

# Clustering Multivariate Functional Data with Phase Variation

Juhyun Park<sup>1,\*</sup> and Jeongyoun Ahn<sup>2,\*\*</sup>

<sup>1</sup>Department of Mathematics and Statistics, Lancaster University, Lancaster LA1 4YF, U.K.

<sup>2</sup>Department of Statistics, University of Georgia, Athens, Georgia 30602-1952, U.S.A.

\**email*: juhyun.park@lancaster.ac.uk

\*\**email*: jyahn@uga.edu

**SUMMARY.** When functional data come as multiple curves per subject, characterizing the source of variations is not a trivial problem. The complexity of the problem goes deeper when there is phase variation in addition to amplitude variation. We consider clustering problem with multivariate functional data that have phase variations among the functional variables. We propose a conditional subject-specific warping framework in order to extract relevant features for clustering. Using multivariate growth curves of various parts of the body as a motivating example, we demonstrate the effectiveness of the proposed approach. The found clusters have individuals who show different relative growth patterns among different parts of the body.

**KEY WORDS:** Curve alignment; Functional clustering; Growth curves; Multivariate functional data; Phase variation.

## 1. Introduction

The rapid development of technology in recording and storing of data makes it possible to collect simultaneous recordings of several longitudinal variables. Often physiological and biological processes are continuously monitored for a number of subjects, resulting in a set of multivariate longitudinal trajectories. When the measurements are viewed as realizations of smooth random functions, the field of functional data analysis (FDA) provides a wealthy suite of techniques to explore and analyze such types of data. For an overview, we refer to Ramsay and Silverman (2002, 2005) and Ferraty and Vieu (2006). Although FDA is proven useful to deal with data in the form of curves, its extension to multivariate functional data is not yet fully developed and it is an active area of ongoing research. In this work, we focus on the problem of clustering multivariate functional data.

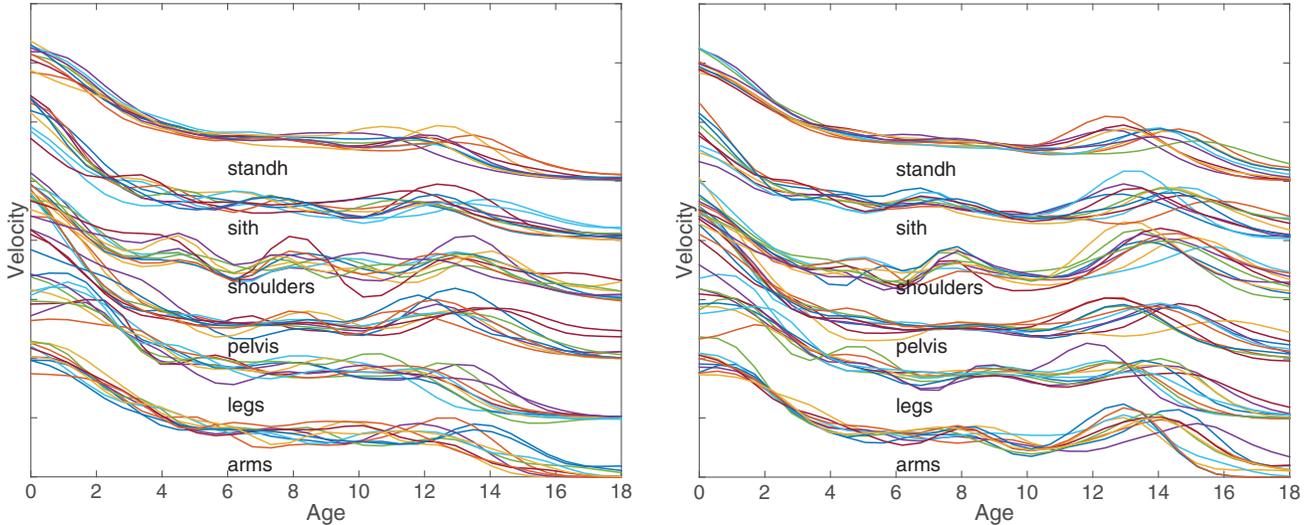
Multivariate functional data can arise in many different settings. The three-dimensional curves from the coordinates of a geometrical object are a good example. Alternatively, we may observe multiple response curves from the same object sharing the common underlying process. For example, in functional Magnetic Resonance Imaging, simultaneously measured signals from neighboring voxels could be viewed as replicates of brain signals subject to some regional or spatial variability. Different types or techniques of spectroscopy used to study molecules in chemistry rely on the same principle that molecules absorb specific frequencies that are characteristic of their structure. A well-known example comes from a longitudinal growth study, where measurements of various body parts are simultaneously obtained from the same subject. The growth curves have served as a benchmark dataset for functional data, but mostly discussed in the context of one-dimensional curves (e.g., Ramsay and Silverman, 2005).

The particular dataset that we are interested in, the Zürich longitudinal growth study database, includes measurements

on six different variables: height, sitting height, leg height, arm length, bihumeral width (shoulder), and biiliac width (pelvis) (Gasser et al., 1984; Sheehy et al., 1999, 2000). Figure 1 shows an example of the growth response curves, in relative velocity, for the first 10 subjects from boys and girls, respectively. Remarkably, the common characteristics are shared by all the variables.

There is no doubt that the growth process of arms is different from that of legs, for example. Nevertheless, it may be assumed that there is a common growth process that coordinates differential developments of different body parts. Then, it is of great interest to characterize the growth process and to identify groups with different growth patterns. Earlier works were successful in characterizing the growth process in the analysis of single body part measurements, however, the analysis of joint behaviors of growth of various body parts has been scarcely developed. Sheehy et al. (1999) is the earlier reference to attempt to analyze the multiple curves simultaneously, though more in line with classical multivariate analysis.

When an object is represented by multivariate responses, it is important to take into account the dependence between the responses. There are many standard dependence measures for multivariate Euclidean data, often under certain distributional assumptions. However, it is not so obvious how to embed such concepts into functional data as an infinite-dimensional object, with various sources of variation influencing the structure of the curves. Multivariate Functional Principal Component Analysis (MPCA) provides a standard tool to capture the variation in a parsimonious way, utilizing the dependence through covariance function, which is suitable to describe a linear dependence between the curves. Nevertheless, the notion of dependence can be closely linked to the scientific questions associated with the data and thus the formulation requires a careful consideration.



**Figure 1.** Growth curves in relative velocity for a sample of subjects for different body parts. First 10 subjects in the girls (left) and the boys (right).

Going back to the example, the growth data have several unique features that standard functional clustering methods do not take into account (Gasser et al., 1984; Gasser and Kneip, 1995). Firstly, the main information of the growth process is contained in the speed of growth, that is, the derivatives of the original measurements. Secondly, the main variation of interest in derivatives occurs in *phase*, rather than in *amplitude*. In fact, amplitude variation could be dominant in size, but considered as a nuisance for comparison of the speed of growth. In addition, due to the natural limit in size for different body parts, it is more natural to compare the variables on a standardized scale. Thirdly, although the variables could be made comparable, there is no natural scale to combine information from different sources once standardization is applied.

With these considerations in mind, we consider ways to describe the joint patterns of growth that differentiate *developmental phases* among the subgroups. For example, one might conjecture that on average,

- legs grow faster than shoulder;
- arm and shoulder tend to grow together;
- nonuniform growth across different body parts is more common compared to uniform growth of all body parts;
- and so on.

To express these ideas in statistical terms, we note that a main signal in growth pattern comparisons is related to the ordering of variables with respect to timing of prominent features, that is, the notion of the *phase* variation. Phase variation is well understood in the context of one-dimensional curves, though often considered as a nuisance. Our focus is on capturing a *relative phase variation across multiple variables* to account for dependence, a distinguishing feature of our methodology in contrast with the existing methods dealing with phase variation. As will become clear, our approach offers a complementary and nonstandard view of the growth pattern.

Our main contributions could be summarized as follows. Although the problem of phase and amplitude variation is

well known in general, its implication in multivariate functional data analysis has not been well studied yet. Even if we focus on estimation of phase variation and curve alignment, the multivariate version does not exist because it essentially requires to define a proper concept of *multivariate monotone functions* at its full generality. The first novelty of our article is the very new formulation of the problem of phase variation in multivariate functional data, and its characterization in terms of different parametrization required. As will be seen later, mathematically as well as practically, the multivariate extension is not unique and it requires serious consideration in the context of the problem at hand. Once said, it sounds obvious, but its implication is far greater. We demonstrate the implications in the context of functional clustering. The second novelty is the new clustering method for phase variation only. As many other new developments, many ingredients of the proposed method are borrowed from existing methods. However, amplitude and phase variations are not comparable to each other, and thus direct translation of amplitude method to phase or a naive mixing and matching strategy does not work. With proper considerations of these issues, we develop a new clustering approach applicable to phase variation only. The third novelty is the application itself. As far as we know, the joint analysis of growth phases has not been done before (possibly due to the difficulty in getting such data), and there is no standard methodology available.

This article is organized as follows. Section 2 introduces the framework for phase variation in multivariate functional curves. Section 3 develops a clustering approach based on multivariate phase variation. Section 4 illustrates numerical performance in terms of simulation studies and a real data example.

## 2. Functional Data and Source of Variation

One of the fundamental challenges in exploring functional data lies in understanding the source of variation. It is well

known that the variation in one-dimensional functional data could be understood in two different directions, vertically (in amplitude) and horizontally (in phase). The amplitude variation can be mostly captured by linear techniques such as principal component analysis (PCA) while phase variation is inherently nonlinear and, without proper adjustment, the analysis as well as its interpretation could be misleading.

2.1. *Source of Variation in One-Dimensional Functional Curves*

Suppose we observe a sample of  $n$  curves on a compact interval  $\mathcal{I} = [0, T]$ , subject to measurement error. The  $i$ th response curve can be represented as  $Y_i(t) = X_i(t) + \varepsilon_i(t)$ , where  $\varepsilon_i(t)$  is the stationary residual process with  $E[\varepsilon_i(t)] = 0$ ,  $E[\varepsilon_i^2(t)] = \sigma^2$  and  $\text{Cov}(\varepsilon_i(t), \varepsilon_i(s)) = 0$ ,  $s \neq t$ . Assume that  $X_i$  is an element of  $L_2(\mathcal{I})$ .

When the curves share common characteristics subject to individual variability, that is, when we assume there exist a common mean  $\mu$  and a common covariance function  $\Gamma$ , we may further decompose  $X_i$  into  $X_i(t) = \mu(t) + V_i(t)$ , where  $E[V_i(t)] = 0$  and  $\text{Cov}(V_i(s), V_i(t)) = \Gamma(s, t)$ .

Often the observed curves exhibit additional variability in the horizontal direction representing delayed or accelerated responses, for which we introduce the time transformation function or warping function  $h$  and its inverse  $\omega = h^{-1}$  as

$$Y_i(t) = X_i(t) + \varepsilon_i(t), \quad X_i(\omega_i(t)) = \mu(t) + V_i(t), \quad (1)$$

where  $\omega_i$ (or  $h_i$ )s are monotonically increasing functions from  $\mathcal{I}$  to  $\mathcal{I}$  satisfying the boundary conditions  $\omega_i(0) = 0$  and  $\omega_i(T) = T$ . It is common to assume that  $\omega_i$ s are continuously differentiable. Conceptually, we imagine that there is a common underlying process, often characterized by common features or landmarks, so that  $\omega_i$  summarizes the timings of the onsets of the common features for the  $i$ th curve.

Most standard approaches developed in the literature to handle this unwieldy problem of phase variation have taken the view that phase variation is a nuisance and can be accounted for in a preprocessing step (Gasser and Kneip, 1995; Ramsay and Li, 1998; Wang and Gasser, 1999, 2009; Gervini and Gasser, 2004; James, 2007; Tang and Müller, 2008). Recent developments attempt to explicitly model the component of phase as the main step of the analysis, although the primary focus is on amplitude in general (Kneip and Ramsay, 2008; Liu and Yang, 2009; Tang and Müller, 2009; Slaets et al., 2012; Peng et al., 2014; Hadjipantelis et al., 2015).

As timings can be measured only relatively, there needs further constraints. For example, if we decide to use a reference curve, say  $X_1$ , then

$$X_1(t) = (\mu \circ h_1)(t) + (V_1 \circ h_1)(t) = \tilde{\mu}_1(t) + \tilde{V}_1(t).$$

Consequently, the timings of the other curves can be expressed as

$$\begin{aligned} X_i(t) &= (\mu \circ h_i)(t) + (V_i \circ h_i)(t) \\ &= (\tilde{\mu}_1 \circ \tilde{h}_i)(t) + (\tilde{V}_1 \circ \tilde{h}_i)(t), \end{aligned}$$

where  $\tilde{h}_i(t) = h_1^{-1} \circ h_i(t)$ . For standard one-dimensional curves, this is often expressed as  $E[\omega_i(t)] = i(t)$ , where

the identity map  $i(t) = t$  and its sample version as  $n^{-1} \sum_{i=1}^n \omega_i(t) = t$ , implying that on average there is no phase variation and we define a reference as the average response times. In other words, there is a representative curve with common features and the  $\omega_i(t)$  measures delays or accelerations relative to the timings of the representative curve at time  $t$ .

2.2. *Source of Variation in Multivariate Functional Curves*

Suppose now that for each subject, there are  $p$  components in the response defined on a common interval  $\mathcal{I}$ , denoted by  $\mathbf{X} = (X_1, \dots, X_p)$  and there might be some measurement errors. A general model could be written as

$$\mathbf{Y}_i(t) = \mathbf{X}_i(t) + \boldsymbol{\varepsilon}_i(t).$$

In the presence of phase variability, we introduce a vector of warping functions, denoted by  $\mathbf{h}_i = (h_{i1}, \dots, h_{ip})$  with its inverse  $\boldsymbol{\omega} = \mathbf{h}^{-1}$  to compensate for horizontal variability.

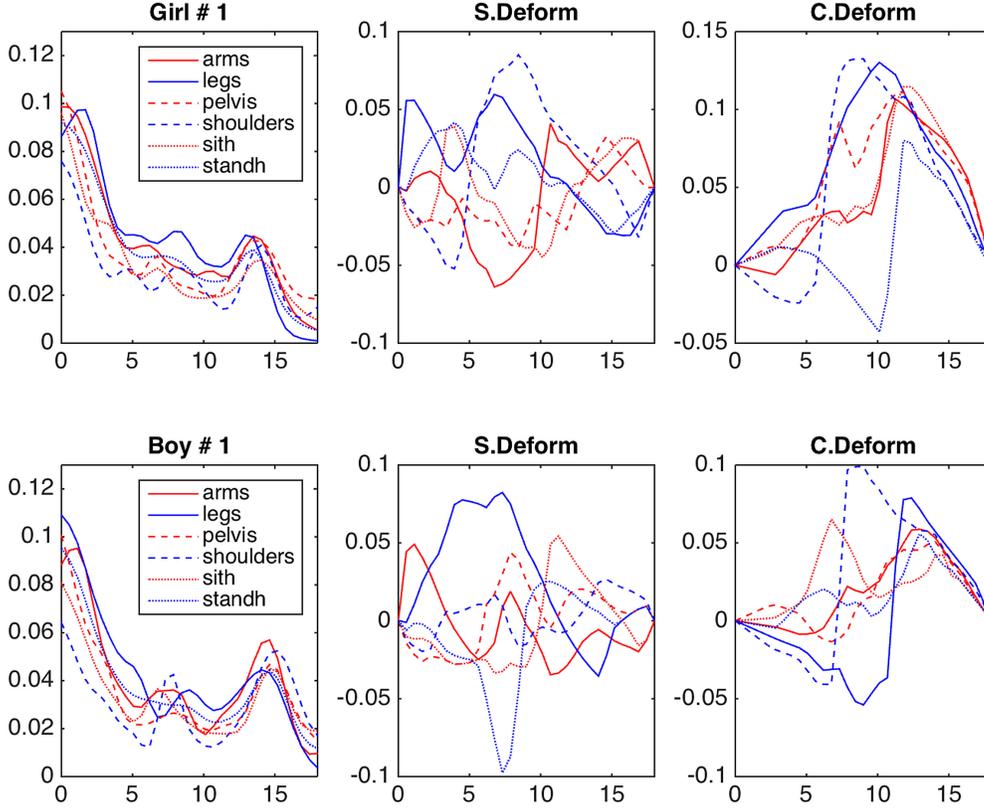
When it comes to multivariate functional data, the issues of phase and amplitude variations still exist but are more difficult to address due to its inherent geometry (Kurtek et al., 2012; Brunel and Park, 2014). Consequently, an extension to multivariate functional curves is not unique and thus needs some careful consideration. In what follows, we list several plausible parametrizations, each of which is based on a different underlying data-generating mechanism.

2.2.1. *Marginal component-wise warping.* The marginal warping is a direct component-wise extension of the one-dimensional curve alignment formulation to multivariate case. Here, we assume that each component has its own average response pattern and phase variation in individual curves vary only relative to the average response time for each component, independent of other components. Then, we would proceed by repeating one-dimensional registration analysis  $p$  times, that is, for each component  $r$ ,  $r = 1, \dots, p$ , the warping function  $\omega_r^c$  satisfies

$$E[(X_{ir} \circ \omega_{ir}^c)(t)] = \mu_r(t), \quad (2)$$

subject to  $n^{-1} \sum_{i=1}^n \omega_{ir}^c(t) = t$  for  $r = 1, \dots, p$ . This strategy is useful as a preprocessing or the primary interest is the amplitude variation, and also it is the approach taken by most multivariate FDA works. However, as individual acceleration/delay is measured relatively within each component variable, this does not allow us to compare relative delays across components. For illustration, an example from the growth data is shown in Figure 2 in the third column.

2.2.2. *Conditional subject-specific warping.* As in the growth data example, when the component variables are comparable, the comparison across the variables are of interest. For example, in the case where a certain component tends to respond systematically later than others across subjects, this cannot be reflected within component-wise standardization. Instead the subject-specific average response curve can be used as a reference for the individual warping function



**Figure 2.** An example of warping profiles extracted from subject-specific warping (middle) and component-wise warping (right) for the first subject in each group. The left column shows the original growth velocity curves.

$\omega_i^s$ . Denote the  $i$ th subject specific average by  $S_i$  and assume that

$$E[(X_{ir} \circ \omega_{ir}^s)(t) | X_i] = S_i(t), \quad (3)$$

subject to  $p^{-1} \sum_{r=1}^p \omega_{ir}^s(t) = t$  for each  $i = 1, \dots, n$ . From the viewpoint of variance decomposition,  $S_i(t)$  can be regarded as the conditional mean of the subject-specific random effect. In particular, if the decomposition in (1) were to hold for each component, the subject-specific reference could be identified as  $\mu(t) + E[V_{ir}(t) | X_i]$ , assuming a common trend for all components.

Now individual warping profiles ( $\omega_i^s$ ) can be compared between subjects, without reference to the pattern of the population average. An example from the growth data is shown in Figure 2 in the second column. This is our model used to extract features of differential growth pattern for each subject for clustering, discussed in Section 3.

*2.2.3. Universal warping for all components.* In contrast to the previous scenarios, often the registration problem occurs in the study of the displacement of an object (e.g., Kurtek et al., 2012; Brunel and Park, 2014) so that the response curves could be identified with respect to some underlying coordinate system. It is then more appropriate to assume a common warping function for all response curves:

$$E[(X_{ir} \circ \omega_i)(t)] = \mu_r(t), \quad r = 1, \dots, p,$$

subject to  $n^{-1} \sum_{i=1}^n \omega_i(t) = t$ . In this case, the unit of analysis is the vector of response curves as a whole and the warping function acts as a nuisance parameter in characterizing the variability in the population of the vector of response curves.

### 3. Multivariate Phase-Cluster Model

Viewing the multiple response curves as a whole unit, we are interested in grouping subjects based on the phase variation differentiating joint growth patterns. When variability in the size or the magnitude of each variable could be of interest, the methodology based on multivariate FPCA could be used (e.g., Jacques and Preda, 2012), and, if necessary, registration methods are combined as a preprocessing (e.g., Ieva et al., 2013). These methods could be efficient but have the same drawbacks as one-dimensional approach when phase variation is of main interest. Sangalli et al. (2010) extend a  $k$ -means-like clustering algorithm to three-dimensional curves with phase variation by explicitly specifying a class of (linear) warping functions in the template so that the curves are aligned within each cluster. Brunel and Park (2014) directly focus on general phase variation in so far as estimating the *structural mean* function but do not address the problem of clustering. In fact, it is a nontrivial problem to find a representation for a cluster center when both variability are present, due to the fact that the functions spaces do not coincide. This is well illustrated in Liu and Müller (2004).

In this section, we present our formulation of the problem of clustering multivariate functional data. Unlike the standard

view of functional clustering, we focus on the case where phase variation prevails and amplitude variation is considered as a nuisance. There are at least two possible approaches to finding clusters based only on phase variation and the two approaches can be explained via characterizing the clusters that they set out to find. In the first approach, one can try to find clusters that have different phase patterns ( $\omega$ ) in certain components. The second approach, which is our focus, is to find clusters that have different phase variability across components within an observation.

Figure 2 illustrates extracted features of the first subject in each growth dataset for a comparison. The first column shows the original velocity curves, and the second column shows the curves defined as  $\hat{\omega}(t) - t$ , called deformation functions, from the proposed conditional subject-specific warping in (3). Here, it is expected that the deformation curves have zero mean from the formulation. The crossing patterns in the deformation curves indicate that a measurement variable reaches a growth peak earlier than some others and then its growth slows down later. The third column shows the deformation curves from marginal component-wise warping in (2). Here, the curves are showing how the subject's growth is delayed or accelerated compared to other subjects in the data, component-wise.

We assume that the population consists of a mixture of  $K$  clusters that have distinct patterns of delays. Our earlier discussion in Section 1 suggests that the dynamic ordering of the warping profiles could be used to define a cluster. However, as the ordering may change continuously over time, this is rather a difficult problem. Instead, we detect the effect of changes in ordering through the relative changes in warping profiles across the component variables. This can be achieved by incorporating the conditional warping as opposed to other types of warping.

### 3.1. A Phase-Cluster Model

Given the multivariate functional curves, we consider a phase-cluster model represented as follows. Our observation model is  $Y_{ir}(t) = X_{ir}(t) + \varepsilon_{ir}(t)$ ,  $i = 1, \dots, n$ ;  $r = 1, \dots, p$  for  $t \in [0, T]$  with

$$E[X_{ir}(\omega_{ir}^{(k)}(t)) | X_i] = S_i(t), \text{ for } k = 1, \dots, K, \quad (4)$$

subject to  $p^{-1} \sum_{r=1}^p \omega_{ir}^{(k)}(t) = t$  and  $\omega_{ir}^{(k)}(0) = 0$  and  $\omega_{ir}^{(k)}(T) = T$  and  $K$  is the number of clusters. Under the subject-specific warping ( $p^{-1} \sum_{r=1}^p \omega_{ir}(t) = t$ ), it follows that  $n^{-1} \sum_{i=1}^n \omega_{ir}(t) \neq t$  for  $r = 1, \dots, p$ . In particular, if there is a systematic delay in certain component variable, for example, if shoulder tends to grow later than other body parts on average for subjects in a certain cluster, the corresponding warping function  $\omega$  for shoulder would tend to lie above the identity line for each individual. Then, on average,  $n^{-1} \sum_{i=1}^n \omega_{ir}(t) - t$  tends to be positive for shoulder. Motivated by this observation, we define a cluster based on relative delays or deformation function for the  $i$ th subject defined by  $\mathbf{u}_i(t) = \boldsymbol{\omega}_i(t) - t$ . It is clear that the relative phase variations can be translated into the differences in mean of the deformation functions. In addition, the space of the deformation function is better approximated by a linear space and thus alleviates the issues of directly dealing with

the space of the warping functions, which is non-Euclidean and not a linear vector space.

### 3.2. Characterization of Phase Clusters

Defining clusters may or may not require the information on variability. In order to characterize the cluster in the vector of warping functions  $\boldsymbol{\omega}$ , we formulate a latent functional mixture model through the deformation functions.

Viewing the deformation functions as an element of  $L_2(\mathcal{I})$ , we assume that  $\mathbf{u}$  belongs to a Hilbert space  $H$ , of  $p$ -dimensional vectors of functions in  $L_2(\mathcal{I})$ . We define a cluster by its mean function in  $H$ , allowing for different variability across the cluster groups. Chiou and Li (2007) proposed a functional  $K$ -center clustering for univariate functional data which takes into account variation of the curves through FPCA in defining the cluster center. The same principle can be applied to multivariate functional data, if we replace FPCA by its multivariate version.

To define multivariate functional principal component analysis, we need to define the corresponding norm in the extended function space. For a multivariate function  $\mathbf{f} = (f_1, \dots, f_p)$  with  $\|f_r\|^2 < \infty$  for  $r = 1, \dots, p$ , define the  $p$ -variate norm by  $\|\mathbf{f}\|^2 = \sum_{r=1}^p \|f_r\|^2$ . Define the cross-covariance function by  $\Gamma_{r,s}(t, u) = \text{cov}(u_r(t), u_s(u))$ ,  $r, s = 1, \dots, p$ . Denote by  $\mathbf{\Gamma}(t, u)$  the  $p \times p$  matrix of the cross-covariance function whose  $(r, s)$ th element is given by  $\Gamma_{r,s}(t, u)$ . Then, the corresponding eigenvalues and eigenfunctions could be defined and the covariance function has the representation as  $\mathbf{\Gamma}(t, u) = \sum_{j=1}^{\infty} \lambda_j \boldsymbol{\phi}_j(s) \boldsymbol{\phi}_j(t)^T$ , where  $\lambda_j$ s are ordered eigenvalues and  $\boldsymbol{\phi}_j(t) = (\phi_{j1}(t), \dots, \phi_{jp}(t))^T$  are the unit-norm eigenfunctions satisfying

$$\sum_{s=1}^p \int \Gamma_{r,s}(t, u) \phi_{js}(u) du = \lambda_j \phi_{jr}(t).$$

Now, we assume that there are  $K$  subgroups in  $\mathbf{u}$ . Based on the eigenfunction expansion, we assume that each cluster has its own variability expressed through cluster-dependent eigenvalues and eigenfunctions:

$$\mathbf{u}^{(k)}(t) = \boldsymbol{\alpha}^{(k)}(t) + \sum_{j=1}^{\infty} \xi_j^{(k)} \boldsymbol{\phi}_j^{(k)}(t), \quad (5)$$

where  $\xi_j^{(k)}$  are uncorrelated and  $E[(\xi_j^{(k)})^2] = \lambda_j^{(k)}$  for  $k = 1, \dots, K$ .

Putting things together, the phase-cluster model (4) is refined to incorporate the additional information on variability for each cluster through (5). Note that within each cluster, amplitude function as well as phase function can vary. In particular, the amplitude function  $\mathbf{X}$  can have its own representation in terms of multivariate FPCA similar to the equation (5).

*Remark:* Although the FPCA representation is commonly assumed in joint analyses with the additive model structure (e.g., Kneip and Ramsay, 2008), the monotonicity of the warping function may not strictly be adhered to and may bring about some mild issues with interpretability. Peng et al. (2014) provide an example of nonmonotone warping functions to describe economic cycle in housing market. If one

insists, we can use a monotone parametrization by assuming that  $\frac{d}{dt} \log \omega$  can be expressed by the basis function expansion. Here, we keep the additive structure for simplicity as this provides a good approximation to describe the main variability in phase.

### 3.3. Estimation of Functional Phase-Cluster Model

Although our formulation resembles a standard mixture model, due to identifiability constraints for the latent warping functions, simultaneous estimation of this type of model is not straightforward. On the other hand, the problem is formulated in such a way that estimation of the latent warping functions is not directly related to how they are clustered. In fact, we consider estimation of the warping profiles as the step for feature extraction in clustering multivariate functional curves. Thus, we propose the following two-stage algorithm and give some details in the following.

- (i) [Feature extraction by conditional warping] For each subject  $i$ ,  $i = 1, \dots, n$ , estimate  $\omega_i$  by subject-specific warping and obtain deformation function  $\mathbf{u}_i(t) = \omega_i(t) - t$ .
- (ii) [Functional clustering] Apply multivariate functional clustering to  $\mathbf{u}$ .

*3.3.1. Feature extraction by conditional warping.* The subject-specific warping amounts to registration of multiple response curves for each subject. Let  $\Omega = \{\omega : \mathcal{I} \rightarrow \mathcal{I}, \omega \geq 0, \omega(0) = 0, \omega(T) = T, \omega \text{ is a diffeomorphism}\}$  be the set of warping functions. For  $i = 1, \dots, n$ ,

$$(S_i, \omega_i) = \arg \inf_{S \in L_2(\mathcal{I})} \inf_{(\gamma_1, \dots, \gamma_p) : \gamma_r \in \Omega} \sum_{r=1}^p d(X_{ir} \circ \gamma_{ir}, S).$$

This problem falls into the general problem of curve alignment for one-dimensional curves, a well-studied area in FDA. Most standard algorithms that treat phase variation as a nuisance are applicable to our case (Gasser and Kneip, 1995; Wang and Gasser, 1997; Ramsay and Li, 1998; Wang and Gasser, 1999; Gervini and Gasser, 2004; Liu and Müller, 2004; James, 2007; Tang and Müller, 2008; Tucker et al., 2013).

The main difficulty in directly solving this optimization problem is that the reference function  $S$  is also unknown so that some form of iteration would be necessary. In addition, the  $\gamma$  is an infinite-dimensional object that should satisfy monotonicity constraints. Different strategies in the choice of the metric  $d(\cdot, \cdot)$ , the parametrization of  $\gamma \in \Omega$ , the scheme of updating the reference  $S$ , as well as the optimization algorithm itself lead to different methods. One of the original methods known as landmark registration (Gasser and Kneip, 1995) exploits the fact that features that could identify  $S$  are available or could be well estimated using appropriate smoothing techniques for estimating derivatives. The standard monotone registration method in Ramsay and Silverman (2005) uses an explicit monotone parametrization with spline functions and solves iteratively a nonlinear least squares problem. Most methods developed an algorithm based on the standard  $L_2$  distance as  $d(X_1, X_2) = \|X_1 - X_2\|^2$  for  $X_1, X_2 \in L_2(\mathcal{I})$  whereas Tucker et al. (2013) based their al-

gorithm on Fisher–Rao metric that compares a normalized derivative function defined as

$$q_X(t) = \dot{X}(t) / \sqrt{|\dot{X}(t)|} \cdot (|\dot{X}(t)| \neq 0),$$

and  $d_{\text{FR}}(X_1, X_2) = \|q_1 - q_2\|$ . This formalizes the intuition that curves are better aligned based on the features in the derivatives. They developed an efficient algorithm based on dynamic programming. Comparison of different registration methods is beyond the scope of the article.

We used the registration method based on the Fisher–Rao metric (Tucker et al., 2013) in our analysis. It is algorithmically efficient and is known to be one of the most successful methods for registration. In our simulation studies, we evaluate the effectiveness of the registration method and its impact on clustering in comparison to the monotone registration method (Ramsay and Silverman, 2005) as a standard reference.

*3.3.2. Estimation of functional clusters in deformation function.* It turns out that the computation of multivariate FPCA can be simplified for the warping functions. We first define an extended curve as the concatenation of the multiple curves over the extended period of time. For each curve  $\mathbf{u}_r(t)$ ,  $r = 1, \dots, p$ ,  $\mathbf{u}_r(0) = \mathbf{u}_r(T) = 0$ . Let  $U(t)$ ,  $t \in [0, pT]$  be the concatenation of the curves  $(\mathbf{u}_1, \dots, \mathbf{u}_p)$  defined as

$$U(t) = \mathbf{u}_r(t), \quad \text{for } t \in [(r-1)T, rT].$$

Denote the mean function by  $\alpha$  and the covariance function by  $\Gamma(t, u) = \text{Cov}(U(t), U(u))$ . Then,  $U$  can be identified as a random process in  $L_2([0, pT])$ . In terms of procedural definition, the multivariate FPCA is equivalent to a univariate FPCA with these extended curves. Normally, the one-dimensional view requires additional continuity assumption at end points of each curve, which in general is not always satisfied by multivariate functions. However, viewing the warping function as a diffeomorphism with boundary conditions imposed on implies that  $\mathbf{u}_r(0) = \mathbf{u}_r(T) = 0$  and thus preserves continuity in concatenation. Although not required, further smoothness condition could be considered at the boundaries. Now, we reformulate the clustering problem with the extended curves of the deformation functions as follows. For the population of the  $k$ th cluster, we assume that

$$U^{(k)}(t) = \alpha^{(k)}(t) + \sum_{j=1}^{\infty} \xi_j^{(k)} \phi_j^{(k)}(t).$$

Another issue with multivariate functional principal component analysis is the comparability of the measure of variability across the variables, say  $X_r$ ,  $r = 1, \dots, p$ . Chiou et al. (2014) proposed to define the principal component analysis based on the standardized variables  $Z_r(t) = \Gamma_{r,r}(t, t)^{-1/2}(X_r(t) - \mu_r(t))$  where  $\mu$  is the mean function and  $\Gamma$  is the covariance function of  $X$ . However, we note that the deformations functions are comparable across the variables and do not require further standardization.

## 4. Numerical Studies

### 4.1. Simulation Studies

We identify two influential and inseparable factors that affect the performance being the success of the estimation of the warping profiles and the success of functional clustering method on the curves. In this section, we use simulated data to investigate relative performances of each component, with a standard implementation as a reference. For warping function estimation, the registration method by the Fisher–Rao metric is compared with the standard monotone registration method (Ramsay and Silverman, 2005). For clustering, Chiou and Li (2007)’s functional clustering with concatenated deformation function is compared with high-dimensional curve clustering method implemented in CCToolbox by Gaffney (2004).

As for the settings, we use total six different settings. Among those two are “null” settings for which the signals in data are not necessarily what the proposed method sets out to find. In one of the null settings, the two clusters are different only in their amplitude patterns, for which usual (multivariate) functional clustering approaches would target at. The other null setting is based on the marginal component warping model discussed in Section 2. These two settings are to illustrate that the proposed approach sets out to solve fundamentally different scientific problems than existing methods to begin with, thus for the null settings we expect to see random assignments in clustering with our approach.

The other four settings are designed to see not only whether the proposed approach works as intended but also which warping method and clustering method works best under various scenarios. As the warping profiles are latent and the amplitude information is a nuisance for clustering, we consider four distinct scenarios to evaluate the method of extracting phase variation in the presence of amplitude dilution as well as phase variation. The level of dilution is introduced by varying the covariance structure for phase functions in each cluster.

Each observation, belonging to one of the two clusters of size  $n = 25$  each, is a set of four curves ( $p = 4$ ) that are analogous to the different body parts in the growth data example. All curves are defined on the time range  $-3 \leq t \leq 3$  and generated as  $y(t) = S(h(t))$ , where  $S(t)$  represents the amplitude pattern and  $h(t)$  for phase variation. The amplitude of the  $r$ th component curve in an individual,  $S_r(t)$ ,  $r = 1, \dots, 4$  is one of the following:

$$\begin{aligned} S_{1r}(t) &= Z_{1r} \exp\{-.05(t-2)^2\} + Z_{2r} \exp\{-10(t+1)^2\}/6, \\ S_{2r}(t) &= Z_{1r} \exp\{-.05(t-2)^2\} + Z_{2r} \exp\{-10(t+1)^4\}/6, \end{aligned}$$

where  $Z_{kr}$  are  $N(1, .2^2)$ ,  $k = 1, 2$ ,  $r = 1, \dots, 4$ .

The phase function  $h_r(t)$ ,  $r = 1, \dots, p$  is constructed based on a functional PCA model. For scenarios 1-5, let  $a = (-3, -1, 1, 3) + N_p(0, .5^2 I_p)$ . The mean functions of  $h_r(t)$ ,  $r = 1, \dots, 4$ , are either not crossed

$$m_r^u(t) = \frac{6 \exp\{a_r(t+3)/6\} - 1}{e^{a_r} - 1} - 3,$$

or crossed

$$\begin{aligned} m_r^c(t) &= I(t \leq 0) \left( \frac{3 \exp\{a_r(t+3)/6\} - 1}{e^{a_r} - 1} - 3 \right) \\ &+ I(t > 0) \left( \frac{3 \exp\{-a_r(t+3)/6\} - 1}{e^{a_r} - 1} \right). \end{aligned}$$

As for the covariance structures of  $h_r(t)$ , we consider three cases: (i)  $\phi_1^1(t) = 0$ ,  $\phi_2^1(t) = 0$ , that is, identity, (ii)  $\phi_1^2(t) = \sin(\pi t/3)/\sqrt{3}$ ,  $\phi_2^2(t) = \cos(\pi t/2)/\sqrt{3}$ , (iii)  $\phi_1^3(t) = \{\cos(\pi t/3) + 1\}/3$ ,  $\phi_2^3(t) = \sin(\pi t/3)/\sqrt{3}$ .

Then, we set  $h_r^{j,u}(t) = m_r^u(t) + \xi_1 \phi_1^j(t) + \xi_2 \phi_2^j(t) + \epsilon(t)$  for crossed warping and  $h_r^{j,c}(t) = m_r^c(t) + \xi_1 \phi_1^j(t) + \xi_2 \phi_2^j(t) + \epsilon(t)$  for uncrossed warping, where  $\xi_k \sim N(0, \theta_k^2)$ ,  $\theta_1 = 1$  and  $\theta_2 = .5$  and  $\epsilon(t)$  are white noise with variance  $0.03^2$ .

The six scenarios are introduced below. The first four scenarios have different mean warping structures within an observation, for which the proposed method aims for:

- Scenario 1: both clusters have identical amplitude patterns and the common covariance structure for phase variation,  $S_{1r}(h_r^{1,u}(t))$  versus  $S_{1r}(h_r^{1,c}(t))$ . That is, the two clusters are the same except the crossing pattern of the mean warping functions.
- Scenario 2: both clusters have identical amplitude patterns, but have different covariance structure as well as mean for phase variation,  $S_{1r}(h_r^{2,u}(t))$  versus  $S_{1r}(h_r^{3,c}(t))$ .
- Scenario 3: clusters have different amplitude patterns, and different mean but identical covariance structure for phase variation,  $S_{1r}(h_r^{3,u}(t))$  versus  $S_{2r}(h_r^{3,c}(t))$ .
- Scenario 4: clusters have different amplitude patterns, and have different covariance structure for phase variation,  $S_{2r}(h_r^{2,u}(t))$  versus  $S_{1r}(h_r^{3,c}(t))$ .
- Scenario 5: (null case) clusters have different amplitude patterns, but have identical phase variation.  $S_{2r}(h_r^{2,u}(t))$  versus  $S_{1r}(h_r^{2,u}(t))$ .
- Scenario 6: (a component-wise warping model) let  $b$  be an equal-sized grid of  $n = 25$  on  $[-3, 3]$  with added Gaussian noise with variance  $0.05$ . Then,  $\tilde{a}$  is the randomly permuted version of  $b$  differently for each component and each cluster. For the  $r$ th component curve,  $r = 1, \dots, 4$ , the  $j$ th observation,  $j = 1, \dots, n$  is  $Y_j(t) = S_{1j}(h_j^{1,u}(t))$  for Cluster 1 and  $Y_j(t) = S_{2j}(h_j^{1,u}(t))$  for Cluster 2, where  $h_j^{1,h}(t) = m_r^u(t)$  except that  $a_r$ ,  $r = 1, \dots, 4$  is replaced with  $\tilde{a}_j$ ,  $j = 1, \dots, n$ .

For a numerical summary for comparison of the clustering outcomes, we use the Adjusted Rand Index (ARI) score proposed by Rand (1971) and Hubert and Arabie (1985). The index is essentially the proportion of the agreements between two partitions. Its value is always between  $-1$  and  $1$ , with  $-1$  corresponding to the perfect disagreement and  $1$  to the perfect agreement. The value of zero can be achieved between two purely random partitions.

Choosing the number of clusters, say  $K$ , is a critical problem in any clustering analysis. In this simulation study, we run the clustering methods with  $K = 2, 3, 4$ , and used Calinski–Harabasz (CH) index (Calinski and Harabasz, 1974) to choose

**Table 1**

Mean warping errors and mean clustering accuracies, measured in adjusted Rand Index, for two choices of warping methods and two choices of clustering methods

Scenario	Warping	MSE( $\hat{h}$ )	Chiou		CCToolbox	
			ARI <sub>2</sub>	ARI <sub>K</sub>	ARI <sub>2</sub>	ARI <sub>K</sub>
1	F-R	0.0864	1.0000	0.9676	0.9984	0.8732
	Ramsay	0.0981	0.9888	0.7422	0.9936	0.8782
2	F-R	0.2013	0.9231	0.7960	0.9400	0.8388
	Ramsay	0.2881	0.4781	0.4053	0.6424	0.4743
3	F-R	0.1768	0.7846	0.6709	0.8546	0.7150
	Ramsay	0.2568	0.2442	0.2627	0.4614	0.3715
4	F-R	0.2130	0.9225	0.8095	0.9228	0.8206
	Ramsay	0.3064	0.3289	0.2808	0.4489	0.3526
5	F-R	0.2632	-0.0010	-0.0039	0.0001	-0.0015
	Ramsay	0.3644	0.0197	0.0146	0.0203	0.0158
6	F-R	N/A	0.0013	0.0115	0.0031	0.0041
	Ramsay	N/A	-0.0026	-0.0027	-0.0035	-0.0049

$K$ . The CH index is analogous to the  $F$ -ratio statistic, that is, the ratio of between cluster variation and within cluster variation, each adjusted for respective degrees of freedom.

Table 1 summarizes the simulation results based on 50 repetitions. First, we report the mean squared error of the estimated warping function computed as

$$\text{MSE}(\hat{h}) = (np)^{-1} \sum_{i=1}^n \sum_{r=1}^p \int_{\mathcal{I}} (\hat{h}_{ir}(t) - h_{ir}(t))^2 dt,$$

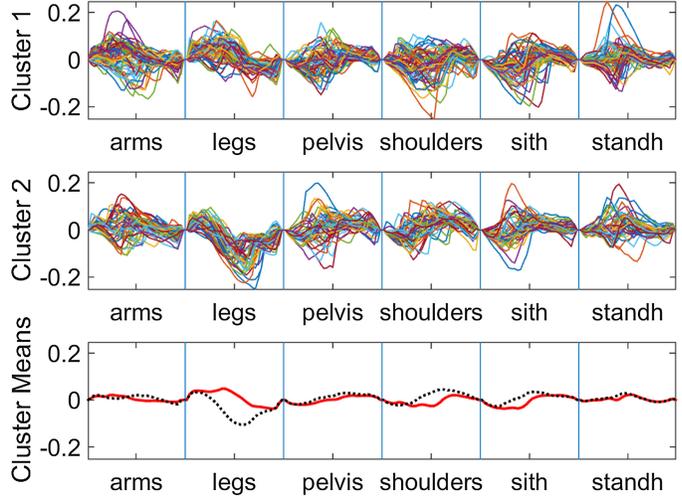
to compare the accuracy in estimation of warping profiles. Note that in Scenario 6, we do not report MSE, since it is not relevant to the considered model. Also, we report ARI when we fix  $K = 2$  (ARI<sub>2</sub>) and also when we let CH index choose  $K$  (ARI<sub>K</sub>).

The findings can be summarized as follows:

- F-R warping is more accurate than Ramsay's method, in terms of MSE, which affects the subsequent clustering results.
- Under the null settings, that is, Scenarios 5 and 6, the proposed method does not yield a meaningful clustering, which is desirable.
- Chiou's clustering is better in Scenario 1 while CCToolbox is better in Scenario 3, especially in ARI<sub>2</sub>. With F-R warping, the difference between clustering methods is not substantial in Scenarios 2 and 4.
- Performance patterns are similar when  $K$  is fixed and when  $K$  is adaptively chosen, the latter generally yields a bit lower ARI.

#### 4.2. An Application to Multivariate Growth Curves

We analyze the growth data, introduced in Section 1 with a goal to identify subgroups that have different comparative acceleration/delay patterns among different body measurement growths. We standardize the original curves and consider the



**Figure 3.** Clusters found in the girls. Deformation curves are shown. Cluster mean curves show differences in relative growths among legs, shoulders, and sitting heights.

velocity as the response curves. In any clustering analysis, verifying whether found clusters are “meaningful” is not a straightforward problem. In this work, in order to verify the clusters, we randomly split data into two groups, say A and B, run the clustering in A, and using the cluster centers in A, we assign the individuals in B. Then, we run the clustering algorithm in B and compare the two cluster assignments in B in terms of ARI. We repeated this 50 times. As for the number of clusters, we choose  $K = 2$ , which yields the lowest ARIs. For the girls, mean ARIs with Chiou and CCToolbox are, respectively, 0.6205 (0.1745) and 0.2024 (0.2344) with standard deviation inside the parentheses. For the boys, the ARIs are 0.0859 (0.1707) and 0.0469 (0.0801), which indicates no meaningful clusters.

Results show that for girls, there are two clusters that show clearly different patterns in the warping functions. Main distinction is in the relative delay pattern in leg length, shoulders, and sitting, as shown in Figure 3. The girls in the first cluster ( $n = 60$ ) share the pattern that legs grow rapidly at the earlier ages while the growth spurt of legs of the girls in the second cluster ( $n = 42$ ) relatively constant and longer lasting.

It is of interest to see how the above cluster solution is compared with the clustering with original velocity curves or the warped velocity curves that contain (supposedly) only amplitude variation left after warping. We also compare results based on the component warping approach. Table 2 shows the disagreements among the cluster solutions based on different types of curves. As a reference, Figure 4 shows the found clusters with original curves. It is clear that the usual clustering with multivariate functional curves has a fundamentally different goal than our problem presented here.

## 5. Discussion

The problem of phase variation and curve alignment is a long-standing issue in functional data analysis. Although phase variation is commonly observed in practice, the treatment is

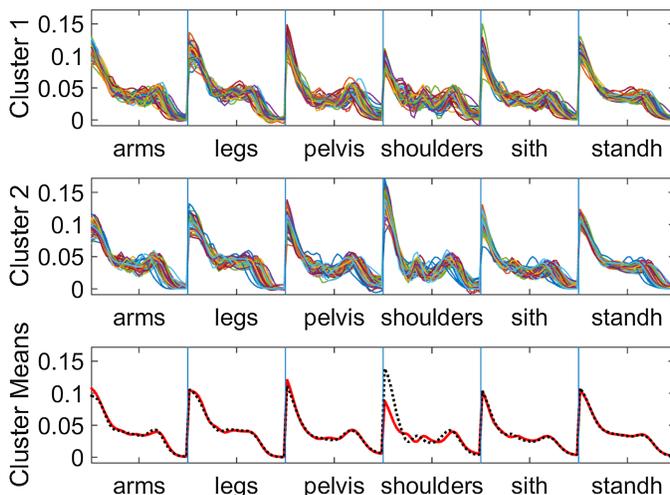
**Table 2**  
Cross tabulated clustering results

Girls	Original	Amplitude	C. Amplitude	C. Phase	Total
Phase	33	27	33	27	33
	21	21	21	21	21
Total	54	48	54	48	55

Boys	Original	Amplitude	C. Amplitude	C. Phase	Total
Phase	31	8	27	12	28
	43	17	31	29	35
Total	74	25	58	41	63

rather limited and phase variation is rarely the focus of the main analysis. Notably, many statistical procedures developed for FDA are inherently linear, relying on the linear representation of an object in a Euclidean space, while phase variation is an exception. Recent interests in geometrical approaches in data analysis will become more relevant (e.g., Kurtek et al., 2012), though its exploration in the context of FDA is still in its infancy. We demonstrated the importance and utility of proper consideration of the problem in the context of clustering multivariate functional data. In contrast to standard clustering problems, our targets are implicitly defined groups based on the patterns of phase variation only, which could be contaminated by amplitude variations. Our solution exploits the parametrization of the curves with respect to amplitude and phase variations. We used the growth data as an illustration and believe that a similar approach will be appropriate to analyze many biological and physiological signals collected in modern times.



**Figure 4.** Clusters found in the girls with original velocity curves. Mean curves show that the differences between the two clusters are mostly in the shoulder growths.

## 6. Supplementary Materials

The accompanying matlab codes to implement our method as in Section 4 are available with this article at the *Biometrics* website on Wiley Online Library.

## ACKNOWLEDGEMENTS

We thank the Editor, Associated Editor, and two anonymous referees for their careful reading and useful suggestions that lead to many improvements in the article.

## REFERENCES

- Brunel, N. and Park, J. (2014). Removing phase variability to extract a mean shape for juggling trajectories. *Electronic Journal of Statistics* **8**, 1848–1855.
- Calinski, T. and Harabasz, J. (1974). A dendrite method for cluster analysis. *Communications in Statistics* **3**, 1–27.
- Chiou, J.-M., Chen, Y.-T., and Yang, Y.-F. (2014). Multivariate functional principal component analysis: A normalization approach. *Statistica Sinica* **24**, 1571–1596.
- Chiou, J.-M. and Li, P.-L. (2007). Functional clustering and identifying substructures of longitudinal data. *Journal of the Royal Statistical Society, Series B* **69**, 679–699.
- Ferraty, F. and Vieu, P. (2006). *Nonparametric Functional Data Analysis*. Springer Series in Statistics. Theory and Practice. New York: Springer.
- Gaffney, S. (2004). *Probabilistic Curve-Aligned Clustering and Prediction with Mixture Models*. PhD thesis, University of California, Irvine.
- Gasser, T. and Kneip, A. (1995). Searching for structure in curve sample. *Journal of the American Statistical Association* **90**, 1179–1188.
- Gasser, T., Müller, H. G., Köhler, W., Molinari, L., and Prader, A. (1984). Nonparametric regression analysis of growth curves. *Annals of Statistics* **12**, 210–229.
- Gervini, D. and Gasser, T. (2004). Self-modeling warping functions. *Journal of the Royal Statistical Society, Series B* **66**, 959–971.
- Hadjipantelis, P. Z., Aston, J., Müller, H. G., and Evans, J. P. (2015). Unifying amplitude and phase analysis: A compositional data approach to functional multivariate mixed-effects modeling of Mandarin Chinese. *Journal of the American Statistical Association* **110**, 545–559.
- Hubert, L. and Arabie, P. (1985). Comparing partitions. *Journal of Classification* **2**, 193–218.
- Ieva, F., Paganoni, A. M., Pigoli, D., and Vitelli, V. (2013). Multivariate functional clustering for the morphological analysis of electrocardiograph curves. *Journal of the Royal Statistical Society Series C* **62**, 401–418.
- Jacques, J. and Preda, C. (2012). Model-based clustering for multivariate functional data. *Computational Statistics and Data Analysis* **1**, 1–24.
- James, G. (2007). Curve alignment by moments. *Annals of Applied Statistics* **1**, 480–501.
- Kneip, A. and Ramsay, J. O. (2008). Combining registration and fitting for functional models. *Journal of the American Statistical Association* **103**, 1155–1165.
- Kurtek, S., Srivastava, A., Klassen, E., and Ding, Z. (2012). Statistical modeling of curves using shapes and related features. *Journal of the American Statistical Association* **107**, 1152–1165.

- Liu, X. and Müller, H. G. (2004). Functional convex averaging and synchronization for time-warped random curves. *Journal of the American Statistical Association* **99**, 687–699.
- Liu, X. and Yang, C. (2009). Simultaneous curve registration and clustering for functional data. *Computational Statistics and Data Analysis* **53**, 1361–1376.
- Peng, J., Paul, D., and Müller, H. G. (2014). Time-warped growth processes, with applications to the modeling of boom-bust cycles in house prices. *Annals of Applied Statistics* **8**, 1561–1582.
- Ramsay, J. O. and Li, X. (1998). Curve registration. *Journal of the Royal Statistical Society, Series B* **60**, 351–363.
- Ramsay, J. O. and Silverman, B. W. (2002). *Applied Functional Data Analysis: Methods and Case Studies*. Springer Series in Statistics. New York: Springer-Verlag.
- Ramsay, J. O. and Silverman, B. W. (2005). *Functional Data Analysis*. Springer Series in Statistics, second edition. New York: Springer.
- Rand, W. M. (1971). Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical Association* **66**, 846–850.
- Sangalli, L., Secchi, P., Vantini, S., and Vitelli, V. (2010). K-means alignment for curve clustering. *Computational Statistics and Data Analysis* **54**, 1219–1233.
- Sheehy, A., Gasser, T., Molinari, L., and Largo, R. H. (1999). An analysis of variance of the pubertal and mid-growth spurts for length and width. *Annals of Human Biology* **26**, 309–331.
- Sheehy, A., Gasser, T., Molinari, L., and Largo, R. H. (2000). Contribution of growth phases to adult size. *Annals of Human Biology* **27**, 281–298.
- Slaets, L., Claeskens, G., and Hubert, M. (2012). Phase and amplitude-based clustering for functional data. *Computational Statistics and Data Analysis* **56**, 2360–2374.
- Tang, R. and Müller, H. G. (2008). Pairwise curve synchronization for functional data. *Biometrika* **95**, 875–889.
- Tang, R. and Müller, H. G. (2009). Time-synchronized clustering of gene expression trajectories. *Biostatistics* **10**, 32–45.
- Tucker, J. D., Wu, W., and Srivastava, A. (2013). Generative models for functional data using phase and amplitude separation. *Computational Statistics and Data Analysis* **61**, 50–66.
- Wang, K. and Gasser, T. (1997). Alignment of curves by dynamic time warping. *Annals of Statistics* **25**, 1251–1276.
- Wang, K. and Gasser, T. (1999). Synchronizing sample curves non-parametrically. *Annals of Statistics* **27**, 439–460.

Received August 2015. Revised March 2016.

Accepted April 2016.