Computational Statistics and Data Analysis 125 (2018) 10–26

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

Robust template estimation for functional data with phase variability using band depth

Jason Cleveland, Weilong Zhao, Wei Wu *

Department of Statistics, Florida State University, 117 N Woodward Ave, Tallahassee, FL, 32306-4330, USA

ARTICLE INFO

Article history:
Received 3 July 2017
Received in revised form 5 March 2018
Accepted 10 March 2018
Available online 29 March 2018

Keywords:
Functional data analysis
Template estimation
Registration
Time warping
Band depth

ABSTRACT

Registration, or alignment, of functional observations has been a fundamental problem in functional data analysis. The creation of a template was the key step for alignment of a group of functions. Recent studies have defined the template with the notion of "mean" in the given observations. However, the mean can be sensitive to the, commonly observed, outlier functions in the data. To deal with this problem, a new approach is proposed to adopt the notion of "median" using the time warping functions in the alignment process, based on the recently developed band depth in functional data. A semi-parametric model is provided with an algorithm that yields a consistent estimator for the underlying median template. The robustness of this depth-based registration is illustrated using simulations and two real data sets. In addition, a new depth-based boxplot is proposed for outlier detection in functional data with phase variability.

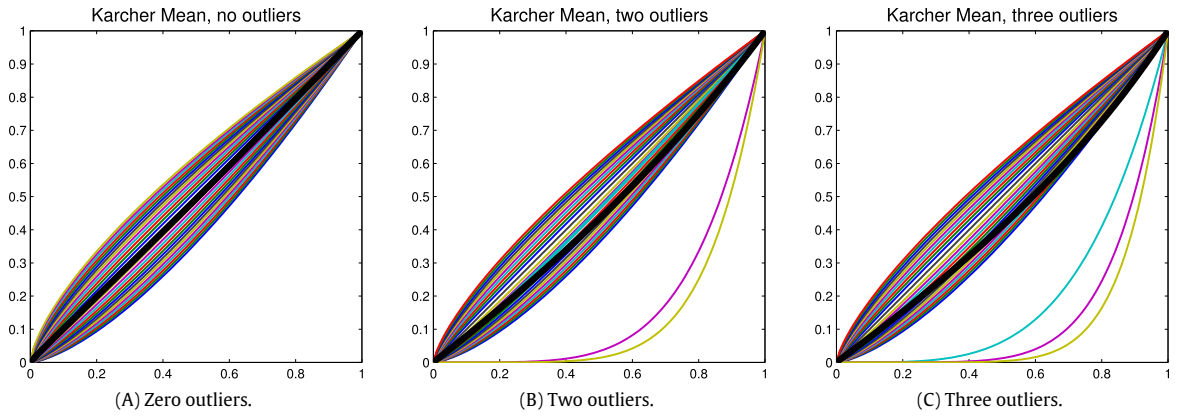
1. Introduction

The field of functional data analysis (FDA) has made great strides in mathematically modeling time-dependent observations. Progress towards handling temporal or phase variability along the time axis has addressed some of the issues once facing conventional analysis methods, such as spline smoothing, covariance or correlation analysis, functional principal component analysis, and functional linear or nonlinear regressions (Kneip and Gasser, 1992; Gasser and Kneip, 1995; Wang and Gasser, 1997; Ramsay and Li, 1998). The general solution to the phase variability, or time warpings, has been to apply a computational approach that removes such warpings prior to analysis of the data (Ramsay and Silverman, 2005; Ferraty and Vieu, 2006). Various approaches over the past two decades depend on finding a way to perform an optimal matching process (a procedure meant to match important functional features in the observations) called registration or curve alignment (Liu and Muller, 2004; Tang and Muller, 2008; Gervini and Gasser, 2004; Kneip and Ramsay, 2008; James, 2007; Capra and Muller, 1997; Ramsay and Silverman, 2005; Kneip and Gasser, 1992; Ramsay and Li, 1998; Hall et al., 2007; Srivastava et al., 2011). The basic procedure to register a set of functional observations is to find a template and then align each function to the template. The template is often defined as the *mean* or *average* of the data under a given metric. However, it is well known that the mean can be sensitive to outliers, which commonly appear in practical observations.

For example, the Karcher mean template with the elastic metric in the Fisher–Rao registration framework (Srivastava et al., 2011) is sensitive to outliers thus making it a poor choice for estimation. We can see in Fig. 1 the Karcher mean of a set of time warping functions can significantly change with respect to a small number of outliers in the data. In that regard, we will need a robust signal estimation with respect to potential outliers. The median is a naturally desired candidate template for our signal estimation. The challenging question is, "How do we find the median of functional observations"?

* Corresponding author.
E-mail addresses: jcleveland@stat.fsu.edu (J. Cleveland), weilong.zhao@stat.fsu.edu (W. Zhao), wwu@stat.fsu.edu (W. Wu).

https://doi.org/10.1016/j.csda.2018.03.009
0167-9473/



(A) Zero outliers. (B) Two outliers. (C) Three outliers.

Fig. 1. Karcher mean with respect to outliers. **A.** Observed time warping functions with no outliers present. Each line denotes one function and the bold black line denotes the Karcher mean. **B.** Same as Panel A except that two outliers are added. **C.** Same as Panel A except that three outliers are added.

This question has been well addressed with the notion of depth on functional data over the past decade. The methods include h-depth (Cuevas et al., 2007), random Tukey depth (Cuesta-Albertos and Nieto-Reyes, 2008), band depth, modified band depth (Lopez-Pintado and Romo, 2009), half-region depth, modified half-region depth (López-Pintado and Romo, 2011), integrated depth (Fraiman and Muniz, 2001), multivariate functional halfspace depth (Claeskens et al., 2014), functional spatial depth (Chakraborty et al., 2014), etc. Of these methods, the band depth by Lopez-Pintado and Romo in Lopez-Pintado and Romo (2009) is probably the most commonly used. This depth is analogous to the well-known simplicial depth (Liu, 1990) on multivariate data and has plenty of computational and theoretical advantages. We will adopt this depth to estimate the template for functional observations with phase variability. Note that a robust method on functional observations has been studied with a Karcher median method (Kurtek et al., 2013), an elastic-metric-based measure of centrality for functional data. We will compare the proposed band depth approach with the Karcher median method in this manuscript.

In theory, functional observations with phase variability have been stated in the following general framework:

$$f_i(t) = (X_i \circ \gamma_i)(t) + e_i(t), \quad t \in [0, 1], \quad i = 1, \ldots, n, \tag{1}$$

where $\{f_i(t)\}$ are observations of $X_i(t)$ with compositional noise $\{\gamma_i(t)\}$ and additive noise $\{e_i(t)\}$. Various approaches have been proposed to address this problem (Tang and Muller, 2008; Gervini and Gasser, 2004; James, 2007; Kneip and Gasser, 1992; Ramsay and Li, 1998). In particular, if $X_i(t) = g(t)$ (independent of index i), then the model is in a template-based form.

In this paper, we focus on observations with nonlinear phase variability and investigate the robust estimation in the following semi-parametric, template-based framework (Kurtek et al., 2011):

$$f_i(t) = c_i(g \circ \gamma_i)(t) + e_i, \quad t \in [0, 1], \quad i = 1, \ldots, n, \tag{2}$$

where $g(t)$ is the underlying template, $\{f_i(t)\}$ are observations of $g(t)$ with random time warpings $\{\gamma_i(t)\}$, scaling coefficients $\{c_i\} \in \mathbb{R}^+$, and vertical translations $\{e_i\} \in \mathbb{R}$. This model has scale and translation variability on amplitude. We will provide an algorithm to estimate g from $\{f_i\}$ with the generative model in Eq. (2) and prove the estimator is asymptotically consistent.

We will also extend the notion of boxplot to time warping functions. Recent work has introduced a boxplot definition for functional observations (Sun and Genton, 2011). However, as the time warping functions are constrained (see the method section for details), the so-defined boxplot cannot capture the degree of variability over the time domain or properly identify outliers in the data. We propose a coordinate-rotation-based method to address this issue. We will demonstrate the important mathematical properties in the new method and compare it with the previous functional boxplot.

The rest of the paper is organized as follows: In Section 2, we provide details of our framework on phase band depth median, compare it with the Karcher median, and then propose a robust estimation algorithm with consistency theory supported. The effectiveness of this new framework is demonstrated using simulations and real data sets in Section 3. In Section 4, we extend the depth-based framework to a definition of boxplot for time warping functions. Section 5 provides discussion and conclusion. Finally, the Appendices give all mathematical details throughout the paper.

2. Methods

In this section we will propose a new framework for estimating the functional template by using the notion of depth in phase variability. We will, at first, briefly review the background information on phase variability in functional data.

2.1. Basic principles on phase variability in functional data

Let Γ be the set of all orientation-preserving diffeomorphism on [0, 1]. That is, $\Gamma = \{\gamma : [0,1] \to [0,1] | \gamma(0) = 0,\ \gamma(1) = 1,\ 0 < \dot{\gamma} < \infty\}$. Each element in Γ is called a *time warping* function. Elements of Γ form a group with function composition and the identity in this group is the self-mapping $\gamma_{id}(t) = t$. The goal in our methodology is to find the optimal time warpings from the registration process and use those for our median computation.

In this paper, we adopt the well-known Fisher–Rao framework for phase variability estimation, which was shown to have superior registration performance over other methods (Srivastava et al., 2011). The Fisher–Rao registration separates functional data into phase and amplitude components. We are interested in using this method to obtain the phase component. This framework makes use of the square-root velocity function (SRVF) transformation which transforms the Fisher–Rao Riemannian metric into the standard $\mathbb{L}^2$ metric. Let $\mathcal{F}$ denote the set of all absolutely continuous functions on [0,1]. Then the SRVF of a function $f \in \mathcal{F}$ is $q : [0,1] \to \mathbb{R}$ defined by

$$q(t) = \text{sign}(\dot{f}(t))\sqrt{|\dot{f}(t)|} \in \mathbb{L}^2,$$

where the dot on the top of a function denotes differentiation. While the Fisher–Rao framework is quite extensive (Srivastava et al., 2011), we simply use it as a pre-processing step to obtain phase. One of the nice results of our methodology is that we only need pair-wise alignment to obtain our phase components. Here we present the pairwise alignment, or registration, in the Fisher–Rao framework in the following definition.

Definition 1. For any functions $f_1, f_2 \in \mathcal{F}$ with SRVFs $q_1, q_2 \in \mathbb{L}^2$, respectively, the **pairwise registration** of f_1 and f_2 is equivalent to finding the optimal time warping function

$$\gamma^* = \underset{\gamma \in \Gamma}{\text{argmin}} \|q_1 - (q_2 \circ \gamma)\sqrt{\dot{\gamma}}\|.$$

For registration over a set of functions, we can take one as a temporary template and then compute the optimal time warping from it for each function in the set. This can be computationally done with dynamic programming (Srivastava et al., 2011). Once this is done, we need to find the centrality of all time warping functions. We impose a functional depth to handle the task of finding center-outwards rankings for the optimal time warping functions. In particular, we impose functional band depth on the phase (Lopez-Pintado and Romo, 2009).

2.2. Framework for phase band depth

This subsection will study the band depth methodology in the functional space. We will also provide important mathematical properties inherited from the time warping space for band depth and demonstrate the robustness of the band depth median in the time warping space.

2.2.1. Notation of band depth in function space

Denote the set of all continuous functions on $I = [0,1]$ as $C(I)$. We start with the notation of the graph and band (Lopez-Pintado and Romo, 2009).

Definition 2. The **graph** of a function $x \in C(I)$ is the subset of the plane

$$G(x) = \{(t, x(t)) : t \in I\}.$$

Definition 3. For $x_1, \ldots, x_k \in C(I)$, the **band** of these functions is defined as

$$B(x_1, \ldots, x_k) = \{(t,y) : t \in I,\ \min_{i=1,\ldots,k} x_r(t) \le y \le \max_{i=1,\ldots,k} x_r(t)\}.$$

Based on the notation in Lopez-Pintado and Romo (2009), we use $BM(x_1, \ldots, x_n)$ and $MBM(x_1, \ldots, x_n)$ to denote the *sample band depth median* and the *sample modified band depth median*, respectively. We will focus our efforts in this paper on band depth only (although both depths will be used in Section 3). In this study, we propose to adopt the band depth on the time warping functions. We will show that such an approach will lead to a robust estimation for the underlying template.

2.2.2. Mathematical properties of band depth in time warping functions

We will provide a detailed explanation of how to use depth on the time warping functions. In pursuit of that goal, we will first introduce four important properties where the group action is taken into account. The proofs of these properties are shown in Appendix A.

The first property states the concept of the graph is immune to an arbitrary warping.

Proposition 1. *For any* $\gamma, \gamma_1, \ldots, \gamma_n, \tilde{\gamma} \in \Gamma$,

$$G(\gamma) \in B(\gamma_1, \ldots, \gamma_n) \Leftrightarrow G(\gamma \circ \tilde{\gamma}) \in B(\gamma_1 \circ \tilde{\gamma}, \ldots, \gamma_n \circ \tilde{\gamma}).$$

The second property states the concept of the graph is immune to inverse mappings.

Table 1
Median property comparison.

Method	Property
Band Depth Median	$BM\{\gamma_1 \circ \gamma, \ldots, \gamma_n \circ \gamma\} = BM\{\gamma_1, \ldots, \gamma_n\} \circ \gamma$
Band Depth Median	$BM\{\gamma_1^{-1} \cdots, \gamma_n^{-1}\} = [BM\{\gamma_1, \ldots, \gamma_n\}]^{-1}$
Karcher Median	$KM\{\gamma_1 \circ \gamma, \ldots, \gamma_n \circ \gamma\} = KM\{\gamma_1, \ldots, \gamma_n\} \circ \gamma$
Karcher Median	$KM\{\gamma_1^{-1} \cdots, \gamma_n^{-1}\} \neq [KM\{\gamma_1, \ldots, \gamma_n\}]^{-1}$

Proposition 2. *For any* $\gamma, \gamma_1, \ldots, \gamma_n \in \Gamma$,

$$G(\gamma) \in B(\gamma_1, \ldots, \gamma_n) \Leftrightarrow G(\gamma^{-1}) \in B(\gamma_1^{-1}, \ldots, \gamma_n^{-1}).$$

This next property states we can apply a time-warping function to our band median and account for full observational shifts via the same time-warping function.

Proposition 3 (*Composition Property*)**.** *The band depth median of shifted warping functions is the shifted band depth median of the warping functions, i.e., for any* $\gamma, \gamma_1, \ldots, \gamma_n \in \Gamma$,

$$BM\{\gamma_1 \circ \gamma, \ldots, \gamma_n \circ \gamma\} = BM\{\gamma_1, \ldots, \gamma_n\} \circ \gamma. \tag{3}$$

The last property states the band depth median computation and the inversion of time warping functions are commutative.

Proposition 4 (*Inverse Property*)**.** *The band depth median of inverse warping functions is the inverse of the band depth median of the warping functions, i.e., for any* $\gamma_1, \ldots, \gamma_n \in \Gamma$,

$$BM\{\gamma_1^{-1} \cdots, \gamma_n^{-1}\} = [BM\{\gamma_1, \ldots, \gamma_n\}]^{-1}. \tag{4}$$

These four properties play an integral role in the formation of our framework and the theoretical results therein. They will be utilized in Section 2.3, where we will discuss the estimation algorithm and the consistency of the estimator. Our next step, however, is to elucidate why we chose the band depth median over the Karcher median (Kurtek et al., 2013).

2.2.3. Robustness of band depth median on phase

The Karcher median was recently developed (Kurtek et al., 2013) and is a metric-based measure of centrality for functional data using a very similar methodology as the Karcher mean. The Karcher median of a set of warping functions $\gamma_1, \ldots, \gamma_n$ is defined as:

$$KM(\gamma_1, \ldots, \gamma_n) = \underset{\gamma \in \Gamma}{\operatorname{argmin}} \sum_{i=1}^{n} \cos^{-1}\left(\int_0^1 \sqrt{\dot{\gamma}_i(t)}\sqrt{\dot{\gamma}(t)}dt\right). \tag{5}$$

While we will integrate the Fisher–Rao methodology into our framework, we chose to adopt band depth as our tool for finding centrality. The mathematical properties are pivotal for our signal estimation consistency theory and to the comparison of the Karcher median to the band depth median. Table 1 gives a quick breakdown of which properties hold true for each methodology.

We can see from Table 1 the composition property holds true for both methods, whereas the inverse property only holds true for the band depth median. This is because in general $d_{FR}(\gamma_1, \gamma_2) \neq d_{FR}(\gamma_1^{-1}, \gamma_2^{-1})$, where $d_{FR}(\cdot, \cdot)$ denotes the Fisher–Rao distance between two time warping functions. This inverse property ends up being critical in our theory of consistency and is a key reason why we chose the band depth median methodology over the Karcher median.

In addition to comparison on mathematical properties, we compare the robustness with respect to the outliers in the time warping space. For illustration, we generate 11 time warping functions as follows: Let $s \in [-3, 3]$ and define a function $h_i(s) : [-3, 3] \to [-3, 3]$ by:

$$h_i(s) = \begin{cases} 6\dfrac{\exp[(b_{[i]}a_i)(s+3)/6] - 1}{\exp[(b_{[i]}a_i)] - 1} - 3 & a_i \neq 0, \\ s & a_i = 0 \end{cases}$$

where the 11 coefficients a_i are equally spaced between -1 and 1, and the term $b_{[i]}$ refers to the ordering of $b_i \sim \text{Exp}(1.5)$, i.e. $b_{[1]} \geq b_{[2]} \geq \cdots \geq b_{[11]}$. Let the function $T(\cdot) = (\cdot + 3)/6$. Then $\gamma_i(t) = T(h_i(T^{-1}(t))) : [0, 1] \to [0, 1]$ is a time warping function in Γ, $i = 1, \ldots, 11$. The three outlier functions were generated through:

$$\gamma_{i+11}(t) = \arctan((2t - 1)\beta_i)/(2\arctan(\beta_i)) + 0.5,$$

where $\beta_i = 3 + 2i$, $i = 1, 2, 3$.

In Fig. 2A & C, all 14 warping functions are shown. In Fig. 2B & D, only the first 11 warping functions are shown. We superimpose the band depth median or Karcher median on the given functions in each case. It is apparent the band depth

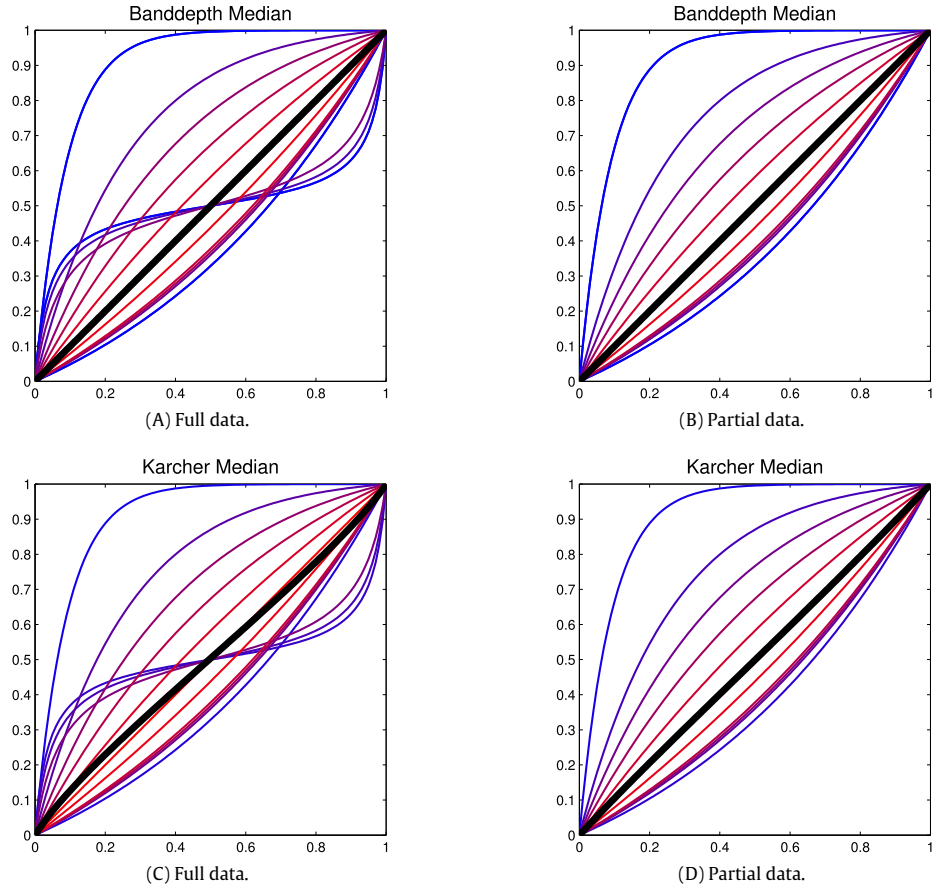


Fig. 2. Median methods and their robustness with respect to outliers. The median function using each method is displayed in the panel via the bold black line. **A.** 14 time warping functions with outliers present (band depth median superimposed). **B.** 11 time warping functions with 3 outliers being removed (band depth median superimposed). **C.** Same as Panel A except using Karcher median. **D.** Same as Panel B except using Karcher median.

median is robust with respect to the 3 outliers in the data (identity function in both cases), whereas the Karcher median varies when the three outliers are removed from the data. More significant results can be obtained when more drastic outliers are used in comparison.

2.3. Robust template estimation and large sample theory

Next we construct our signal estimation and provide illustrations of its robustness.

2.3.1. Estimation algorithm

As stated in Introduction, the generative model of the given functional data is assumed to have the following semi-parametric form:

$$f_i(t) = c_i(g \circ \gamma_i)(t) + e_i, \quad t \in [0, 1], \quad i = 1, \ldots, n, \tag{6}$$

where g is the underlying template, and $\{f_i\}$ are the observations of g with random warpings $\{\gamma_i\} \in \Gamma$, scaling coefficients $\{c_i\} \in \mathbb{R}^+$, and vertical translation $e_i \in \mathbb{R}$. In this study, we assume the population band depth median of $\{\gamma_i\}$ is the identity γ_{id}.

One advantage in our approach is that we only need one round of pairwise alignment. We use the time warping space and band depth to find the median template in phase space and relate that back to the original functional observation space. The following algorithm outlines the procedure, where the output function $\hat{g}$ is the estimated template.

Algorithm 1. Given a set of observed functions f_i, $i = 1, \ldots, n$:

1. Compute the SRVFs q_i of f_i, $i = 1, \ldots, n$.
2. Choose any q_i as the alignment objective template function m.

3. Estimate $\gamma_i^* = \text{argmin}_{\gamma\in\Gamma}\|m - (q_i, \gamma)\|$.
4. Compute $\bar{\gamma}_n = BM(\gamma_1^*, \ldots, \gamma_n^*)$.
5. Return estimated warping functions, aligned functions, and the estimator: $\tilde{\gamma}_i = (\gamma_i^* \circ \bar{\gamma}_n^{-1})^{-1}, \tilde{f}_i = f_i \circ \gamma_i^* \circ \bar{\gamma}_n^{-1}$, and $\hat{g} = \frac{1}{n}\sum_{i=1}^n \tilde{f}_i$.

In practice, the observed function data may not exactly follow the model in Eq. (6). In this case, there will be no underlying true function g. To estimate a phase-based template in the given data, we slightly modify Algorithm 1 by combining the Fisher–Rao alignment (Srivastava et al., 2011) for the initial selection of objective template. This new algorithm (referred to as Algorithm 2) is given as follows, where the output function f_j is the estimated template whose time warping is the median in the alignment.

Algorithm 2. Given a set of observed functions f_i, $i = 1, \ldots, n$:

1. Compute the SRVFs q_i of f_i, $i = 1, \ldots, n$.
2. Apply the Complete Alignment Algorithm in Srivastava et al. (2011) on SRVFs $\{q_i\}$. Denote the output template as μ_n.
3. Choose μ_n as the alignment objective template function m.
4. Estimate $\gamma_i^* = \text{argmin}_{\gamma\in\Gamma}\|m - (q_i, \gamma)\|$.
5. Compute $\bar{\gamma}_n = BM(\gamma_1^*, \ldots, \gamma_n^*)$.
6. The estimated warping functions and aligned functions are given as: $\tilde{\gamma}_i = (\gamma_i^* \circ \bar{\gamma}_n^{-1})^{-1}, \tilde{f}_i = f_i \circ \gamma_i^* \circ \bar{\gamma}_n^{-1}, i = 1, \ldots, n$.
7. Let $\tilde{\gamma}_j, j \in \{1, \ldots, n\}$ denote the median warping function in the set $\{\tilde{\gamma}_i\}$, then $\tilde{f}_j$ is the phase-based template in the original set.

2.3.2. Illustration of the estimation

Here we present one example to illustrate this estimation process. We start by letting the underlying signal g be an enveloped sine-wave function (shown in Fig. 3A):

$$g(t) = 4\cos(5\pi t)(\frac{1}{4} - (t - \frac{1}{2})^2), \qquad t \in [0, 1].$$

Then, we generate 24 observation functions by Eq. (6) using random scaling coefficients c_i, warping functions $\gamma_i(t)$, and translation shifts $e_i(t)$. In particular, each scaling coefficient follows the exponential distribution with mean 1. Each translation shift is generated from a normal distribution with mean 0 and variance 1. Finally, 24 warping functions are generated with median γ_{id} as follows:

$$\gamma_i(t) = t^{\alpha_i},$$

where $\alpha_1 = 4$, $\alpha_2 = 3$, $\alpha_i = 1.65 - 0.05(i - 3)$ for $i = 3, \ldots, 10$, and $\alpha_i = 1/(1 + 0.05(i - 11))$, for $i = 11, \ldots, 21$. The last three warping functions are generated as outliers:

$$\gamma_{i+21}(t) = \arctan((2t - 1)\beta_i)/(2\arctan(\beta_i)) + 0.5,$$

where $\beta_i = 3 + 2i$, $i = 1, 2, 3$.

The true warping functions $\{\gamma_i\}$, the observations $\{f_i\}$, and the aligned functions $\{\tilde{f}_i\}$ using the Estimation Algorithm are shown in Fig. 3B–D, respectively. In this case, $E(\bar{c}) = 1$. The estimated warping functions $\{\tilde{\gamma}_i\}$ are shown in Panel E. We see the estimated warping functions are very close to true ones in Panel B. The estimated template $\hat{g} = \frac{1}{n}\sum_{i=1}^n \tilde{f}_i$ and the true underlying template are shown in Panel F. This estimate is highly accurate despite large variability in both phase and amplitude in the observations. For comparison, we also show the estimation results using the Karcher mean and Karcher median methods on the same data set in Panel G and Panel H, respectively. It is apparent that both methods fail to capture the phase variability and do not accurately reconstruct the underlying template.

We also want to present the robustness of our signal estimation algorithm with respect to outliers. We follow up Fig. 3 with the cases of only two potential outliers or zero outliers present in the data set. Based on the estimation result shown in Fig. 4, we can see our signal estimation is hardly affected by the removal of outliers.

2.3.3. Large sample theory

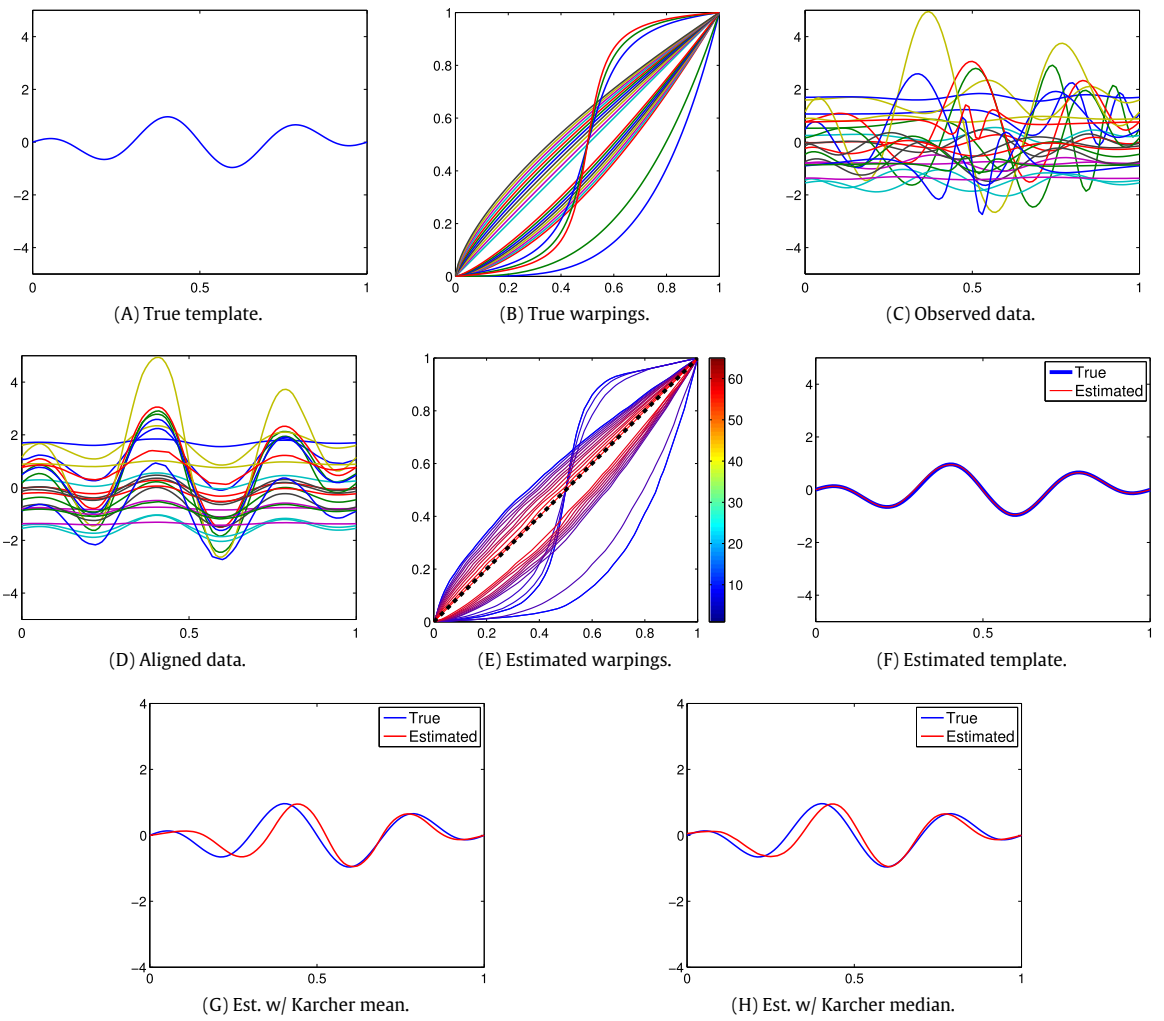
Theorem 1. *Assume independent observations $\{f_i\}$ follow a semi-parametric model:*

$$f_i(t) = c_i(g \circ \gamma_i)(t) + e_i, \quad i = 1, 2, \ldots, n,$$

where g is the underlying template, and independent random warpings $\{\gamma_i\} \in \Gamma$, independent scaling coefficients $\{c_i\} \in \mathbb{R}^+$, and independent vertical translation $e_i \in \mathbb{R}$. Let $\{\tilde{f}_i\}$ be the aligned functions in Algorithm 1. If (1) the population band depth median of $\{\gamma_i\}$ is γ_{id} (a.s.), and (2) $E(c_1)$ and $E(e_1)$ exist and are finite, then

$$\frac{1}{n}\sum_{i=1}^n \tilde{f}_i \to E(c_1)g + E(e_1). \ \ a.s.$$

The proof of this theorem is based on the four properties in Section 2.2. See details in Appendix B.



(A) True template. (B) True warpings. (C) Observed data.

(D) Aligned data. (E) Estimated warpings. (F) Estimated template.

(G) Est. w/ Karcher mean. (H) Est. w/ Karcher median.

Fig. 3. A. True underlying template. **B.** Time warping functions which represent compositional noises **C.** Observed functions after random scaling, translation, and compositional noise applied to the true template in Panel A. **D.** Aligned functions by Algorithm 1. **E.** Estimated time warping functions. The colorbar indicates the depth of each warping function and the thick dashed line is the band depth median function. **F.** Estimated template (red) with true template (blue) superimposed. **G.** and **H.** Same as Panel F except that the estimate is obtained using the Karcher mean and Karcher median methods, respectively.

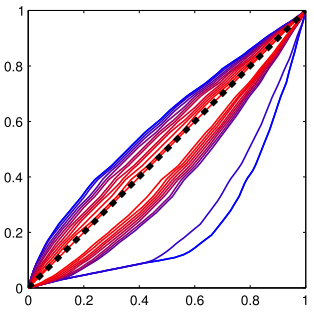
3. Results

In this section, we discuss the results of the proposed template estimation methodology. We start off with two simulations and then move onto two real data demonstrations. These data do not exactly follow the model in Eq. (6) and their templates will be estimated using Algorithm 2.

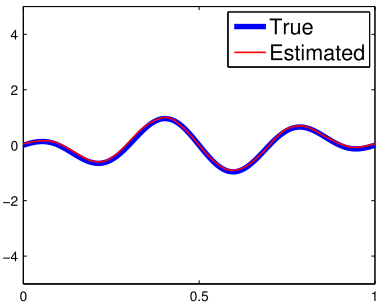
3.1. Simulations

3.1.1. Example 1: two-peak data

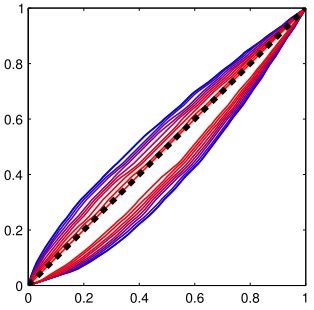
In our first simulation study, we presented a bimodal set of functional data. The functions are generated by: $u_i(t) = z_{i,1}e^{-(10t-6.5)^2/2} + z_{i,2}e^{-(10t-3.5)^2/2}$, $t \in [0, 1]$, $i = 1, 2, \ldots, 21$, where $z_{i,1}$ and $z_{i,2}$ are *i.i.d* normal with mean 1 and standard deviation 0.25. Each function is then warped according to: $\gamma_i(t) = \frac{e^{a_i t}-1}{e^{a_i}-1}$ if $a_i \neq 0$, otherwise $\gamma_i = \gamma_{id}$. Here a_i are equally spaced between -1 and 1 and the observed functions are computed using $x_i(t) = u_i(\gamma_i(t))$. The functions are sampled with 101 uniform time points in [0, 1] for computational purposes. A set of 21 such functions composes the original data and is shown in Fig. 5A. The data has two apparent peaks with a small dip in the middle of each function.



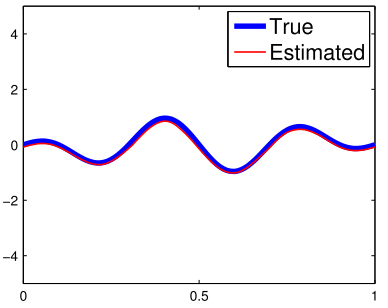
(A) Estimated warpings with three outliers removed.



(B) True and estimated signal.



(C) Estimated warpings with all outliers removed.



(D) True and estimated signal.

Fig. 4. Signal estimation with respect to outliers. **A.** Estimated time warping functions with three of the prominent outliers removed. The line properties are the same as those in Fig. 3E. **B.** Estimated template (red) with true template (blue) superimposed. **C.** Same as Panel A but with all five outliers removed. **D.** Same as Panel B but with Panel C's data set.

We wish to discuss the difference between the regular amplitude band depth approach (and modified band depth approach) and our phase band depth approach (and phase modified band depth approach). This comes down to visually comparing Panel B with E (and Panel C with F) in Fig. 5. One might want to believe the dashed line median template in Fig. 5B might be preferable to Fig. 5E but the reader should take note of the color bar associated with each approach, which represents the depth score (centrality) of each function. In Fig. 5B, most functions have low depth values of roughly 0.15 while the functions in Fig. 5E have more distributed depth values and reach up to 0.72. The measure for centrality in our method far surpasses the regular band depth approach for this set of data.

Next, let us discuss the difference between the modified band depth and our phase modified band depth approach. It might appear these two are quite close and therefore the reader might want to take the approach of modified band depth. However, we can still show our process yields higher measures of centrality. At first glance, Fig. 5C and F are quite similar in depth scores due to identical color bars but note that there are only a couple of functions with depth scores around 0.7 in Fig. 5C while there are about six or seven functions with depth scores around 0.7 in Fig. 5F. In a similar vein of argument as with band depth versus phase band depth, the measures of centrality in our method surpasses the regular modified band depth approach.

3.1.2. Example 2: data with an outlier

In the second simulation study, we present a set of Gaussian functional data with one outlier. The underlying function used to generate the data is: $x(t) = \frac{1}{\sqrt{2\sigma^2\pi}} e^{-\frac{(t-\mu)^2}{2\sigma^2}}$, $t \in [0, 1]$, $\mu = 0.5$, and $\sigma = 0.1$. Then we generate 26 observation functions by Eq. (6) using random scaling coefficients c_i, warping functions $\gamma_i(t)$, and translation shifts $e_i(t)$. In particular, each scaling coefficient follows the exponential distribution with mean 1. Each translation shift is generated from a normal distribution with mean 0 and standard deviation 0.07. Then warping functions are generated with population mean γ_{id} in the following manner:

$$\gamma_i(t) = t^{\alpha_i},$$

where $\alpha_i = 1.47 - 0.035(i-1)$ for $i = 1, \ldots, 13$, and $\alpha_i = 1/(1 + 0.035(i-13))$, for $i = 14, \ldots, 26$.

It is important to note there is an obvious outlier in Fig. 6A. The last observation function, the outlier, follows the same set up above except the peak is inverted in a very specific manner. This specific manner is found through the algebraic

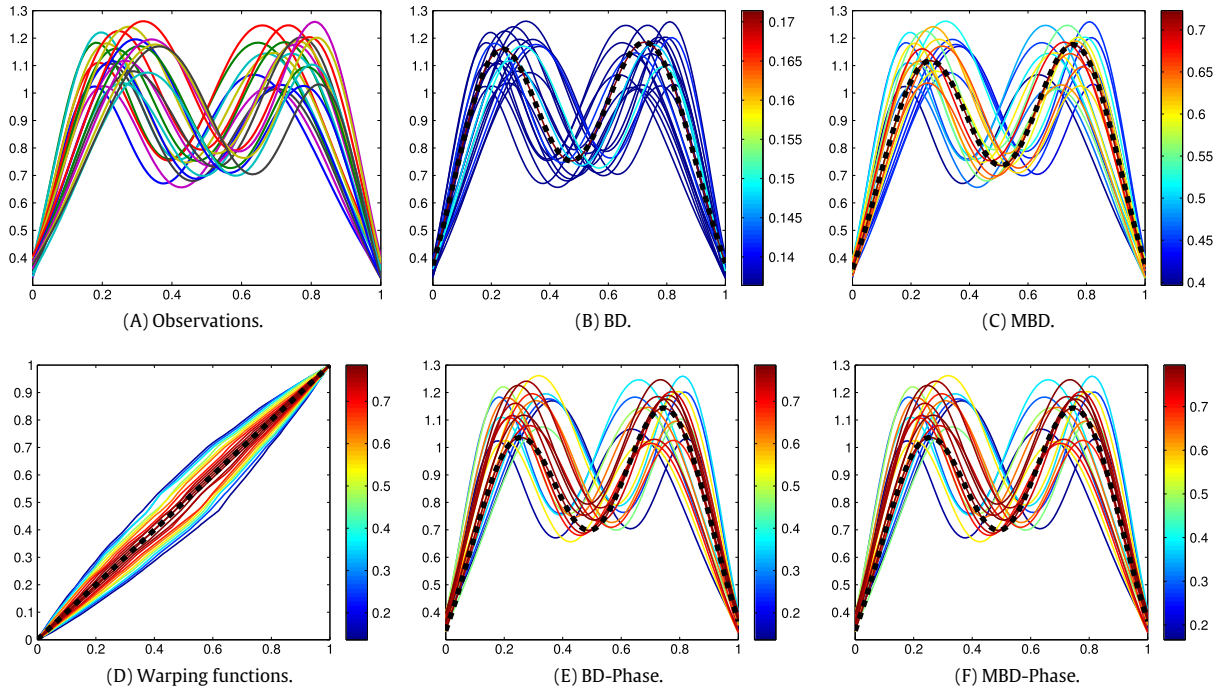


Fig. 5. Simulation One. **A.** Observed data. **B.** Functions color-labeled with traditional amplitude-based band depth (BD). Here the color bar indicates the depth of each function and the thick dashed line is the median function. **C.** Same as Panel B except the functions are color-labeled with amplitude-based modified band depth (MBD). **D.** Estimated warping functions for the alignment of observed data. The color bar indicates the depth of each warping function using MBD and the thick dashed line is the median function. **E.** Same as Panel B except using Phase-based BD. Here the thick dashed line is the function whose time warping is the median in the alignment. **F.** Same as Panel E except that the functions are color-labeled with MBD-Phase.

manipulation:

$$x_{27}(t) = \begin{cases} 176.12(t-0.5)^2 + 1.32 & t \in [0.45, 0.55], \\ \dfrac{1}{\sqrt{2\sigma^2\pi}} e^{-\frac{(t-\mu)^2}{2\sigma^2}}, \mu = 0.5, \sigma = 0.1 & t \in [0, 0.45) \cup (0.55, 1]. \end{cases}$$

We will, again, discuss the differences between our approaches and regular band depth approaches. In Fig. 6B the color bar goes no higher than the value of roughly 0.156 while the color bar in Fig. 6E gets close to 0.73. The measure for centrality in our method far surpasses the regular band depth approach for this set of data since the regular band depth approach has depths in the range of 0.11–0.155 while our approach ranges from 0.2 to 0.73.

Fig. 6C and F are quite similar in depth scores due to identical color bars. However, it is worth noting the median template picked in Fig. 6C is the outlier we designed. We note in our method, displayed in Fig. 6D, the outlier curve does not end up being the deepest function which is why we get a more central template in Fig. 6F. In a similar vein of argument as with band depth versus phase band depth, the measures of centrality in our method surpass the regular modified band depth approach and can detect non-typical outliers potentially in the middle of a set of data.

3.2. Real data

In this section, we look at two real world data sets and analyze them with our methodology.

3.2.1. Berkeley growth data

The first real data set we want to explore is the Berkeley growth data, in particular the growth velocity of 39 males. This data set is derived from an original, discrete, set of height observations taken over the first eighteen years (Ramsay and Silverman, 2005). The data is shown as 39 curves in Fig. 7A.

Similar to the simulations, we can rank the data with traditional (or phase-based) BD and MBD depth values. We can see from Fig. 7B that all observations have the same BD score. This indicates the traditional BD scores do not rank the data appropriately (we can also note this because no median template was picked for display since all of them have the same depth score). Clearly the phase-based BD in Fig. 7E outperforms band depth since a median template is actually chosen due to varying scores of center instead of identical scores.

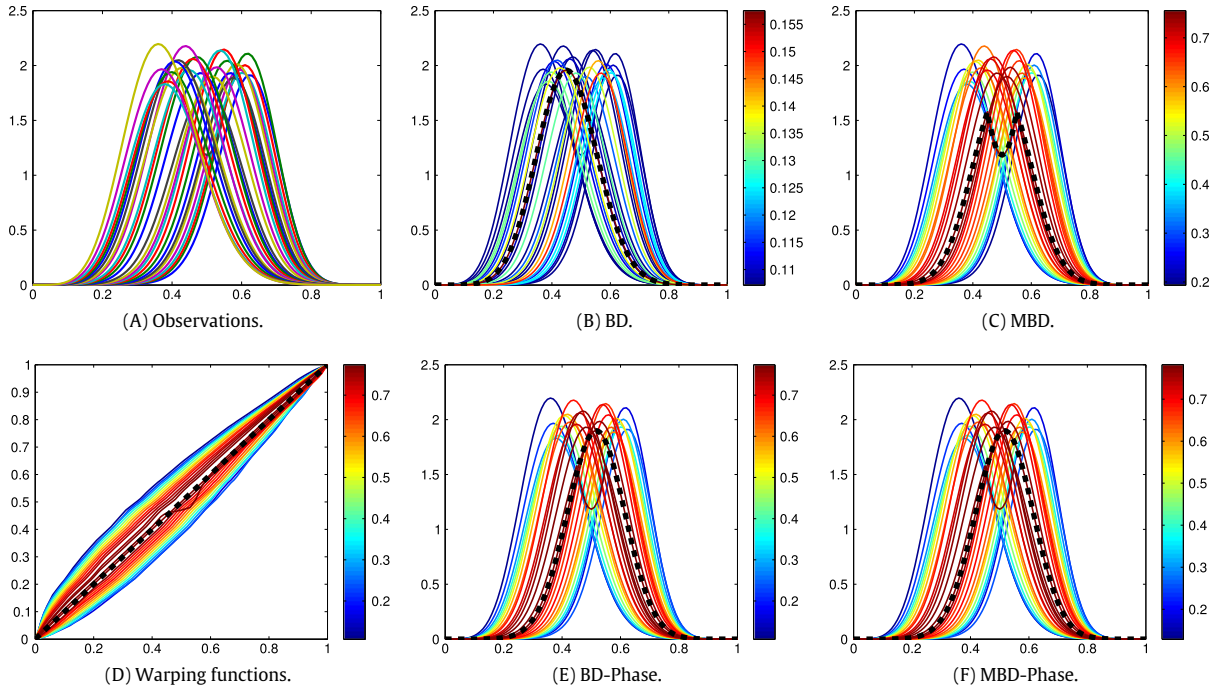


Fig. 6. Simulation Two. All line properties and colorbars are the same as those in Fig. 5.

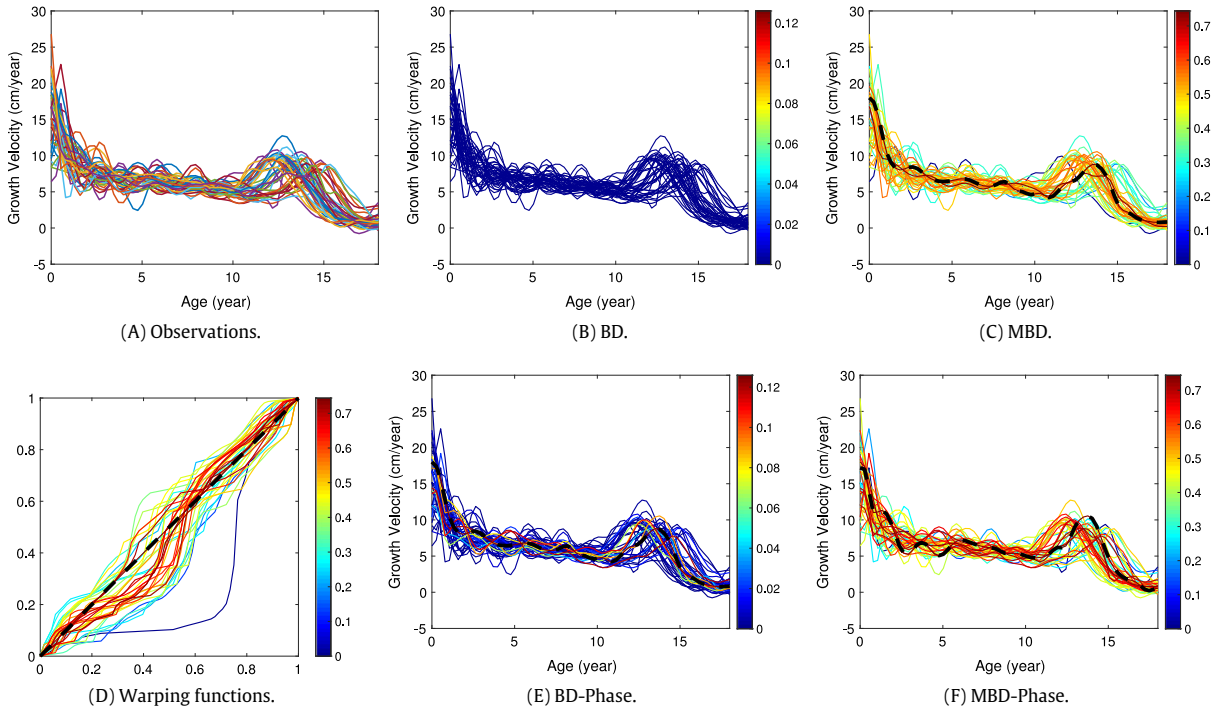


Fig. 7. Berkeley growth data. All line properties and colorbars are the same as those in Fig. 5.

We further look at results using MBD and phase-based MBD. In Fig. 7D, we are given the estimated time warping functions that will be used to measure depth of this data set. In Fig. 7C (for MBD) and Fig. 7F (for phase-based MBD), we notice the

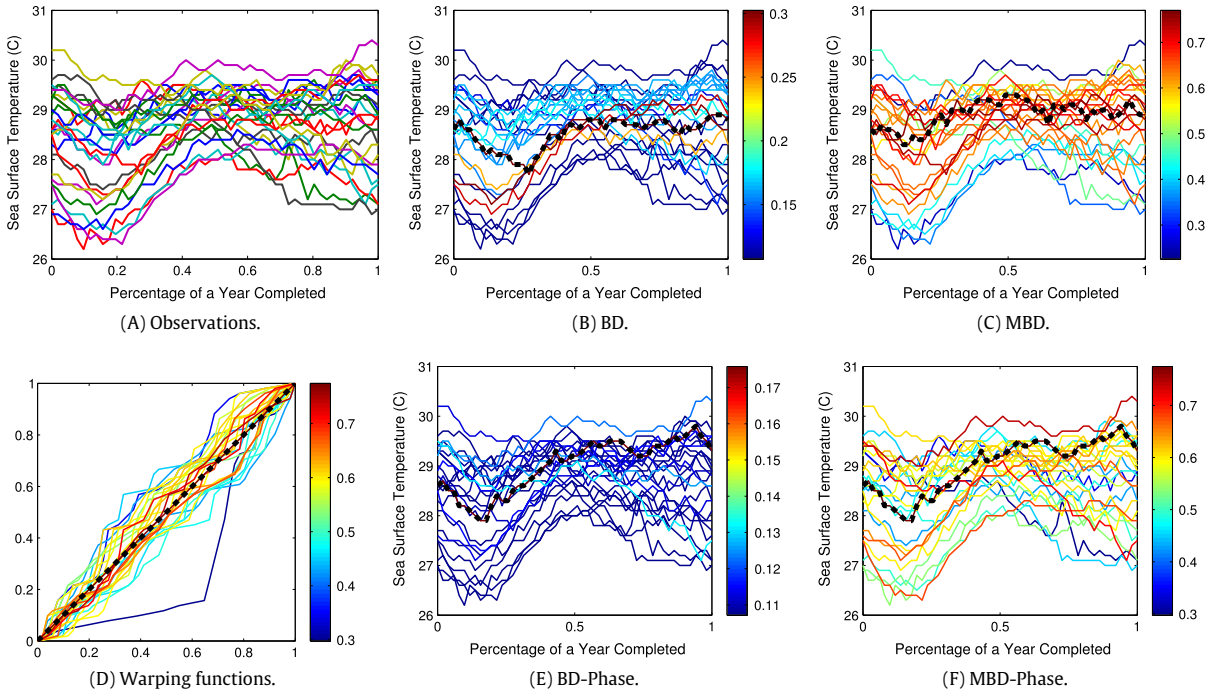


Fig. 8. El Nino data. All line properties and colorbars are the same as those in Fig. 5.

identical color bars but many more 0.5–0.7 depth score observations in Fig. 7F than in Fig. 7C. In addition, the median in Panel F has more variability during 2–7 years old, which better represents typical pattern in the given observations than that in Panel C. This confirms the measures of centrality in our phase-based method surpasses the regular modified band depth approach for the Berkeley growth data.

3.2.2. El Nino data

The second real data set we want to examine is El Nino data, in particular the subset of data labeled region 4 which covers Australia to the western Pacific about the equator (i.e. latitude 5N–5S and longitude 150W–160E). This data set was obtained from an original, discrete, set of sea surface temperature observations taken every week of the year during the years 1989 to 2016. Thus we have twenty-seven functions with fifty-two discrete time points. The data was obtained from NOAA (National Oceanic and Atmospheric Administration).

The sea surface temperature curves over twenty-seven years are shown in Fig. 8A. We can easily see from Fig. 8B and E band depth scores for both the original space and the phase space are low but Fig. 8B has slightly higher band depth scores. Upon investigating the data, we can notice the general trend of the sea surface temperature is to dip in the first 15% to 20% of the year and then gradually spike back upwards until the end of the year. The pattern in Fig. 8B has a dip but way more gradual of a decline which displays very different pattern from most of the curves.

We further investigate the data with modified band depth. Recall that Fig. 8A will be the representative space for Panels B and C while Fig. 8D will be the representative space for the remaining panels (E and F). We can see from Fig. 8C and F both have similar depth scores. In Fig. 8C and F, we notice quite similar color bars but many more 0.6–0.65 depth score observations in Fig. 8C than in Fig. 8F. However, in Fig. 8C the dipping then rising structure is lost and replaced with a balancing act of staying consistently around twenty-eight and twenty-nine degrees Celsius. In Fig. 8F we still observe the same pattern as in Fig. 8E and one consistent to the general pattern.

4. Extension to boxplot

In this section, we discuss the recent extension of the boxplot from the multivariate model to functional data. We discuss some of the drawbacks for this boxplot extension with respect to the time warping space Γ and then propose a new approach to deal with the problem. The section is organized as follows: Section 4.1 covers the conventional boxplot model in the functional space, Section 4.2 covers our approach to the boxplot model in the time warping space Γ, and Section 4.3 compares these two different approaches.

4.1. Traditional functional boxplot

An extension from traditional multivariate boxplot to the functional boxplot is not new (Sun and Genton, 2011). The extension uses the band depth framework and approaches functional data in a similar manner as classical statistics. Basically, we take the middle fifty percent of the data, create a box from that middle proportion, compute the inter-quartile range (IQR), and use a detection criteria based off the IQR to create whiskers (finally detect outliers). We use the band depth to find the 50% central region from the sample. This 50% central region from the sample can be found by

$$C_{0.5} = \{(t, y(t))|_{\min_{r=1,\ldots,[n/2]} y_{(r)}(t) \le y(t) \le \max_{r=1,\ldots,[n/2]} y_{(r)}(t)}\},$$

where $[r]$ is the ordering based off the band depth and $[n/2]$ is the smallest integer not less than $n/2$. The next step is to find the inter-quartile range at every time point in the domain, i.e.

$$IQR(t) = \max_{r=1,\ldots,[n/2]} y_{(r)}(t) - \min_{r=1,\ldots,[n/2]} y_{(r)}(t).$$

In pursuit of outlier detection, whiskers are generated. The difference between classical whiskers and these whiskers are that these whiskers are band-based and outlier detection is based on staying within the band created by the whiskers. The whiskers are computed in the following manner:

$$W = \{(t, y(t))|_{\min_{r=1,\ldots,[n/2]} y_{(r)}(t) - \mathcal{C}(IQR(t)) \le y(t) \le \max_{r=1,\ldots,[n/2]} y_{(r)}(t) + \mathcal{C}(IQR(t))}\}.$$

While the traditional outlier detection relies on the whiskers being extended outwards by $\mathcal{C} = 1.5$ times the IQR, there is nothing in the functional boxplot model that states this coefficient $\mathcal{C}$ is set in stone. We stated this type of boxplot is not appropriate for the set of time warping functions. This will be illustrated in Section 4.3 when we compare this conventional methodology to our new approach.

4.2. Boxplot on the phase variability

Before the comparison, we provide details of our approach, which is for functions in the time warping space. Quite simply, we utilize the band in the same manner as above but we project our warping functions to a new space, compute the inter-quartile range (IQR), compute the whiskers, and then project the results back to our original time warping space. It is important to remember the time warping space is given by

$$\Gamma = \{\gamma : [0, 1] \to [0, 1] | \gamma(0) = 0, \gamma(1) = 1, \dot{\gamma} > 0\}.$$

Our goal is to find a projection that will allow us to rotate the graph of our time warping functions in a clockwise manner by $\frac{\pi}{4}$. That is, we have another space

$$\Lambda = \{\lambda : [0, \sqrt{2}] \to \mathbb{R} | \lambda(0) = 0, \lambda(\sqrt{2}) = 0, \dot{\lambda} \text{ exists}\}.$$

We define a mapping $P : \Gamma \to \Lambda$, $P(\gamma) = \lambda$ by the rotation matrix shifting the graph down $\frac{\pi}{4}$. That is, for any $t \in [0, 1]$, we have

$$\begin{pmatrix} s \\ \lambda(s) \end{pmatrix} := \begin{pmatrix} \cos(\pi/4) & \sin(\pi/4) \\ -\sin(\pi/4) & \cos(\pi/4) \end{pmatrix} \begin{pmatrix} t \\ \gamma(t) \end{pmatrix} = \begin{pmatrix} (\gamma(t) + t)/\sqrt{2} \\ (\gamma(t) - t)/\sqrt{2} \end{pmatrix}.$$

Therefore, for any $s \in [0, \sqrt{2}]$,

$$\lambda(s) = \frac{1}{\sqrt{2}}(\gamma(\tilde{\gamma}^{-1}(s)) - \tilde{\gamma}^{-1}(s)),$$

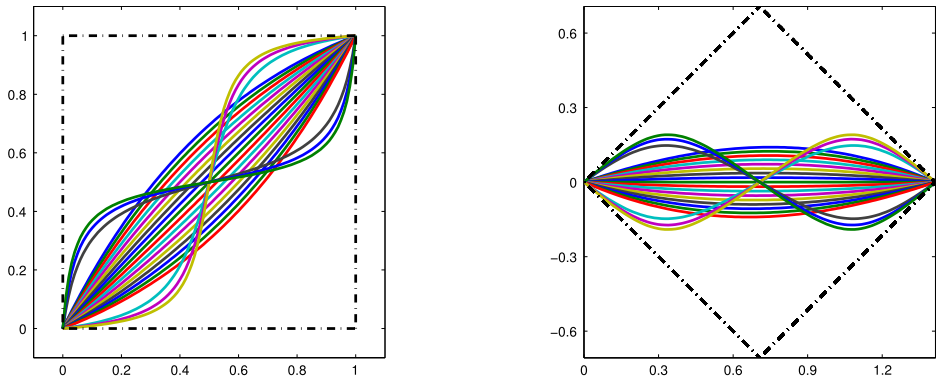
where $\tilde{\gamma}(t) = \frac{1}{\sqrt{2}}(\gamma(t) + t)$ is a strictly increasing function from $[0, 1]$ to $[0, \sqrt{2}]$ (so its inverse exists). We present Fig. 9 to demonstrate the graphical effects of our mapping.

The 23 time warping functions in Fig. 9A were generated in a similar manner to those in Fig. 1. The first 17 time warping functions are generated as follows. Let $s \in [-3, 3]$ and define a function $h_i(s) : [-3, 3] \to [-3, 3]$ by:

$$h_i(s) = \begin{cases} 6\dfrac{\exp[(0.55a_i)(s+3)/6] - 1}{\exp[(0.55a_i)] - 1} - 3 & a_i \neq 0, \\ s & a_i = 0 \end{cases}$$

where the coefficients a_i are equally spaced between -1 and 1. Recall the linear transformation $T(\cdot) = (\cdot + 3)/6$, $T(s) : [-3, 3] \to [0, 1]$ is mapping into $t \in [0, 1]$, and hence $T \circ h_i \circ T^{-1} \in \Gamma$. The outliers, the last six functions, were generated through:

$$\gamma_{i+17}(t) = \arctan((2t - 1)\beta_i)/(2\arctan(\beta_i)) + 0.5,$$



(A) Original observations. (B) Mapped observations.

Fig. 9. **A** The original observations in the space Γ. **B** The mapped observations in the space Λ.

where $\beta_i = 3 + 2i$, $i = 1, 2, 3$ and

$$\gamma_{i+20}(t) = \gamma_{i+17}^{-1}(t)$$

with $i = 1, 2, 3$.

Our goal is to use this projection to find a symmetric around the median template band for outlier detection. In pursuit of that goal, we need to show that projection into this new space does not affect our time warping functions. This can be proven with the following two lemmas and one theorem. The detailed proofs are given in Appendix C.

Our first lemma demonstrates the rotation mapping preserves comparative functional magnitude.

Lemma 1. *For $\gamma_1, \gamma_2 \in \Gamma$, if $\gamma_1(t) \leq \gamma_2(t)$ for any $t \in [0, 1]$, then $P(\gamma_1)(s) \leq P(\gamma_2)(s)$ for any $s \in [0, \sqrt{2}]$.*

This next lemma informs us two important ranking functions, minimum and maximum of a collection of observations, are preserved during the rotation mapping.

Lemma 2. *For $\gamma_1, \ldots, \gamma_n \in \Gamma$, if $\gamma_{\min} = min_{i=1,\ldots,n}\gamma_i$, then $P(\gamma_{\min}) = min_{i=1,\ldots,n}P(\gamma_i)$. Analogously, if $\gamma_{\max} = max_{i=1,\ldots,n}\gamma_i$, then $P(\gamma_{\max}) = max_{i=1,\ldots,n}P(\gamma_i)$.*

It is important to note that the band depth is pivotal to the construction of the functional boxplot. This next theorem states the rotation mapping and the concept of the functional boxplot can be combined without losses, i.e. the concept of the band is invariant to our rotation mapping.

Theorem 2. *For $\gamma_1, \ldots, \gamma_n \in \Gamma$, if $G(\gamma) \subset B(\gamma_1, \ldots, \gamma_n)$, then $G(P(\gamma)) \subset B(P(\gamma_1), \ldots, P(\gamma_n))$.*

Since we have invariance of the band under our rotation mapping, our next task is to build the framework for the boxplot in the time warping setting. The basic outline is: put the time warping functions through rotation, apply the traditional functional boxplot framework to these rotated functions, and return relevant information. The actual process is provided below:

Construct the boxplot of warping functions $\{\gamma_1, \ldots, \gamma_n\}$

1. Rotate the space using the mapping P, which results in n functions $P(\gamma_1), \ldots, P(\gamma_n) \in \Lambda$.
2. Compute the 50% central region and whiskers on $\{P(\gamma_i)\}$ using traditional functional boxplot.
3. Rotate the space back using P^{-1}.
4. Return the rotated 50% central region and whiskers and display them with the original $\{\gamma_i\}$.

We point out that the whisker bands in the proposed boxplot may not be a warping function. By construction, the whisker functions in the rotated (clockwise) coordinate must be in the space Λ. When we map it back (rotate counterclockwise) to the time wrapping space, these whiskers may not even be a function as one x could correspond to multiple y's. That is, the boxplot is actually built in the rotated space, and we merely rotate it back in the original space for display purposes. This display will help identify the range for possible outliers, whereas the identification is actually conducted in the rotated space.

A related boxplot study for functional data with phase variability was recently developed by Xie and colleagues in Xie et al. (2017). They proposed a metric based method to independently detect outliers for amplitude and phase components of elastic functional data. They rank the distances between each component and its geometric median, and then use them to introduce the concept of boxplot visualization for functional data.

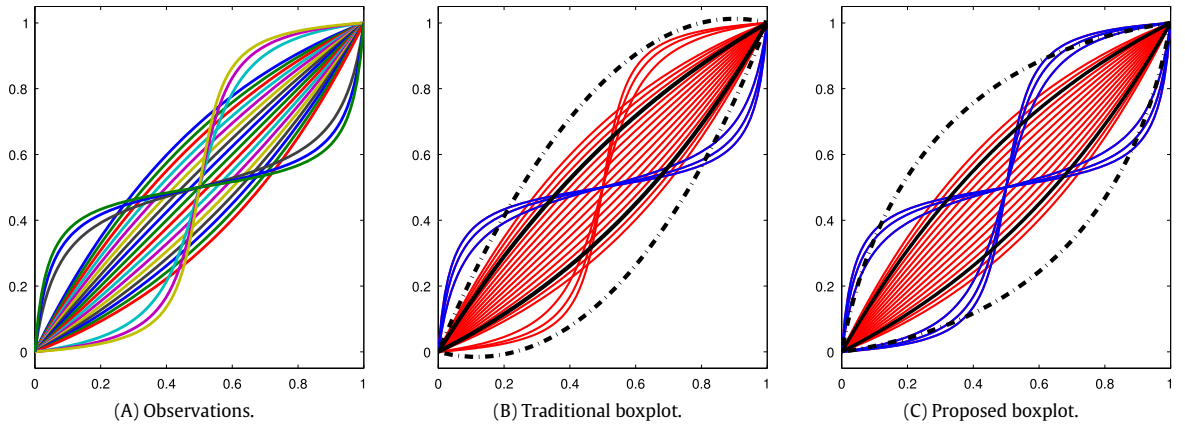


Fig. 10. **A** Functional observations given in Fig. 9A. **B** The functional observations with the traditional boxplot methodology applied with 0.7 times the IQR for whiskers (black, dashed lines). The two black, solid lines are the 50% central region. The detected 3 outliers are in blue. **C** Same as Panel B except using the proposed boxplot. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

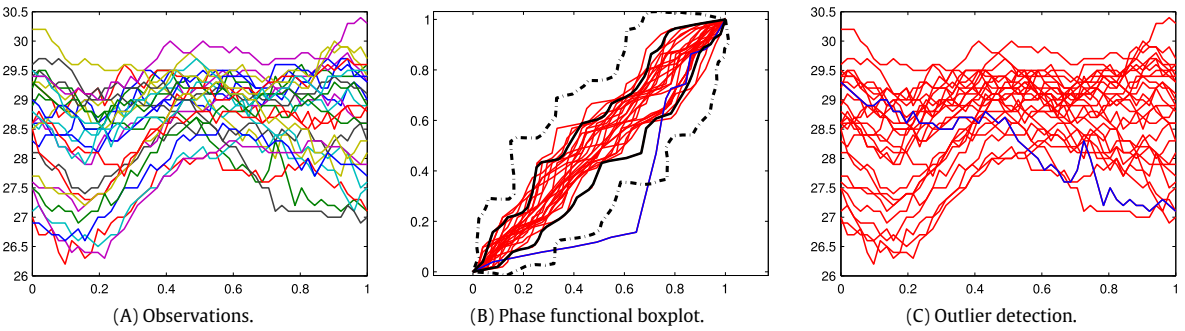


Fig. 11. El Nino data set. **A** The original functional observations given in Fig. 8A. **B** The red curves denote all time warping functions for alignment on the observed data. The fifty percent central region (black solid lines), the whiskers (black, dashed lines), and any observations that jump outside the whiskers (solid, blue lines). **C** The detected outlier (blue) back in the original functional observation space. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

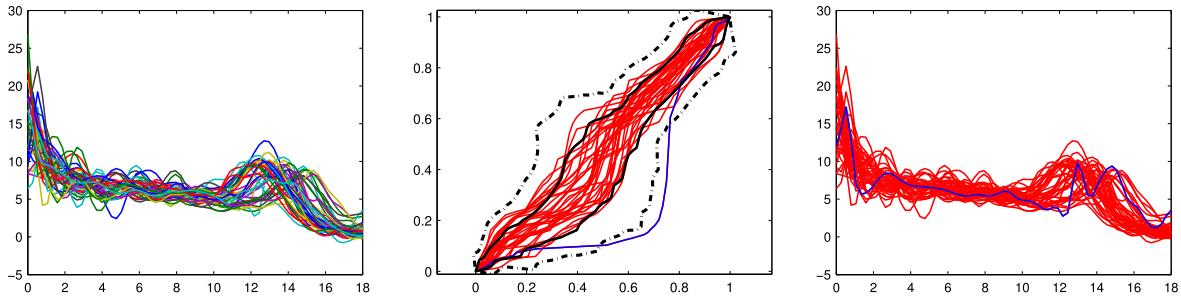
4.3. Illustration

In this section, we discuss the difference between the two approaches for extending the model of the boxplot and which should be used for warping functions. We make this distinction because the preferred approach depends entirely on what type of functional data is used. In the time warping function space Γ, it is easy to see the proposed boxplot is preferable. We use Fig. 10 to demonstrate this point.

We use the coefficient of $\mathcal{C} = 0.7$ for our whisker computation because the traditional coefficient 1.5 will often fail to catch any outliers in our phase approach due to the domain and range width of the time warping space Γ. We can easily see three of the six outliers are present in Fig. 10B while six outliers are present in Fig. 10C. The question is whether it is legitimate to consider only three of these functions outliers or not. Since this is simulated data, we can tell the construction of these six functions were to create three outlier functions and then use their inverses for the other three outliers. The problem with the traditional approach is it will never catch the functions that dip first and then rapidly speed up towards the end due to the curvature of the whiskers in that approach. The whiskers are not symmetric about an observation, but rather they are created based on the curvature of the IQR at each time point. In our approach, we specifically rotate all observations, so that we ensure our median matches our space and construct symmetric whiskers about this median signal.

4.4. Real data

In this section we look at the results of our boxplot extension on the two real data sets in Section 3. We start off by running the El Nino data set through our outlier detection methodology. We can see from Fig. 11 that only one observation ends up being detected as an outlier. This detected function does not have a dip in the early year time and is apparently an outliers in the set. Next, we use the Berkeley Growth data set and see if we can detect any outliers. The result is shown in Fig. 12. In this



(A) Observations. (B) Phase functional boxplot. (C) Outlier detection.

Fig. 12. Berkeley growth velocity data set given in Fig. 7. All line properties are the same as those in Fig. 11.

case, only one observation ends up being detected as an outlier. The function has growth spurts in a two peak form around 12 to 16 years old and is a typical outlier. These real experimental results show the new boxplot method can effectively detect outliers in time warping functions.

5. Conclusion and discussion

In this paper we have studied the problem of robust signal estimation in the field of functional data analysis. We introduced a framework for phase band depth. Our methodology builds upon the inherent properties of Fisher–Rao registration and band depth while also providing insight into a better design for the functional boxplot in the time warping setting.

One major contribution in this new method is that it provides an effective approach for robust signal estimation when extreme values and outliers are present. Our proposed algorithm, under a semi-parametric framework, provides a consistent estimator for the underlying signal. Simulations and real data sets were used to show that even when outliers and extremes are present in the data, our methodology leads to a robust signal estimation. We point out the time warping space required a more developed approach for outlier detection. Thus the boxplot was extended to operate within the space of time warping functions. We have shown our boxplot on the phase variability preserves ranking induced by the band depth and creates a more symmetric construction of the boxplot with respect to time warping functions.

The current study has focused on phase variability only in the given functional data. In the future we will study the ranking-based on the combination of phase and amplitude variability. In addition, the boxplot on time warping functions is only in the given function space. In the future, we will explore the variability in their derivatives (e.g. using Sobolev norms), so that we better capture the shape variation in the warping functions (Xie et al., 2017).

Appendix A. Proofs of the four mathematical properties

A.1. Proof of Proposition 1

Proof. $\Rightarrow$ Suppose $G(\gamma) \in B(\gamma_1, \ldots, \gamma_n)$. Therefore $\min_{i=1,\ldots,n}\gamma_i(t) \leq \gamma(t) \leq \max_{i=1,\ldots,n}\gamma_i(t)$ for all $t \in [0, 1]$. Let $s = \tilde{\gamma}(t)$, where $\tilde{\gamma} \in \Gamma$. Then $s \in [0, 1]$, and

$$\begin{aligned} & \min_{i=1,\ldots,n}(\gamma_i \circ \tilde{\gamma})(t) \leq (\gamma \circ \tilde{\gamma})(t) \leq \max_{i=1,\ldots,n}(\gamma_i \circ \tilde{\gamma})(t) \\ \Leftrightarrow \quad & \min_{i=1,\ldots,n}\gamma_i(\tilde{\gamma}(t)) \leq \gamma(\tilde{\gamma}(t)) \leq \max_{i=1,\ldots,n}\gamma_i(\tilde{\gamma}(t)) \\ \Leftrightarrow \quad & \min_{i=1,\ldots,n}\gamma_i(s) \leq \gamma(s) \leq \max_{i=1,\ldots,n}\gamma_i(s). \end{aligned}$$

The inequalities in the last line certainly hold. Therefore, $G(\gamma \circ \tilde{\gamma}) \in B(\gamma_1 \circ \tilde{\gamma}, \ldots, \gamma_n \circ \tilde{\gamma})$.

$\Leftarrow$ Suppose $G(\gamma \circ \tilde{\gamma}) \in B(\gamma_1 \circ \tilde{\gamma}, \ldots, \gamma_n \circ \tilde{\gamma})$. Then $\min_{i=1,\ldots,n}(\gamma_i \circ \tilde{\gamma})(t) \leq (\gamma \circ \tilde{\gamma})(t) \leq \max_{i=1,\ldots,n}(\gamma_i \circ \tilde{\gamma})(t)$ for all $t \in [0, 1]$. Let $s = \tilde{\gamma}^{-1}(t)$. Then $t = \tilde{\gamma}(s)$, and

$$\begin{aligned} & \min_{i=1,\ldots,n}\gamma_i(t) \leq \gamma(t) \leq \max_{i=1,\ldots,n}\gamma_i(t) \\ \Leftrightarrow \quad & \min_{i=1,\ldots,n}\gamma_i(\tilde{\gamma}(s)) \leq \gamma(\tilde{\gamma}(s)) \leq \max_{i=1,\ldots,n}\gamma_i(\tilde{\gamma}(s)) \\ \Leftrightarrow \quad & \min_{i=1,\ldots,n}\gamma_i \circ \tilde{\gamma}(s) \leq \gamma \circ \tilde{\gamma}(s) \leq \max_{i=1,\ldots,n}\gamma_i \circ \tilde{\gamma}(s). \end{aligned}$$

The inequalities in the last line hold. Therefore, $G(\gamma) \in B(\gamma_1, \ldots, \gamma_n)$. $\square$

A.2. Proof of Proposition 2

Proof. $\Rightarrow$ If $G(\gamma) \in B(\gamma_1, \ldots, \gamma_n)$, then $\min_{i=1,\ldots,n}\gamma_i(t) \leq \gamma(t) \leq \max_{i=1,\ldots,n}\gamma_i(t)$ for all $t \in [0, 1]$. For any $s \in [0, 1]$, let $t = \gamma^{-1}(s)$, or $s = \gamma(t)$. There exists $j \in \{1, \ldots, n\}$ such that $\gamma_j(t) \leq s$. Hence, there exists $a \in [t, 1]$ such that $\gamma_j(a) = s$ and $\gamma_j^{-1}(s) = a \geq t = \gamma^{-1}(s)$. This shows that $\gamma^{-1}(s) \leq \max_{i=1,\ldots,n}\gamma_i^{-1}(s)$. Similarly, we can have $\gamma^{-1}(s) \geq \min_{i=1,\ldots,n}\gamma_i^{-1}(s)$.
$\Leftarrow$ The proof is analogous to that in $\Rightarrow$. The detail is omitted. □

A.3. Proof of Proposition 3

Proof. Suppose we have a collection of warping functions, $\{\gamma_1, \ldots, \gamma_n\}$. Let $\gamma \in \Gamma$ that we will use to shift the others. Consider these sets of functions M_S and S_M:

$$M_S \stackrel{d}{=} \{(\gamma_M \circ \gamma)(t) : \gamma_M \in BM(\{\gamma_1, \ldots, \gamma_n\})\}$$
$$S_M \stackrel{d}{=} \{\gamma_S(t) : \gamma_S \in BM(\{\gamma_1 \circ \gamma, \ldots, \gamma_n \circ \gamma\})\}.$$

Our goal now is to show the set M_S is equivalent to the set S_M. Using Proposition 1 we see:

$$\tilde{\gamma} \in S_M \Leftrightarrow \tilde{\gamma} \in \operatorname*{argmax}_{\gamma^*} \sum_{j=2}^{J} \binom{n}{j}^{-1} \sum_{1\leq i_1<\cdots<i_j\leq n} I\{G(\gamma^*) \subseteq B(\gamma_{i_1} \circ \gamma, \ldots, \gamma_{i_j} \circ \gamma)\}$$
$$\Leftrightarrow \tilde{\gamma} \in \operatorname*{argmax}_{\gamma^*} \sum_{j=2}^{J} \binom{n}{j}^{-1} \sum_{1\leq i_1<\cdots<i_j\leq n} I\{G(\gamma^* \circ \gamma^{-1}) \subseteq B(\gamma_{i_1}, \ldots, \gamma_{i_j})\}$$

[Let $\gamma^* \circ \gamma^{-1}$ become γ^{**}.]

$$\Leftrightarrow \tilde{\gamma} \circ \gamma^{-1} \in \operatorname*{argmax}_{\gamma^{**}} \sum_{j=2}^{J} \binom{n}{j}^{-1} \sum_{1\leq i_1<\cdots<i_j\leq n} I\{G(\gamma^{**}) \subseteq B(\gamma_{i_1}, \ldots, \gamma_{i_j})\}$$
$$\Leftrightarrow \tilde{\gamma} \in M_S. \quad \square$$

A.4. Proof of Proposition 4

Proof. Consider these sets of functions M_I and I_M:

$$M_I \stackrel{d}{=} \{(\gamma_M^{-1})(t) : \gamma_M \in BM(\{\gamma_1, \ldots, \gamma_n\})\}$$
$$I_M \stackrel{d}{=} \{\gamma_I(t) : \gamma_I \in BM(\{\gamma_1^{-1}, \ldots, \gamma_n^{-1}\})\}.$$

Our goal now is to show the set M_I is equivalent to the set I_M. Using Proposition 2 we see:

$$\tilde{\gamma}^{-1} \in I_M \Leftrightarrow \tilde{\gamma}^{-1} \in \operatorname*{argmax}_{\bar{\gamma}} \sum_{j=2}^{J} \binom{n}{j}^{-1} \sum_{1\leq i_1<\cdots<i_j\leq n} I\{G(\bar{\gamma}) \subseteq B(\gamma_{i_1}^{-1}, \ldots, \gamma_{i_j}^{-1})\}$$
$$\Leftrightarrow \tilde{\gamma}^{-1} \in \operatorname*{argmax}_{\bar{\gamma}} \sum_{j=2}^{J} \binom{n}{j}^{-1} \sum_{1\leq i_1<\cdots<i_j\leq n} I\{G(\bar{\gamma}^{-1}) \subseteq B(\gamma_{i_1}, \ldots, \gamma_{i_j})\}$$

[Let $\bar{\gamma}^{-1}$ become γ^*.]

$$\Leftrightarrow \tilde{\gamma} \in \operatorname*{argmax}_{\gamma^*} \sum_{j=2}^{J} \binom{n}{j}^{-1} \sum_{1\leq i_1<\cdots<i_j\leq n} I\{G(\gamma^*) \subseteq B(\gamma_{i_1}, \ldots, \gamma_{i_j})\}$$
$$\Leftrightarrow \tilde{\gamma}^{-1} \in M_I. \quad \square$$

Appendix B. Proof of consistency

Proof. Our observations, $f_i(t) = c_i(g \circ \gamma_i)(t) + e_i$, can be represented as SRVF as follows: $q_i = \sqrt{c_i}(q_g, \gamma_i)$. WLOG let q_1 to be the median template (i.e. $m = q_1$). The alignment yields the following:

$$\gamma_i^* = \operatorname*{argmin}_{\gamma} \|q_1 - (q_i, \gamma)\| = \operatorname*{argmin}_{\gamma} \|\sqrt{c_1}(q_g, \gamma_1) - \sqrt{c_i}(q_g, \gamma_i \circ \gamma)\| = \gamma_i^{-1} \circ \gamma_1.$$

Thus, $\tilde{\gamma_n} = BM\{\gamma_1^*, \ldots, \gamma_n^*\} = BM\{\gamma_1^{-1} \circ \gamma_1, \ldots, \gamma_n^{-1} \circ \gamma_1\} = BM\{\gamma_1^{-1}, \ldots, \gamma_n^{-1}\} \circ \gamma_1 = BM\{[\gamma_1, \ldots, \gamma_n\}]^{-1} \circ \gamma_1 \to [\gamma_{id}]^{-1} \circ \gamma_1 = \gamma_1$. a.s.

Then, $\tilde{f_i} = f_i \circ \gamma_i^* \circ \tilde{\gamma_n}^{-1} = (c_i[g \circ \gamma_i] + e_i) \circ (\gamma_i^{-1} \circ \gamma_1) \circ \tilde{\gamma_n}^{-1} = c_i[g \circ \gamma_1 \circ \tilde{\gamma_n}^{-1}] + e_i$.

Therefore $\frac{1}{n}\sum_{i=1}^{n} \tilde{f_i} = \bar{c}[g \circ \gamma_1 \circ \tilde{\gamma_n}^{-1}] + \bar{e} \to E(c_1)g + E(e_1)$. a.s. □

Appendix C. Proof of the lemmas and theorem on the boxplots

C.1. Proof of Lemma 1

Proof. Assume $\gamma_1(t) \le \gamma_2(t)$ for $t \in [0, 1]$. Then $\forall s \in [0, \sqrt{2}]$, $\exists t \in [0, 1]$ s.t. $\tilde{\gamma_i}(t) = s$ and $P(\gamma_i(t)) = \lambda_i(s) = \frac{1}{\sqrt{2}}(\gamma_i(t) - t)$. Therefore $P(\gamma_1(t)) = \frac{1}{\sqrt{2}}(\gamma_1(t) - t) \le \frac{1}{\sqrt{2}}(\gamma_2(t) - t) = P(\gamma_2(t))$. □

C.2. Proof of Lemma 2

Proof. Consider $F(a) = a + \gamma_{\min}(a) - (\gamma(t) + t)$ for $a \in [t, 1]$. Then $F(t) = t + \gamma_{\min}(t) - (\gamma(t) + t) \le 0$ and $F(1) = 1 + \gamma_{\min}(1) - (\gamma(t) + t) \ge 0$. So $\exists a \in [t, 1]$ s.t. $a + \gamma_{\min}(a) = \gamma(t) + t$. Thus $\forall s \in [0, \sqrt{2}]$, $\exists t \in [0, 1]$ and $\exists a \in [t, 1]$ s.t. $\tilde{\gamma}(t) = s \Rightarrow \tilde{\gamma}_{\min}(a) = \frac{1}{\sqrt{2}}(\gamma_{\min}(a) + a) = s$, $\tilde{\gamma}_{\min}^{-1}(s) = a$. $P(\gamma_{\min}(t)) = \frac{1}{\sqrt{2}}(\gamma_{\min}(\tilde{\gamma}_{\min}^{-1}(s)) - \tilde{\gamma}_{\min}^{-1}(s)) = \frac{1}{\sqrt{2}}(\gamma_{\min}(a) - a)$ and $P(\gamma(t)) = \frac{1}{\sqrt{2}}(\gamma(t) - t)$. Therefore $P(\gamma_{\min}(t)) \le P(\gamma(t)) \Leftrightarrow \gamma_{\min}(a) - a \le \gamma(t) - t \Leftrightarrow t + \gamma(t) - 2a \le \gamma(t) - t \Leftrightarrow t \le a$. □

C.3. Proof of Theorem 2

Proof. By the definitions of $\gamma_{\min}$ and $\gamma_{\max}$, $\forall t \in [0, 1]$, $\gamma_{\min}(t) \le \gamma(t) \le \gamma_{\max}(t) \Rightarrow \forall s \in [0, \sqrt{2}]$, $\exists t \in [0, 1]$ s.t. $P(\gamma_{\min}(t)) \le P(\gamma(t)) \le P(\gamma_{\max}(t))$ from the previous lemmas. □

References

Capra, W., Muller, H., 1997. An accelerated-time model for response curves. J. Amer. Statist. Assoc. 92, 72–83.

Chakraborty, A., Chaudhuri, P., et al., 2014. The spatial distribution in infinite dimensional spaces and related quantiles and depths. Ann. Statist. 42 (3), 1203–1231.

Claeskens, G., Hubert, M., Slaets, L., Vakili, K., 2014. Multivariate functional halfspace depth. J. Amer. Statist. Assoc. 109 (505), 411–423.

Cuesta-Albertos, J.A., Nieto-Reyes, A., 2008. The random tukey depth. Comput. Statist. Data Anal. 52 (11), 4979–4988.

Cuevas, A., Febrero, M., Fraiman, R., 2007. Robust estimation and classification for functional data via projection-based depth notions. Comput. Statist. 22 (3), 481–496.

Ferraty, F., Vieu, P., 2006. Nonparametric Functional Data Analysis, Theory and Practice. In: Springer Series in Statistics.

Fraiman, R., Muniz, G., 2001. Trimmed means for functional data. Test 10 (2), 419–440.

Gasser, T., Kneip, A., 1995. Searching for structure in curve sample. J. Amer. Statist. Assoc. 90, 1179–1188.

Gervini, D., Gasser, T., 2004. Self-modeling warping functions. J. R. Stat. Soc. Ser. B Stat. Methodol. 66, 959–971.

Hall, P., Lee, Y., Park, B., 2007. A method for projecting functional data onto a low-dimensional space. J. Comput. Graph. Statist. 16, 799–812.

James, G., 2007. Curve alignments by moments. Ann. Appl. Stat. 1 (2), 480–501.

Kneip, A., Gasser, T., 1992. Statistical tools to analyze data representing a sample of curves. Ann. Statist. 20, 1266–1305.

Kneip, A., Ramsay, J.O., 2008. Combining registration and fitting for functional models. J. Amer. Statist. Assoc. 103 (483).

Kurtek, S., Srivastava, A., Wu, W., 2011. Signal estimation under random time-warpings and nonlinear signal alignment, in: Proceedings of Neural Information Processing Systems, NIPS.

Kurtek, S., Su, J., Grimm, C., Vaughan, M., Sowell, R., Srivastava, A., 2013. Statistical analysis of manual segmentations of structures in medical images. Comput. Vis. Image Underst..

Liu, R.Y., 1990. On a notion of data depth based on random simplices. Ann. Statist. 18, 405–414.

Liu, X., Muller, H.G., 2004. Functional convex averaging and synchronization for time-warped random curves. J. Amer. Statist. Assoc. 99, 687–699.

Lopez-Pintado, S., Romo, J., 2009. On the concept of depth for functional data. J. Amer. Statist. Assoc. 104, 718–734.

López-Pintado, S., Romo, J., 2011. A half-region depth for functional data. Comput. Statist. Data Anal. 55 (4), 1679–1695.

Ramsay, J.O., Li, X., 1998. Curve registration. J. R. Stat. Soc. Ser. B Stat. Methodol. 60, 351–363.

Ramsay, J.O., Silverman, B.W., 2005. Functional Data Analysis, second ed. In: Springer Series in Statistics.

Srivastava, A., Wu, W., Kurtek, S., Klassen, E., Marron, J.S., 2011. Registration of functional data using fisher-rao metric. J. Amer. Statist. Assoc. (submitted for publication). arXiv:1103.3817v2.

Sun, Y., Genton, M.G., 2011. Functional boxplot. J. Comput. Graph. Statist. 20, 316–334.

Tang, R., Muller, H.G., 2008. Pairwise curve synchronization for functional data. Biometrika 95 (4), 875–889.

Wang, K., Gasser, T., 1997. Alignment of curves by dynamic time warping. Ann. Statist. 25 (3), 1251–1276.

Xie, W., Kurtek, S., Bharath, K., Sun, Y., 2017. A geometric approach to visualization of variability in functional data. J. Amer. Statist. Assoc. 112 (519), 979–993.