sqrt{\dot{\gamma}}$
- Important advantage of this transformation is that set of all such ψs is a Hilbert sphere $\mathbb{S}_\infty$

Analysis of Aligned Functions (Amplitude)

- Vertical fPCA: Analysis of Aligned Functions
  - Given the observed SRVF have been aligned
  - They can be analyzed in a standard way ($\mathbb{L}^2$) in SRVF space, since we have a proper distance
  - Need variability associated with the initial values ($\{f_i(0)\}$)
    - Analyze the aligned pair $\tilde{h} = [\tilde{q}_i \;\; f_i(0)]$ such that mapping from the function space $\mathcal{F}$ to $\mathbb{L}^2 \times \mathbb{R}$ is a bijection

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

  - Taking the SVD, $K_h = U_h \Sigma_h V_h^T$ we can calculate the directions of principle variability

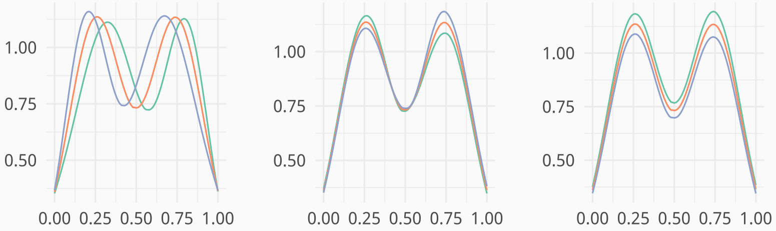


Combined Elastic fPCA

- Recently [Lee 2017] extended the horizontal and vertical fPCA approach
- Uses a combined function

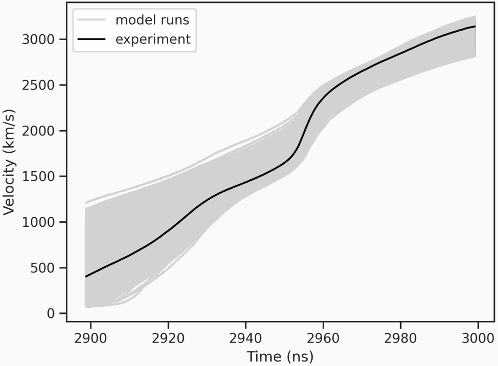
$$g^C(t) = \begin{cases} q^*(t), & t \in [0,1) \\ Cv(t-1), & t \in [1,2] \end{cases}$$

- where C is again used to adjust for the scaling imbalance between q^* and v
- Taking the SVD, $K_g^C = U_g^C \Sigma_g^C (V_g^C)^{\mathsf{T}}$, accounts for correlation between amplitude and phase



Elastic Functional Bayesian Model Calibration

Elastic Model Calibration



- We wish to calibrate a computer model with parameters θ to a experiment simulation
- The data is functional in nature and has phase and amplitude variability
- Utilize elastic metrics in a Bayesian Model Calibration Framework

Elastic Model Calibration

- Decompose observation into aligned functions and warping functions

$$y_i^E(t) = y_i^E(t^*) \circ \gamma_i^E(t)$$

- and decompose the simulations

$$y^M(t, x_j) = y^M(t^*, x_j) \circ \gamma^M(t, x_j)$$

To facilitate modeling, we transform the warping functions into shooting vector space with

$$v_i^E = \exp_\psi^{-1}\left(\sqrt{\dot{\gamma}_i^E}\right)$$

$$v^M(x) = \exp_\psi^{-1}\left(\sqrt{\dot{\gamma}^M(x)}\right)$$

Elastic Model Calibration

- Calibrate the aligned data and shooting vectors using the following model

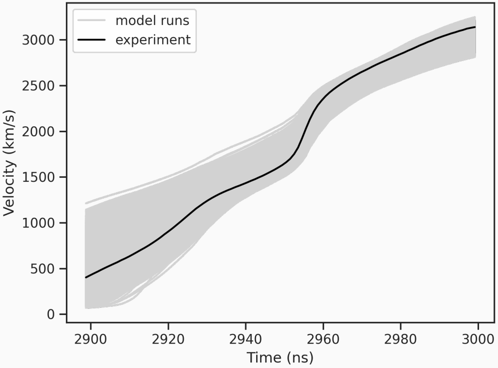
$$y^E(t^*) = y^M(t^*, \theta) + \delta_y(t^*) + \epsilon_y(t^*),\ \epsilon_y(t^*) \sim \mathcal{N}(0, \sigma_y^2 I)$$

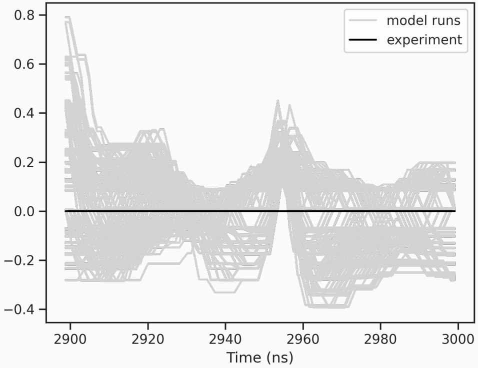
$$v^E = v^M(\theta) + \delta_v + \epsilon_v,\ \epsilon_v \sim \mathcal{N}(0, \sigma_v^2 I)$$

- Note: The shooting vector will be identity if the data is aligned to the observation (experiment)
- There if θ is calibrated correctly the shooting vectors will be identity

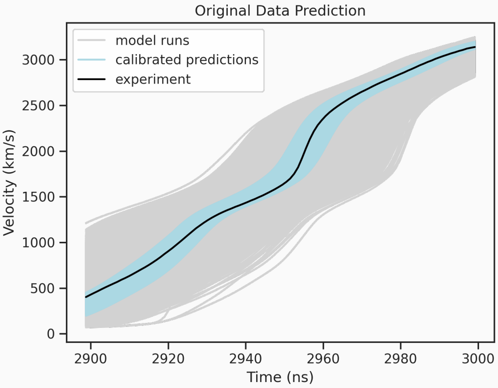
Calibration of Tantalum

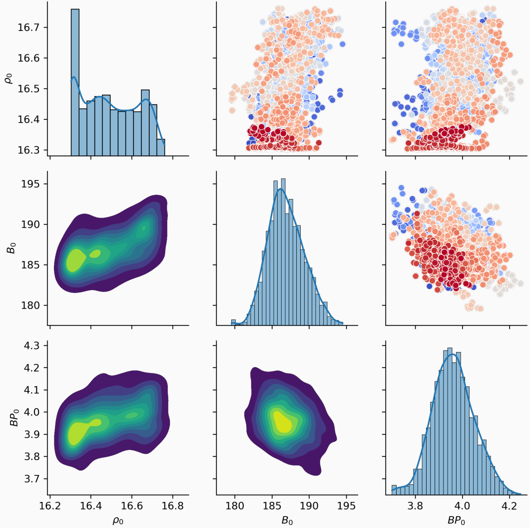
- Calibration of stress-strain model of Tantalum





Calibration of Tantalum





Summary

- FDA is a very rapidly growing area in statistics with the increase in sensors and dimensionality of data
- Can perform statistics using functions, but have to be aware of different set of issues/nuances
- Functional data often comes with phase variability that cannot be handled using standard $\mathbb{L}^2$ framework
- Elastic FDA provides more flexibility than classical FDA
    - Provides excellent alignment results
    - Provides joint solutions for inferences along with alignment
- Theory and methods work for functions, curves, surfaces, and images

Questions?

jdtuck@sandia.gov
http://research.tetonedge.net