Biometrics XX, 1–26 DOI: xx
2020

Cross-Component Registration for Multivariate Functional Data, With Application to Growth Curves

Cody Carroll[1*], Hans-Georg Müller[1], and Alois Kneip[2]

[1] Department of Statistics, University of California, Davis, One Shields Ave, Davis, CA 95616, U.S.A.

[2] Department of Economics, Universität Bonn, Regina-Pacis-Weg 3, 53113 Bonn, Germany

email: cjcarroll@ucdavis.edu

Summary: Multivariate functional data are becoming ubiquitous with the advance of modern technology and are substantially more complex than univariate functional data. We propose and study a novel model for multivariate functional data where the component processes are subject to mutual time warping. That is, the component processes exhibit a similar shape but are subject to systematic phase variation across their time domains. To address this previously unconsidered mode of warping, we propose new registration methodology which is based on a shift-warping model. Our method differs from all existing registration methods for functional data in a fundamental way. Namely, instead of focusing on the traditional approach to warping, where one aims to recover individual-specific registration, we focus on shift registration across the components of a multivariate functional data vector on a population-wide level. Our proposed estimates for these shifts are identifiable, enjoy parametric rates of convergence and often have intuitive physical interpretations, all in contrast to traditional curve-specific registration approaches. We demonstrate the implementation and interpretation of the proposed method by applying our methodology to the Zürich Longitudinal Growth data and study its finite sample properties in simulations.

Key words: Component processes; Functional data analysis; Growth curves; Multivariate functional data; Shift registration; Time warping.

This paper has been submitted for consideration for publication in *Biometrics*

1. Introduction

Multivariate functional data are often encountered in biological or chemical processes that are continuously measured for a group of subjects or observational units. Such processes arise in many longitudinal studies, especially in the biomedical sciences, ranging from growth studies to longitudinal measurements of protein levels during metabolic processes (Park and Ahn, 2017; Dubin and Müller, 2005). With the increasing ubiquity of multivariate functional data, the study of how to treat such data has recently become a very active field, in particular in the context of clustering (Brunel and Park, 2014; Jacques and Preda, 2014; Park and Ahn, 2017), functional regression (Chiou, 2012; Chiou et al., 2016), and in terms of general modeling of functional data (Claeskens et al., 2014; Di Salvo et al., 2015). Common approaches for analyzing multivariate functional data have focused on dimension reduction via multivariate functional principal components (MFPCA) (Zhou et al., 2008; Chiou et al., 2014; Happ and Greven, 2018) or decomposition into component-specific processes and their interactions (Chiou et al., 2016).

In applications such as growth curves, if we view multivariate longitudinal data as generated by an underlying p-dimensional smooth stochastic process, the component curves of the functional vector may exhibit mutual time warping. If left unchecked, such vector component warping may distort principal components and inflate data variance, while if handled properly, it may yield intuitive physical interpretations and a more parsimonious representation of the data. As far as we know, the idea of explicitly modeling time relations between component processes has not been considered yet for multivariate functional data, which are characterized by the availability of repeated observations of the multivariate process for a cohort of subjects, where one can therefore take advantage of the entire sample of repeated processes.

Typically, for each subject in longitudinal studies one has measurements on a grid of

time points, where the measurements are possibly contaminated with measurement errors. This situation is found in many longitudinal studies with multivariate measurements across the sciences, notably in growth studies (e.g., Han et al., 2018), and suggests to consider functional methods which are geared towards repeatedly sampled multivariate functional data. The analysis of the Zürich Longitudinal Growth study data motivated us to model such multivariate functional data by allowing the components to be mutually time-shifted against each other, as some components of growth may systematically precede others. We emphasize here that by no means our approach is limited to growth data and can be applied also when one has multivariate functional data where there is neither a well defined time origin nor an endpoint, such as is the case when analyzing multivariate blood protein time courses (e.g., Dubin and Müller, 2005).

In many situations, the component processes of multivariate functional data exhibit similarity in their shapes. A well-known instance of this phenomenon are longitudinal studies of children's growth, where the sizes of multiple body parts are measured over time. Each body part's component process follows the same general pattern of growth: a rapid rate of growth during infancy, which then slows to a roughly constant rate of growth in prepubescence until puberty, at which time the rate increases as the subject goes through a growth spurt, followed by a decrease in the rate to zero as an individual reaches adulthood (Gasser et al., 1984). The multivariate aspect of these growth curves allows us to compare the growth processes of different parts of the body. For example, it may be that arms undergo their growth spurt earlier in life than legs do. It is an interesting biological question to search for a common growth process that ordinates the timings of growth spurts across body parts. Another situation where this phenomenon arises is in the afore-mentioned measurements of protein levels during metabolic processes. Certain biological functions are associated with

peaks and valleys of certain protein levels and their relative timings expose the order of the underlying enzymatic mechanisms at work (Dubin and Müller, 2005).

Data from the Zürich Longitudinal Growth Study were used previously to investigate the timing of growth spurts across body parts using a phase-clustering model (Park and Ahn, 2017). In this paper we study the same data but instead of clustering, our emphasis is rather examining phase variations in the component growth velocity curves to establish time relations. In particular, we investigate mutual time warping in the derivatives across the components of the multivariate functional processes during a growth spurt window, as derivatives are more informative about human growth than the growth curves themselves. Specifically, we assume a model which uses relative time shifts between component processes to establish their pairwise time relations. Information about the relative shifts between pairs of components may then be combined to inform the full system of relative timings across body parts.

The organization of this paper is as follows. Section 2 motivates and establishes a shift-warping model for the cross-component warping problem. In Section 3 and 4 we estimate the proposed model components in the pairwise and general settings, respectively. An application to human growth curves is discussed in Section 5. A simulation study is explored in Section 6, which illustrates the stability of the method even in the presence of nuisance peaks and sizeable measurement error. Section 7 contains theoretical results, with accompanying proofs appearing in the appendix. Specifically, we find that under a quadratic curvature assumption, one attains parametric rates of convergence for cross-component shift estimates.

2. A Shift-Warping Model for Multivariate Time Relations

To illustrate the idea of mutual component warping, consider the growth velocities for a handful of representative children in the Zürich Longitudinal Study (Fig. 1), which will be revisited in its entirety in Section 5. We consider pubertal growth, i.e. growth curves are

evaluated in the interval $\mathcal{I} = [9, 18]$ ranging from 9 to 18 years. Each child has four growth velocity curves, each corresponding to a different body part. The peaks represent the moment of maximal rate of growth and can be used as a crude measure of the timing of pubertal growth spurt for that modality. For ease of viewing we mark these locations in time with vertical dashed lines in Figure 1.

A key observation is to recognize that regardless of when the child underwent puberty, the ordering of the spurts is consistent: legs undergo their growth spurts first, then arm length and standing height roughly together, followed by sitting height. This pattern in pubertal spurts was briefly discussed in the descriptive growth studies of Sheehy et al. (1999) and suggests that there is a population-wide mutual component warping occuring across the four modalities. We emphasize that the children displayed are typical of the entire sample; to align these curves across components, we propose a shift-warping model, as it is a simple and interpretable method for cross-component alignment of growth data.

[Figure 1 about here.]

Note that the time differences across modalities are relatively consistent across children, despite individual differences in the age of the pubertal onset. This is worth highlighting for two reasons: (1) It motivates the estimation of a fixed population-wide set of shift parameters, and (2) it shows that cross-component registration makes sense even in the presence of subject-specific time warping, which is the usual mode of warping considered in univariate functional data.

From a methodological point of view our approach builds on basic ideas in the literature on parametric and semi-parametric modeling of growth and related phenomena. In applied work on human growth, empirical studies often utilize parametric models (Milani, 2000). One of the most popular classes of models has been proposed by Preece and Baines (1978); for a recent application, see e.g. Banik et al. (2017). All of these models make use of shift parameters

θ_{ij} to capture the main differences in individual timings. For p-dimensional multivariate functional data, $(X_{i1}(t), \ldots, X_{ip}(t))^T$, $i = 1, \ldots, n$, that we consider here on a domain $\mathcal{I}$ that covers the pubertal period, an extension of the existing models to the multivariate case is as follows. For some function G and some additional parameter vectors $\boldsymbol{\xi}_{ij}$ one posits that, with time shifts θ_{ij}, the growth curve for the j-th component of the i-th subject has the form

$$X_{ij}(t) = G(\boldsymbol{\xi}_{ij}, t - \theta_{ij}), \quad j = 1, \ldots, p, \ i = 1, \ldots, n, \ t \in \mathcal{I}, \tag{1}$$

where previously only cases with $p = 1$ have been considered.

As fully parametric specifications were found to lack accuracy, various semi-parametric extensions have been proposed for the one-dimensional case. For example, for standing height, in the case $p = 1$, Kneip and Engel (1995) assumed a shape-invariant model with $G(\boldsymbol{\xi}_{ij}, t - \theta_{ij}) = \xi_{ij;2} f(\xi_{ij;1}(t - \theta_{ij}) + \xi_{ij;3}$ for real valued parameters $\xi_{ij;1}, \xi_{ij;2}, \xi_{ij;3}$, and an unknown real valued function f which is estimated from the data. The k-mean alignment introduced by Sangalli et al. (2010) may be seen as a generalization of this framework, where it is assumed that the population can be decomposed into K disjoint clusters, and individual functions belonging to each cluster can be approximately described by a shape-invariant model with respect to a cluster-specific template function f_g, $g \in \{1, \ldots, K\}$.

In the following we assume that growth data follow a multivariate and flexible version of models of type (1), under the natural assumption that the shift parameters θ_{ij} can be decomposed in the form $\theta_{ij} = \theta_i + \theta_j$, where θ_i is specific for the individual, while θ_j is specific for the component. Then (1) may be rewritten in the form

$$X_{ij}(t) = G(\boldsymbol{\xi}_{ij}, t - \theta_i - \theta_j) \equiv G^*(\boldsymbol{\xi}^*_{ij}, t - \theta_j), \quad \text{where } \boldsymbol{\xi}^*_{ij} = (\boldsymbol{\xi}_{ij}, \theta_i). \tag{2}$$

Motivating our alignment procedure is that for a given individual i the component functions $X_{i1}(t), \ldots, X_{ip}(t)$ can be made more similar when removing the different shift parameters $\theta_1, \ldots, \theta_p$. The most favorable situation arises if shifts constitute the only important difference between components such that $\boldsymbol{\xi}_{ij} \equiv \boldsymbol{\xi}_i$ is independent of $j = 1, \ldots, p$. Then with $Z_i(s) :=$

$G^*(\xi_i^*, s)$ we arrive at

$$X_{ij}(t) = Z_i(t - \theta_j), \quad j = 1, \ldots, p,\ i = 1, \ldots, n,\ t \in \mathcal{I}, \tag{3}$$

so that $E\int_{\mathcal{I}}(X_{ij}(t+\theta_j) - X_{il}(t+\theta_l))^2 dt = E\int_{\mathcal{I}}(Z_i(t) - Z_i(t))^2 dt = 0$ for all $j, l \in \{1, \ldots, p\}$, $j \neq l$; to apply this argument will require some pre-processing in order to eliminate scale differences between the different components (see Section 5).

In the context of growth curves, subject specific warping based on nonparametric warping functions $h_i(t)$ has been studied extensively (Gasser et al., 1990; Kneip and Engel, 1995; Gervini and Gasser, 2004). There is a vast literature on different registration procedure determining warping functions for univariate curve data. Higher dimensional problems of subject-specific registration have been considered through the lens of elastic shape analysis (Srivastava et al., 2010; Srivastava and Klassen, 2016), or reduced to the problem of aligning a univariate curve generated from the component curves (Ramsay et al., 2014). It is now seen from (2) and (3) that in our context such functions $h_i(t)$ do not play any role and may simply be part of the parameter set $\boldsymbol{\xi}_i$. We therefore emphasize that in the non-traditional warping framework presented here, the pertinent issues are fundamentally different from those considered in subject-specific warping that are discussed in the cited articles.

For readers well-versed in traditional registration methods, it may nevertheless be tempting to use a suitable technique to register the curves across individuals for each modality before attempting to register across components. These traditional approaches provide subject-specific warping functions, i.e. one warping function per individual, and raise several issues, ranging from non-identifiability of the warping functions to a lack of satisfactory theoretical results. A first difficulty arises from the fact that in our context, as well as in many other studies, analysis focuses only on a compact subinterval $\mathcal{I}$ of the range of all relevant ages (0 up to approximately 20 years). But nonparametric warping functions are always defined as strictly monotonic functions from a compact interval $\mathcal{I}$ onto $\mathcal{I}$. If $\mathcal{I} = [a, b]$ this implies

$h_i(a) = a$ and $h_i(b) = b$ for all individuals $i = 1, \ldots, n$. For our setup with $a = 9$ (years) this cannot be seen as a tenable assumption. Therefore fitting nonparametric warping functions has to be based on a much larger age interval which faces the risk that highly variable, irregular growth in early childhood may heavily influence results.

More importantly, there are conceptual problems of standard alignment which, for example, were discussed in Kneip and Ramsay (2008) and Marron et al. (2015). A basic issue is a lack of identifiability of "true" underlying warping functions. As shown in Wagner and Kneip (2019) a unique identification is only possible under very restrictive conditions on the functional structure of sample curves. This is, however, not an issue in the context of our approach where the goal is to estimate the shift parameters $\theta_1, ..., \theta_p$ and not the individual warping functions h_i which are not meaningful for our purposes. A more detailed discussion of this issue is presented in the context of the Zürich data in the online appendix. That it bypasses dealing with individual warping functions is a strength of our method and allows us to side-step the myriad of problems associated with subject-specific warping.

It is especially noteworthy that we obtain a $\sqrt{n}$-rate of the estimated time shifts to their targets under mild regularity conditions (see Section 7). Such fast convergence rates cannot be obtained in traditional warping approaches, since these focus on subject-specific warps rather than component-specific warping and therefore require identification of n time alignments, where n is the sample size, whereas in our approach there are only p components that need to be considered, where p is the fixed dimension of the multivariate process. This then enables us to obtain much faster convergence than is possible in traditional warping.

A further distinction between cross-component warping as proposed here and the common subject-specific time warping is that subject-specific registration traditionally views the presence of individual warping functions as a nuisance characteristic of the data to be accounted for in order to correctly analyze underlying functional features of interest, for example curves

will be registered first before conducting a functional principal component analysis (FPCA). In contrast, the perspective on cross-component warping and the shift parameters $\theta_1, ..., \theta_p$ is to provide insight into inter-component relationships and, when applicable, an essential aspect of multivariate functional data that is of genuine interest rather than a nuisance.

3. Bivariate Cross-Component Registration

3.1 *Pairwise-Shift Estimation*

We introduce here the main idea of registering different component times across modalities, which we call Cross-Component Registration (XCR). As explained in the previous section, XCR differs in key aspects from traditional curve warping, which is also known as curve registration or alignment (Ramsay and Silverman, 2005; Wang et al., 2016; Kneip and Gasser, 1992; Silverman, 1995), as it aims at a situation where the component curves of a multivariate functional process are time-shifted versions of one another. We will not give here a comprehensive overview of traditional warping methods; this is a varied field where one encounters complications rather quickly (Gervini and Gasser, 2004; Liu and Yang, 2009; Tang and Müller, 2008; Marron et al., 2015). A key difference is that instead of estimating n subject-specific warping functions, which align curves across individuals and the determination of which is the goal of traditional curve warping methods, our new approach targets a p-vector of shift parameters for the case of p-dimensional functional data, where p is the fixed dimension of the functional vector. These component-wise shifts are then applied in the same way across all subjects to mutually align the component curves.

In the following, we write $(X_1(t), \ldots, X_p(t))^T$ to represent the generic underlying multivariate process and $(X_{i1}(t), \ldots, X_{ip}(t))^T$, $i = 1, \ldots, n$, for a sample of realizations of the functional vector obtained for n subjects. One may assume *a priori* smoothness of curves or may preprocess the data with a smoothing method if the curves are subject to measurement

error. In this subsection we consider the case of multivariate functional data with just $p = 2$ component curves to introduce the main ideas, and will then discuss the extension to $p > 2$. To fix the idea, consider a sample of bivariate functional processes, writing $\{X_{i1}(t), X_{i2}(t)\}_{i=1}^{n}$ for the observed i.i.d. realizations of the bivariate process (X_1, X_2) and assume that the domain of both component processes is a compact interval $\mathcal{T} = [0, T]$. As a criterion for alignment and to determine the optimal shift, we aim to minimize the $\mathcal{L}^2$-distance between functions on a subinterval $\mathcal{I} \subset \mathcal{T}$; see the discussion below. Using a simple shift-warp family under the $\mathcal{L}^2$-norm allows for a straightforward and clear interpretation of the relationship between two components and has been used previously in the context of shape-invariant modeling (Härdle et al., 1990; Kneip and Engel, 1995; Silverman, 1995).

Specifically, we aim for the optimal value of the parameter τ, the pairwise *cross-component (XC) shift* as the minimizer of

$$\Lambda(\tau) = E \int_{\mathcal{I}} [X_1(t) - X_2(t - \tau)]^2 dt, \tag{4}$$

with associated sample version

$$L_n(\tau) = \frac{1}{n} \sum_{i=1}^{n} \int_{\mathcal{I}} [X_{i1}(t) - X_{i2}(t - \tau)]^2 dt \tag{5}$$

and sample-based shift parameter estimate

$$\widehat{\tau} = \operatorname*{argmin}_{\tau} L_n(\tau), \tag{6}$$

targeting $\tau_0 = \operatorname{argmin}_{\tau} \Lambda(\tau)$.

Integrating over a subinterval $\mathcal{I}$ rather than the whole interval is a device that is necessary in order to ensure that both the shifted and unshifted curves are defined on the domain of integration. If we did not specify a suitable subinterval $\mathcal{I} \subset [0, T]$ that stays away from both 0 and T, shifting a curve forward or backward may result in a subinterval of integration in which one of the curves is defined while the other is not, making it impossible to compute their $\mathcal{L}^2$-distance. Specifically, we partition the data domain $\mathcal{T}$ into three disjoint intervals $\mathcal{T} = R_1 \cup \mathcal{I} \cup R_2$, where $\mathcal{I} = [r_1, r_2]$ is the subinterval of integration and $R_1 = [0, r_1)$ and

$R_2 = (r_2, T]$ are the remaining intervals on the boundary. Note that this partitioning implies that the magnitude of pairwise shift estimates cannot exceed the length of the relevant remainder interval, depending on the direction of the shift. This subtlety suggests that the choice of subinterval of integration $\mathcal{I}$ is not trivial and should be done carefully and data-adaptively.

3.2 *Subinterval Selection*

We propose the following guidelines for subinterval selection: $\mathcal{I}$ should be chosen to (1) include the critical features of the sample curves, and (2) avoid censoring estimates of pairwise shifts. For example, in our application to the Zürich data, we choose $\mathcal{I}$ to range from the earliest age of pubertal onset to the age of adulthood. Doing so ensures the inclusion of the main pubertal growth spurt peaks which are the structural features to be aligned across components (Gasser and Kneip, 1995). Unreasonable estimates may occur if the subinterval is too small, as an inappropriately narrow window may discard the features to be aligned for a subset of individuals.

The problem of subinterval selection was discussed previously in Kneip and Engel (1995) and we follow their convention to seek an "overlapping interval" across all individuals, described as follows. Individual intervals $\mathcal{J}_i$ are chosen such that information about structural landmarks for the i^{th} individual are entirely contained in $\mathcal{J}_i$. Then the overlapping interval $\mathcal{J}$ is defined as $\mathcal{J} = \cup_i \mathcal{J}_i$ and guarantees that all individual's structural features are included. One can then either simply use this overlapping interval as the subinterval of integration, i.e., let $\mathcal{I} = \mathcal{J}$, or choose $\mathcal{I}$ such that $\mathcal{J} \subset \mathcal{I}$ and $\mathcal{I}$ has some relevant physical meaning. An example for the latter case is demonstrated in the data application of Section 5.

In the more general setting with more than two components, we will encounter several pairwise time shifts between sets of component curves. To distinguish between these, we write τ_{jk} to denote the relative time shift which moves component k to component j. Note

that the sample and population time shifts are symmetric in the sense that $\tau_{jk} = -\tau_{kj}$. The problem of estimating general cross-component shift parameters $\theta_1, ..., \theta_p$ can be solved after the estimation of all of the pairwise shift parameters τ_{jk} for $1 \leqslant j < k \leqslant p$, as discussed in the following section.

4. General Cross-Component Registration

We now extend the methodology for bivariate cross-component registration (XCR) to the case of p-variate multivariate functional processes, aiming to align more than two component functions. Assume we observe p-variate functional data $(X_{i1}(t), \ldots, X_{ip}(t))^T$ for $i = 1, \ldots, n$, now with $p > 2$. We search for a vector of global XC shifts, $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)$, such that when each modality $X_j(t),\ j = 1, \ldots, p$, is shifted by θ_j, all p curves are aligned. Here it is useful to introduce the idea of an underlying *latent process*.

To fix the ideas, consider only a single observation of simulated multivariate functional data where the components of the multivariate process are just time-shifted replicates. Figure 2 illustrates an example for $p = 4$. A simple approach would be to align the component curves by fixing one component curve and shifting the others via bivariate XCR to align them with the selected component. However, a major problem with this approach is that the resulting XC shifts depend on the choice of the fixed component.

[Figure 2 about here.]

These problems can be overcome by assuming that each curve is a shifted version of an unobserved latent component curve that is visualized with a solid line in Figure 2. The observed components are then time-shifted with respect to this latent component and the shifts will be subject to the constraint $\sum_{j=1}^{p} \theta_j = 0$, so that there is no net XC shift from the latent component curve. This assumption is necessary for the identifiability of the XC shifts.

A key observation is that there is a linear relationship between pairwise XC shifts, τ_{jk}, and the global XC shifts, θ_j and θ_k. Specifically the pairwise shifts can be expressed as the difference of two global shifts as shown in Eq. 7. Furthermore, the linear map L that maps the vector of global shifts to their pairwise counterparts is invertible. Thus, after estimating bivariate XC shifts τ_{jk} between component functions, we can infer the global XC vector θ, and importantly, the linear maps are invariant with respect to the choice of the latent process.

More specifically, the linear map L is given by:

$$\tau_{jk} = \theta_j - \theta_k, \quad j,\ k = 1, \ldots, p, \quad j < k \tag{7}$$

with constraint $\sum_{j=1}^{p} \theta_j = 0$, so that

$$\widehat{\boldsymbol{\tau}}^* = L(\boldsymbol{\theta}) = \mathbf{A}\boldsymbol{\theta}, \tag{8}$$

where $\widehat{\boldsymbol{\tau}}^* = (\tau^T, 0)^T = (\tau_{12}, \tau_{13}, \ldots, \tau_{(p-1)p}, 0)^T$ is the pairwise shift parameter vector stacked with 0, $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)^T$ is the global shift vector of each component function with the latent process and $\mathbf{A}$ is the matrix of the linear map, corresponding to the contrasts in (8). Note that $\mathbf{A}$ is of dimension $(p(p-1)/2 + 1) \times p$, and is always of full column rank. Specifically, we write

$$\mathbf{A} = \begin{pmatrix} 1 & -1 & 0 & 0 & 0 & 0 & \ldots & 0 & 0 \\ 1 & 0 & -1 & 0 & 0 & 0 & \ldots & 0 & 0 \\ 1 & 0 & 0 & -1 & 0 & 0 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 1 & -1 & 0 & 0 & 0 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & \ldots & 1 & -1 \\ 1 & 1 & 1 & 1 & 1 & 1 & \ldots & 1 & 1 \end{pmatrix}.$$

To implement this approach, we must first estimate the stacked vector of bivariate XC

shifts,

$$\widehat{\boldsymbol{\tau}}^* = (\widehat{\boldsymbol{\tau}}^T, 0)^T = (\widehat{\tau}_{12}, \widehat{\tau}_{13}, \ldots, \widehat{\tau}_{(p-1)p}, 0)^T, \tag{9}$$

leading to the model $\widehat{\boldsymbol{\tau}}^* = \mathbf{A}\boldsymbol{\theta} + \varepsilon$, where ε is a vector of random noise with mean 0 and finite variance. Once the pairwise shifts $\widehat{\tau}_{jk}$ are obtained, global shifts $\boldsymbol{\theta}$ can be estimated as

$$\widehat{\boldsymbol{\theta}} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\widehat{\boldsymbol{\tau}}^* \tag{10}$$

by ordinary least squares. The p component curves will be aligned (to the latent curve) once they are time-shifted with their respective estimated global XC shifts, $\widehat{\boldsymbol{\theta}}$.

5. Application to the Zürich Longitudinal Growth Study

From 1954 to 1978, a longitudinal study on human growth and development was conducted at the University Children's Hospital in Zürich. Modalities of growth that were longitudinally measured on a dense regular time grid include standing height, sitting height, arm length, and leg length, so that the resulting data can be naturally viewed as multivariate functional data (Gasser et al., 1984, 1989). The raw trajectories of the $p = 4$ component processes for the children measured are displayed in Figure 3, which also indicates the measurement grid. Component curves are initially observed on the domain $\mathcal{T} = [0, 20]$, which can be artificially extended to the right by assuming measurements stay constant in adulthood, as for almost all subjects growth processes terminate before age 20.

We are especially interested in the timing of pubertal growth spurts, which occur for all individuals between ages 9 and 18 typically. We are using this time window as the subinterval of integration, $\mathcal{I}$, in accordance with the guidelines of Section 2. A common way to study growth velocities is to examine the derivatives of the growth curves instead of the curves themselves (Gasser et al., 1984). The growth velocities have a peak during puberty, with the apex representing the instant when an individual's growth rate is at its maximum. Previous analysis of human growth curves indicates that there is a difference in the ways that boys

and girls undergo puberty (Gasser et al., 1984; Eiben et al., 2005). For example, it is widely known that girls begin puberty at younger ages than boys do on average. Accordingly, for the subsequent analysis we separate boys and girls and for brevity display only the results for boys. We estimate the growth velocities, i.e., the derivatives of the growth trajectories, via local weighted linear smoothing using the package `fdapace` (Chen et al., 2020).

[Figure 3 about here.]

[Figure 4 about here.]

[Figure 5 about here.]

Because different body parts have different physical sizes, their velocities are also on different scales. We eliminate the majority of this amplitude variation by dividing each function by the total area under the curve, resulting in "relative velocities" for each modality. Relative velocities have been previously used in the growth curve literature (see, e.g. Sheehy et al. (1999)) and allow for the comparison of variables that are at different scale. Figure 4 shows the rescaled derivative estimates for the four growth modalities that we consider (growth velocities of standing and sitting heights, and of arm and leg lengths). After this pre-processing, we now have multivariate functional data with component functions such as those shown for a typical individual in Figure 5.

When we apply the proposed shift model to the growth velocities of the four growth modalities of the Zürich data, we obtain the following estimated global XC shifts:

[Table 1 about here.]

One can interpret these shift parameters in a pairwise manner. For example, legs tend to undergo their growth spurts roughly half a year years before arms do ($\widehat{\tau}_{34} = \widehat{\theta}_4 - \widehat{\theta}_4 \approx -0.5$) and sitting height trails roughly half a year behind standing height ($\widehat{\tau}_{21} = \widehat{\theta}_2 - \widehat{\theta}_1 \approx 0.5$). Our shift estimates and their implied order of growth spurts is

consistent with what is known about human growth patterns, as discussed in the descriptive longitudinal studies of Cameron et al. (1982) and Sheehy et al. (1999).

[Figure 6 about here.]

We next investigate some individuals before and after component alignment for a demonstration of how XC alignment affects the curves. Figure 6 (top) shows three individuals who are representative of the "average" ordering of growth spurts across modalities, whereas Figure 6 (bottom) displays those who generally went through pubertal spurts for whom the different body parts were already in sync before alignment. Individuals like those shown in Figure 6 (bottom) for whom alignment moved component curves further away from each other were very rare, and it was common for most individuals to have reduced $\mathcal{L}^2$-distance between the component curves after alignment. To illustrate this further, we use the total cross-component $\mathcal{L}^2$-distance (XD) for an individual as a function of θ,

$$\mathrm{XD}_i(\theta) = \sum_{j<k} \int_{\mathcal{I}} (X_{ij}(t-\theta_j) - X_{ik}(t-\theta_k))^2 dt. \tag{11}$$

Figure 7 displays the distribution of the difference in total cross-component $\mathcal{L}^2$-distance before and after shifting, i.e., $\mathrm{XD}_i(0) - \mathrm{XD}_i(\widehat{\boldsymbol{\theta}})$. Here it is noteworthy that implementing component alignment reduced total $\mathcal{L}^2$-distance in the sample by about 40%.

[Figure 7 about here.]

6. Simulation Study

6.1 *Finite Sample Performance*

We studied the performance of the proposed XC approach in several simulation settings, where we included varying levels of noise, using the base curve $Z(t) = 20 - .5t + 30e^{-\frac{(t-25)^2}{72}}$ on $t \in \mathcal{T} = [0, 50]$ as the underlying process dictating the common shape of the component curves. We then generated a 4-dimensional process with components of the form $X_j(t) =$

$Z(t-\theta_j)$ for $j=1,...,4$, where $\boldsymbol{\theta}=(-3,-1,1,3)$ on a data grid spanning $\mathcal{T}$ by increments of 0.5, displayed in Figure 8.

[Figure 8 about here.]

Starting with this base $4-$dimensional process, we generated $n=100$ noise contaminated versions as

$$X_{ij}(t_k)=X_j(t_k)+\zeta_{ij}\sin(\frac{\pi t_k}{5})+e_{ijk}, \tag{12}$$

where $\zeta_{ij}\overset{iid}{\sim}\mathcal{N}(0,\sigma_\zeta^2)$, $e_{ijk}\overset{iid}{\sim}\mathcal{N}(0,\sigma_e^2)$ and k indexes the 501 points on the data grid described above. The variance components σ_ζ^2 and σ_e^2 represent the amount of functional noise and measurement error we add to the curve, respectively. Adding the functional noise allows for the main peaks to vary in width and maximum height, while also potentially adding nuisance peaks at other time points. Such nuisance peaks can arise in the growth data, for example, in the form of a pre-pubertal spurt, which typically occurs around age 7 in children (Gasser et al., 1985).

After generating the curves on the grid described before, we then smoothed them with a local linear smoother on a finer final output grid spanning $\mathcal{T}$ by steps of size 0.1. A random sample of four such noisy processes is displayed in Figure 9. We evaluated $\mathcal{L}^2$-distances on the subinterval $\mathcal{I}=[10,40]$ and then applied XCR to estimate the shift vector θ. We replicated the simulation $B=1000$ times per setting and calculated the MSEs for each component at various levels of σ_ζ^2 and σ_e^2 to explore how XCR handles the two types of noise, with results displayed in Table 2.

We cap our functional noise level at $\sigma_\zeta^2=100$ since adding noise beyond this level allows the sine function to dominate the base shape, rendering it unrecognizable. Should one observe actual data with such high levels of functional noise, one likely would not suspect mutual time warping since the component processes no longer exhibit a similarity in their shapes. Estimates remain unbiased through all noise settings we consider, indicating that the XCR

method can handle shape and amplitude variations in the main signal, the presence of nuisance peaks, and sizeable measurement error.

[Figure 9 about here.]

[Table 2 about here.]

6.2 *Improved Fit when Modeling with XCR*

We demonstrate here the superior fit of curves aligned by cross-component registration prior to analysis through FPCA. We use the same base curve $Z(t) = 20 - .5t + 30e^{-\frac{(t-25)^2}{72}}$ on $t \in \mathcal{T} = [0, 50]$ as the underlying process dictating the common shape of the component curves but now let $\boldsymbol{\theta} = (-5, -2.5, 2.5, 5)$. As discussed in Section 2, the shift parameters themselves may be subject to noise. We emulate this error alongside the standard functional noise by generating contaminated component curves

$$X_{ij}(t_k) = X_j(t_k - \theta_j + \eta_{ij}) + \zeta_{ij} \sin(\frac{\pi t_k}{5}) + e_{ijk}, \tag{13}$$

where $\eta_{ij} \overset{iid}{\sim} \mathcal{N}(0, \sigma_\eta^2)$, $\zeta_{ij} \overset{iid}{\sim} \mathcal{N}(0, \sigma_\zeta^2)$, $e_{ijk} \overset{iid}{\sim} \mathcal{N}(0, 1)$, and k indexes the points on the data grid spanning $\mathcal{T}$ by increments of 0.5. Here the noise on the time domain is introduced through η_{ij}, while noise on the functional domain is controlled through ζ_{ij} and e_{ijk}, which correspond to a random amplitude sine wave and minor additive measurement error, respectively.

One can consider each of the component curves as a single noisy warped realization of the underlying latent curve Z. We may try to estimate the latent curve by viewing all the component processes for all subjects as a noisy sample of Z and then analyzing them through an established method such as FPCA. We expect that failing to account for the component warping will inflate variances and result in a suboptimal fit, since the cross-component warping masks the features of processes Z, and this is indeed what the simulations described next show. Curves were fit via FPCA using the first two eigenfunctions, with

and without incorporating XCR. Specifically, when incorporating XCR, a sample of $N = 100$ curves was generated, then XC shifts were estimated, whereupon curves were shifted according to these estimates, followed by a FPCA step applied to the thus aligned curves. The first two eigenfunctions were used to fit the sample of aligned curves, and after this fitting step the curves were shifted back by employing the shift estimates. To quantify the benefit of incorporating XCR, we obtained the integrated mean squared error for the approaches with and without XCR. The benefit of including XCR for various noise scenarios was measured through the percent decrease in integrated mean squared error for the sample.

This process was performed $B = 1000$ times under low, medium, and high functional noise settings ($\sigma_\zeta^2 = 25,\ 64,\ 100$), while letting the noise on the time domain start low and increase until it was on the same scale as the shifts themselves. Table 3 shows the average percent decreases across replications for various settings. The improvements in fit are relatively consistent across functional noise levels. It is noteworthy to observe that once the noise on the domain becomes comparable to that of the shifts themselves (i.e. $\sigma_\eta^2 > 0.5$), the benefits of XCR start to decrease. It conforms with expectation that when the within-subject time ordering is highly noise-contaminated, the benefits of performing XCR are lost. At such high shift noise levels there would be little incentive to perform XCR, as exploratory data inspection would then not likely indicate the presence of cross-component warping.

[Table 3 about here.]

An example of the difference in performances for a single set of curves can be seen in Figure 10. FPCA is ill-suited to account for the shift warping, a source of horizontal variation, through its eigenfunctions and their scores, which are geared towards representing vertical variation. In the presence of this horizontal variation, the estimated FPC scores then tend to over- or underestimate the actual amplitude variation, especially near the peaks, as seen in the left side of Figure 10. By accounting for component warping with XCR however, the

burden of modeling time domain variation is moved away from FPCA which can then focus on modeling amplitude variation without the confounding phase variation.

[Figure 10 about here.]

7. Theoretical Results

For bivariate Cross-Component Registration, a key finding is that the centered processes

$$Z_n(\tau) = \sqrt{n}[L_n(\tau) - \Lambda(\tau)]$$

converge weakly to a Gaussian limit process $Z(\tau)$, denoted by $Z_n(\tau) \rightsquigarrow Z(\tau)$, where L_n, Λ are as in (14), (15).

Specifically, we aim for the optimal value of the parameter τ, the pairwise *cross-component (XC) shift* as the minimizer of

$$\Lambda(\tau) = E \int_{\mathcal{I}} [X_1(t) - X_2(t-\tau)]^2 dt, \tag{14}$$

with associated sample version

$$L_n(\tau) = \frac{1}{n} \sum_{i=1}^{n} \int_{\mathcal{I}} [X_{i1}(t) - X_{i2}(t-\tau)]^2 dt \tag{15}$$

and sample-based shift parameter estimate

$$\widehat{\tau} = \underset{\tau}{\operatorname{argmin}}\, L_n(\tau), \tag{16}$$

targeting $\tau_0 = \operatorname{argmin}_\tau \Lambda(\tau)$.

To show the weak convergence, we require the following assumptions on Λ:

(P1) For any $\varepsilon > 0$, $\inf_{\tau : d(\tau,\tau_0) > \varepsilon} \Lambda(\tau) < \Lambda(\tau_0)$.

(P2) There exists $\eta > 0$, $C > 0$ and $\beta > 1$, such that, when $d(\tau, \tau_0) < \eta$, we have

$$\Lambda(\tau) - \Lambda(\tau_0) \geqslant C d(\tau, \tau_0)^\beta.$$

Assumption (P1) ensures that there exists a well-defined minimum, and assumption (P2) describes the local curvature of Λ at the true minimum τ_0, compare, e.g., Petersen and Müller (2016). We also require the following assumptions for the observed random processes:

(A1) $X_j(t)$ is continuously twice differentiable for $j = 1, \ldots, p$,

(A2) $E\left[\int_{\mathcal{I}} X_j^4(t)dt\right] < \infty$, for $j = 1, \ldots, p$,

(A3) $E\left[\int_{\mathcal{I}} X_j'^4(t)dt\right] < \infty$, for $j = 1, \ldots, p$.

These assumptions are standard in the literature. They were for example previously stipulated in Hall and Horowitz (2007) and enable us to obtain asymptotic covariance matrices for our estimates and to derive some crucial bounds.

LEMMA 1: *Under assumptions (P1), (P2), and (A1)-(A3), it holds that*

$$Z_n(\tau) \rightsquigarrow Z(\tau),$$

where $Z(\tau)$ *is a Gaussian process with mean zero and covariance* $G(\tau_1, \tau_2) = \int_{\mathcal{I}} \int_{\mathcal{I}} E((X_1(t) - X_2(t - \tau_1))^2 (X_1(s) - X_2(s - \tau_2))^2) dtds - \Lambda(\tau_1)\Lambda(\tau_2)$.

THEOREM 1: *Under the assumptions of Lemma 1, we have*

$$\widehat{\tau} - \tau_0 = O_p(n^{-1/2(\beta-1)}).$$

In particular, when $\beta = 2$, *the sequence* $\sqrt{n}(\widehat{\tau} - \tau_0)$ *is asymptotically normal with mean zero and variance* $V = 4[\int_{\mathcal{I}} E[(X_1(t) - X_2(t - \tau_0))X_2'(t - \tau_0)]^2 dt]/(\Lambda''(\tau_0))^2$ *where* $\Lambda(\tau) = E\int_{\mathcal{I}}(X_1(t) - X_2(t - \tau))^2 dt$.

The proof is in the appendix and utilizes results for M-estimators (Jain and Marcus, 1975; Van der Vaart and Wellner, 1996; van der Vaart, 1998). We note that when the local geometry around the minimum has a quadratic curvature, i.e. when $\beta = 2$, one obtains the parametric rate $n^{1/2}$.

Our main result for general Cross-Component Registration concerns the rate of convergence of the estimated shift parameter vector and its asymptotic distribution, as follows:

THEOREM 2: *Under assumptions (P1)-(P2) and (A1)-(A3)*

$$\widehat{\theta} - \theta_0 = O_p(n^{-1/2(\beta-1)}).$$

In particular, when $\beta = 2$*, the sequence* $\sqrt{n}(\widehat{\theta} - \theta_0)$ *is asymptotically normal with mean zero and covariance matrix*

$$\Sigma_p = \frac{1}{p^2} A^T \begin{bmatrix} V_{\tau_0}^{-1} E[\nabla m_{\tau_0} \nabla m_{\tau_0}^T] V_{\tau_0}^{-1} & 0 \\ 0 & 0 \end{bmatrix} A,$$

where

$m_{\tau_0} = \left[L_n(\tau_{12}), L_n(\tau_{13}), ..., L_n(\tau_{(p-1)p})\right]^T$ *and* V_{τ_0} *is the Hessian of* $\Lambda(\tau) = E(m_\tau)$ *at* τ_0.

8. Concluding Remarks

Cross-component registration seeks to address mutual component time warping that is often an issue for multivariate functional data arising from longitudinal studies in the biosciences. This issue does not manifest itself for univariate functional data. By focusing on time warping across components, and not on the traditional time warping between individual subjects, we are able to estimate population-wide time shift parameters with fast parametric rates of convergence and obtain a limit distribution under suitable assumptions.

This new cross-component time warping approach leads to insights about the relative timings of the component processes, which is of interest for the analysis of growth data and also other multivariate longitudinal data. After cross-component shift warps have been identified and incorporated into the model, common methods such as functional principal component analysis for multivariate processes can be expected to lead to more meaningful outputs and the resulting principal component component scores can be used for subsequent downstream analysis. The identification and estimation of the underlying latent process may also lead to a more parsimonious representations and is of interest in itself.

There are limitations of the framework we have established here. While the shift-warping model we develop in this paper is appropriate for certain applications such as the Zürich Longitudinal Growth Study data, the cross-component warping phenomena need not be restricted to shifts in general and may emerge in the form of non-linear distortions among

components. Using the shift-warping methodology in such a situation may or may not yield satisfactory results, depending on the nature of the actual time warping. If it is simple structured, a shift parameter may be a sufficient and parsimonious way to discover and approximate the component time relations, especially for practitioners who seek clear and concise interpretations. However the situation for more pronounced or complicated warps is less auspicious. When the data at hand exhibit more complicated component warping beyond shifts, a more flexible warping paradigm should be adopted. The nonlinearity of such cross-component distortions may suggest that such problems warrant an alternative metric to the L^2 norm.

In spite of this, we argue that the limitations of a shift-warping model are not necessarily tied to the general idea of cross-component registration which we have presented here. While in this paper we have used a shift-warping model to introduce the notion of cross-component registration, one can imagine more flexible extensions. The study of nonlinear warping models in cross-component registration is left for future research. Other potential problems of interest concern alternative representations of cross-component warping.

Acknowledgements

Research supported by NSF Grant DMS-1712864.

References

Banik, S. D., Salehabadi, S. M., and Dickinson, F. (2017). Preece-baines model 1 to estimate height and knee height growth in boys and girls from merida, mexico. *Food and Nutrition Bulletin* **38,** 182–195.

Brunel, N. J.-B. and Park, J. (2014). Removing phase variability to extract a mean shape for juggling trajectories. *Electron. J. Statist.* **8,** 1848–1855.

Cameron, N., Tanner, J., and Whitehouse, R. (1982). A longitudinal analysis of the growth of limb segments in adolescence. *Annals of Human Biology* **9,** 211–220.

Chen, Y., Carroll, C., Dai, X., Fan, J., Hadjipantelis, P. Z., Han, K., et al. (2020). *fdapace: Functional Data Analysis and Empirical Dynamics*. R package version 0.5.2.

Chiou, J.-M. (2012). Dynamical functional prediction and classification, with application to traffic flow prediction. *Annals of Applied Statistics* **6,** 1588–1614.

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., Yang, Y.-F., and Chen, Y.-T. (2016). Multivariate functional linear regression and prediction. *Journal of Multivariate Analysis* **146,** 301–312.

Claeskens, G., Hubert, M., Slaets, L., and Vakili, K. (2014). Multivariate functional halfspace depth. *Journal of the American Statistical Association* **109,** 411–423.

Di Salvo, F., Ruggieri, M., and Plaia, A. (2015). Functional principal component analysis for multivariate multidimensional environmental data. *Environmental and Ecological Statistics* **22,** 739–757.

Dubin, J. A. and Müller, H.-G. (2005). Dynamical correlation for multivariate longitudinal data. *Journal of the American Statistical Association* **100,** 872–881.

Eiben, O., Barabás, A., and Németh, Á. (2005). Comparison of growth, maturation, and physical fitness of hungarian urban and rural boys and girls. *Journal of Human Ecology* **17,** 93–100.

Gasser, T. and Kneip, A. (1995). Searching for structure in curve samples. *Journal of the American Statistical Association* **90,** 1179–1188.

Gasser, T., Kneip, A., Binding, A., Largo, R., Prader, A., and Molinari, L. (1989). Flexible methods for nonparametric fitting of individual and sample growth curves. *Auxology* **88,** 23–30.

Gasser, T., Kneip, A., Ziegler, P., Largo, R., and Prader, A. (1990). A method for determining the dynamics and intensity of average growth. *Annals of Human Biology* **17,** 459–474.

Gasser, T., Köhler, W., Müller, H.-G., Kneip, A., Largo, R., Molinari, L., et al. (1984). Velocity and acceleration of height growth using kernel estimation. *Annals of Human Biology* **11,** 397–411.

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.

Gasser, T., Müller, H.-G., Köhler, W., Prader, A., Largo, R., and Molinari, L. (1985). An analysis of the mid-growth spurt and of the adolescent growth spurt based on acceleration. *Annals of Human Biology* **12,** 129–148.

Gervini, D. and Gasser, T. (2004). Self-modeling warping functions. *Journal of the Royal Statistical Society: Series B* **66,** 959–971.

Hall, P. and Horowitz, J. L. (2007). Methodology and convergence rates for functional linear regression. *Annals of Statistics* **35,** 70–91.

Han, K., Hadjipantelis, P. Z., Wang, J.-L., Kramer, M. S., Yang, S., Martin, R. M., et al. (2018). Functional principal component analysis for identifying multivariate patterns and archetypes of growth, and their association with long-term cognitive development. *PloS one* **13,** e0207073.

Happ, C. and Greven, S. (2018). Multivariate functional principal component analysis for data observed on different (dimensional) domains. *Journal of the American Statistical Association* **113,** 649–659.

Härdle, W., Marron, J. S., et al. (1990). Semiparametric comparison of regression curves. *The Annals of Statistics* **18,** 63–89.

Jacques, J. and Preda, C. (2014). Model-based clustering for multivariate functional data. *Computational Statistics and Data Analysis* **71,** 92–106.

Jain, N. C. and Marcus, M. B. (1975). Central limit theorems for C(S)-valued random variables. *Journal of Functional Analysis* **19,** 216–231.

Kneip, A. and Engel, J. (1995). Model estimation in nonlinear regression under shape invariance. *The Annals of Statistics* pages 551–570.

Kneip, A. and Gasser, T. (1992). Statistical tools to analyze data representing a sample of curves. *Annals of Statistics* **20,** 1266–1305.

Kneip, A. and Ramsay, J. O. (2008). Combining registration and fitting for functional models. *Journal of the American Statistical Association* **103,** 1155–1165.

Liu, X. and Yang, M. C. K. (2009). Simultaneous curve registration and clustering for functional data. *Computational Statistics and Data Analysis* **53,** 1361–1376.

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

Milani, S. (2000). Kinetic models for normal and impaired growth. *Annals of human biology* **27,** 1–18.

Park, J. and Ahn, J. (2017). Clustering multivariate functional data with phase variation. *Biometrics* **73,** 324–333.

Petersen, A. and Müller, H.-G. (2016). Fréchet integration and adaptive metric selection for interpretable covariances of multivariate functional data. *Biometrika* **103,** 103–120.

Preece, M. and Baines, M. (1978). A new family of mathematical models describing the human growth curve. *Annals of human biology* **5,** 1–24.

Ramsay, J. O., Gribble, P., and Kurtek, S. (2014). Description and processing of functional data arising from juggling trajectories. *Electronic Journal of Statistics* **8,** 1811–1816.

Ramsay, J. O. and Silverman, B. W. (2005). *Functional Data Analysis*. Springer Series in Statistics. Springer, New York, second edition.

Sangalli, L. M., Secchi, P., Vantini, S., and Vitelli, V. (2010). K-mean alignment for curve

clustering. *Computational Statistics & Data Analysis* **54,** 1219–1233.

Sheehy, A., Gasser, T., Molinari, L., and Largo, R. (1999). An analysis of variance of the pubertal and midgrowth spurts for length and width. *Annals of human biology* **26,** 309–331.

Silverman, B. W. (1995). Incorporating parametric effects into functional principal components analysis. *Journal of the Royal Statistical Society: Series B* **57,** 673–689.

Srivastava, A., Klassen, E., Joshi, S. H., and Jermyn, I. H. (2010). Shape analysis of elastic curves in euclidean spaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **33,** 1415–1428.

Srivastava, A. and Klassen, E. P. (2016). *Functional and shape data analysis*. Springer.

Tang, R. and Müller, H.-G. (2008). Pairwise curve synchronization for functional data. *Biometrika* **95,** 875–889.

Van der Vaart, A. and Wellner, J. (1996). *Weak Convergence and Empirical Processes*. Springer, New York.

van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge University Press.

Wagner, H. and Kneip, A. (2019). Nonparametric registration to low-dimensional function spaces. *Computational Statistics & Data Analysis* **138,** 49–63.

Wang, J.-L., Chiou, J.-M., and Müller, H.-G. (2016). Functional data analysis. *Annual Review of Statistics and its Application* **3,** 257–295.

Zhou, L., Huang, J., and Carroll, R. (2008). Joint modelling of paired sparse functional data using principal components. *Biometrika* **95,** 601–619.

Supporting Information

Supporting Web Appendices referenced in Sections 2 and 5 are available with this paper at the Biometrics website on Wiley Online Library. An example dataset and code are also provided.

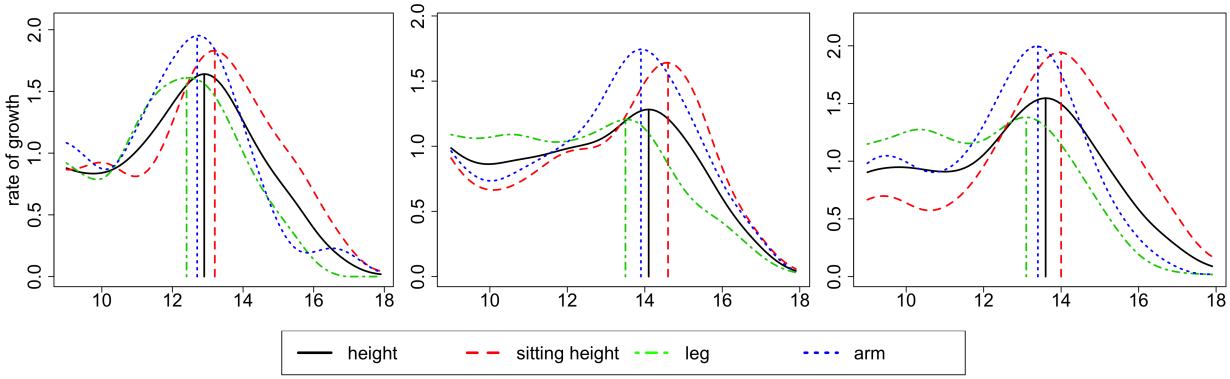

Figure 1. Three children's growth velocities for standing height (black, solid), sitting height (red, dashed), leg length (green, dot-dashed), and arm length (dotted). Peak velocity positions are marked with vertical lines and can be used as rough markers of pubertal onset for each modality. *This figure appears in color in the electronic version of this article, and any mention of color refers to that version.*

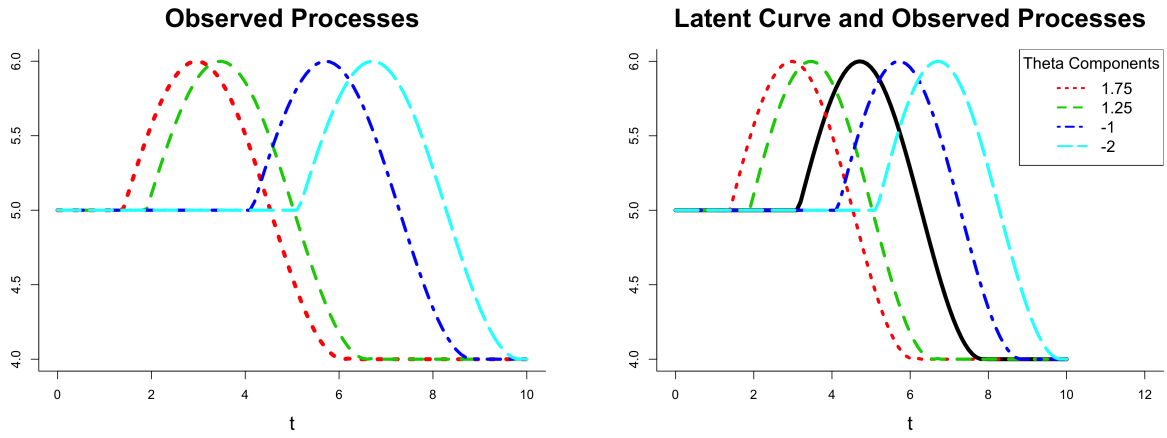

Figure 2. Observed components (dashed, left) and latent curve (solid, right) defined by identifiability constraint. *This figure appears in color in the electronic version of this article, and any mention of color refers to that version.*

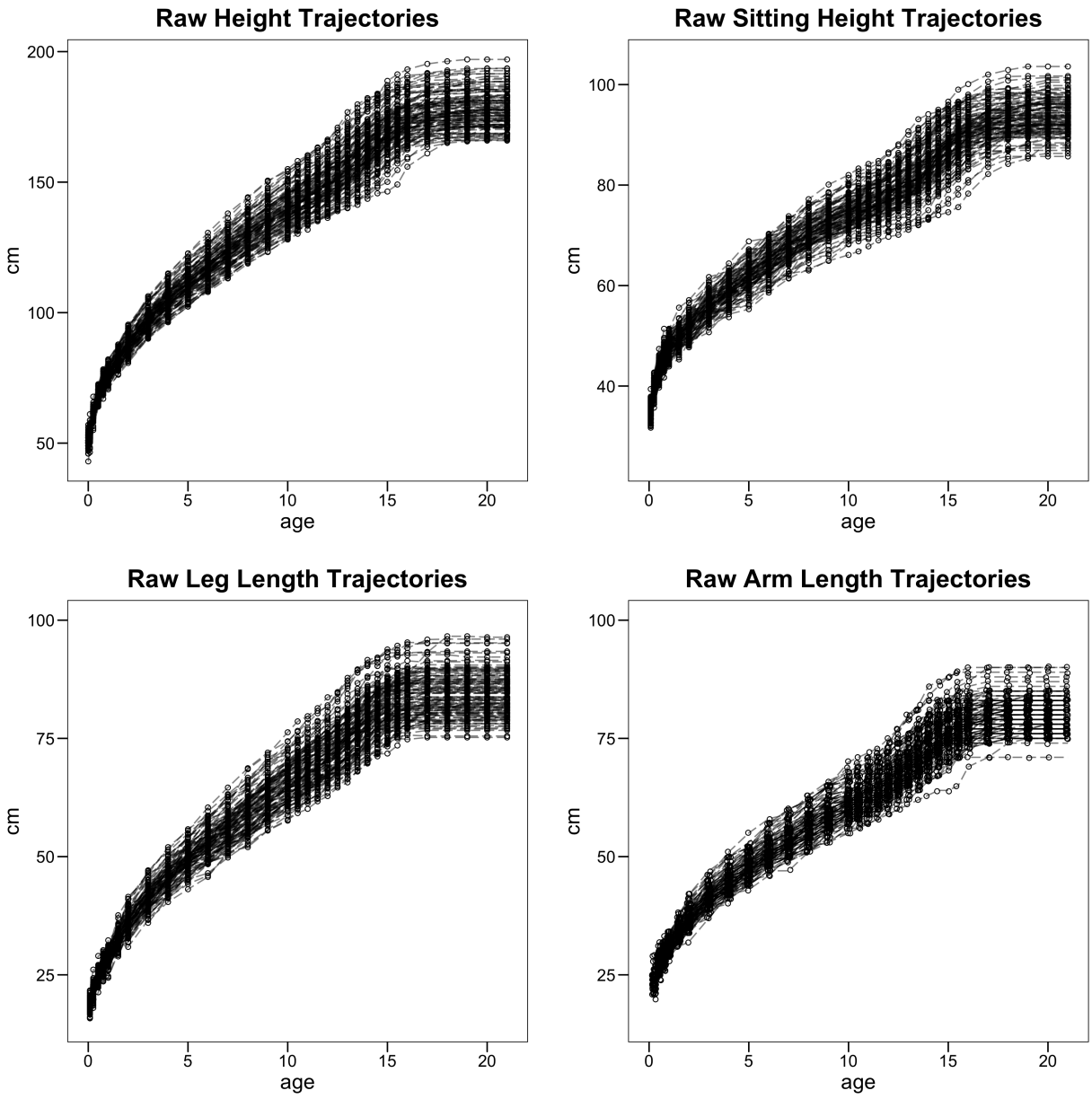

Figure 3. Raw growth trajectories for all Zürich boys.

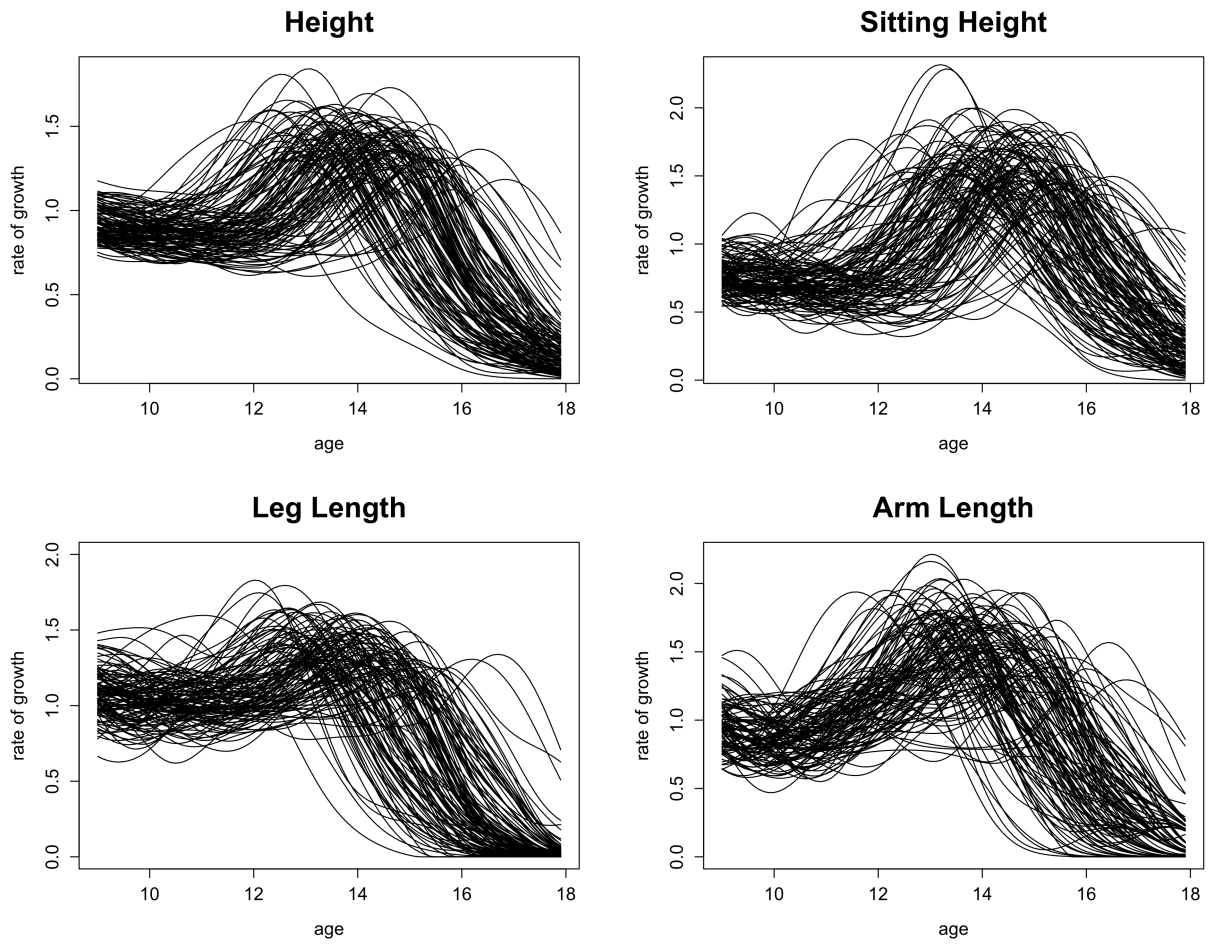

Figure 4. Scaled growth velocity curves for Zürich boys.

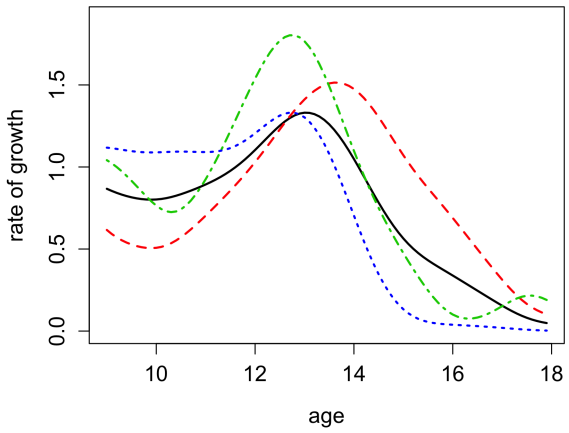

Figure 5. Example of growth component functions for an individual. Growth modalities are standing height (black, solid), sitting height (red, dashed), leg length (green, dot-dashed), and arm length (dotted). *This figure appears in color in the electronic version of this article, and any mention of color refers to that version.*

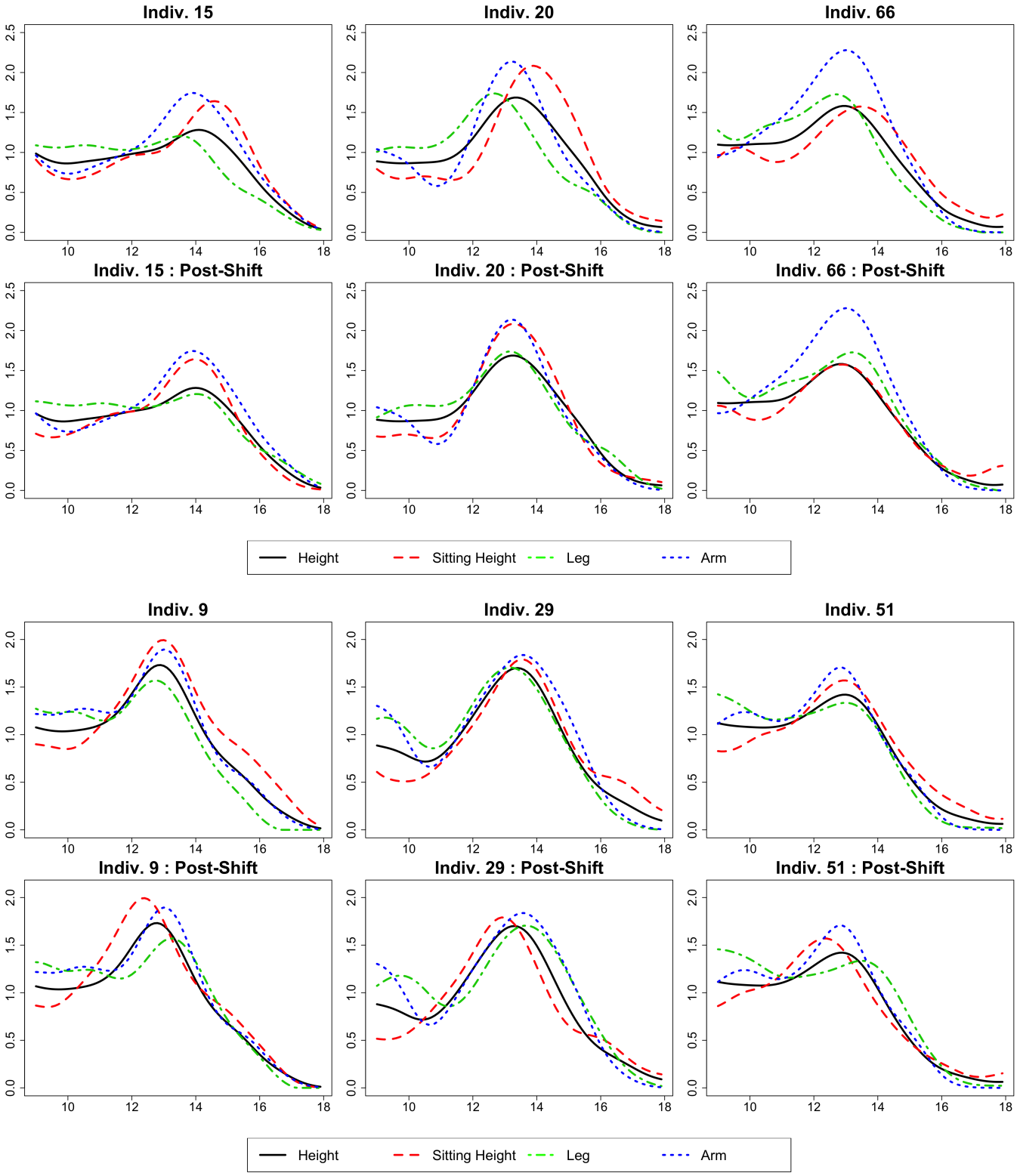

Figure 6. Well-aligned (top) and poorly aligned individuals (bottom) after component alignment. Growth modalities are standing height (black, solid), sitting height (red, dashed), leg length (green, dot-dashed), and arm length (dotted). *This figure appears in color in the electronic version of this article, and any mention of color refers to that version.*

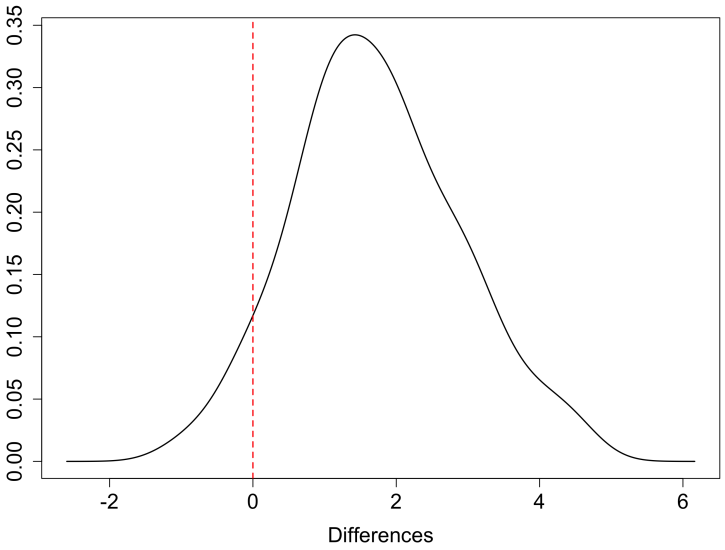

Figure 7. Kernel density estimate for the distribution of total $\mathcal{L}^2$-distance reduction after XCR. The dashed line indicates no change.

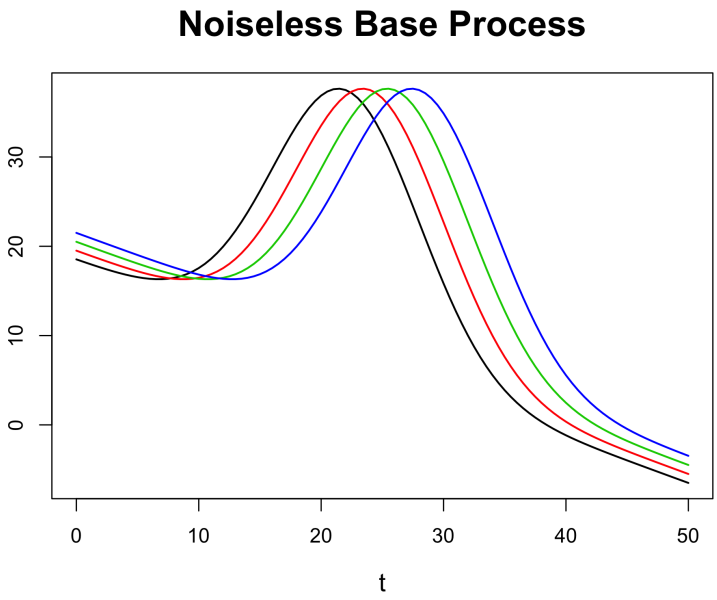

Figure 8. Base 4-dimensional process. *This figure appears in color in the electronic version of this article, and any mention of color refers to that version.*

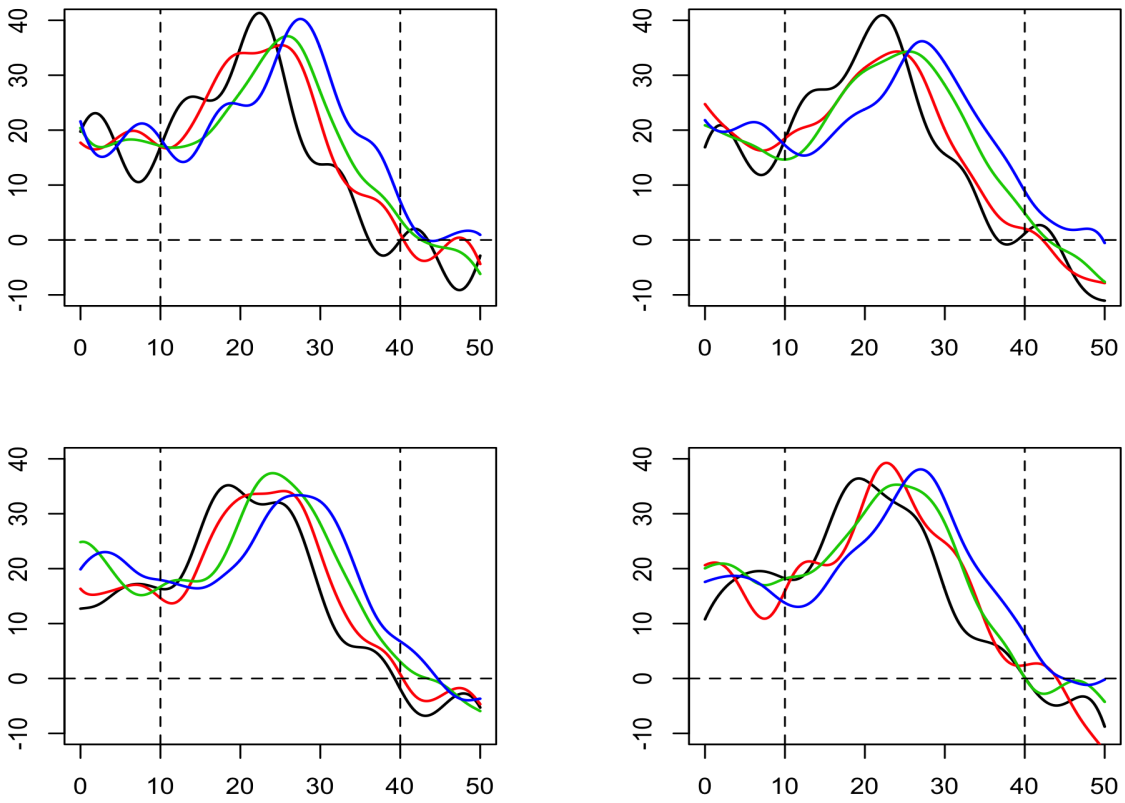

Figure 9. Four randomly selected simulated processes with under noise setting $\sigma_\zeta^2 = 25$, $\sigma_e^2 = 25$. The subinterval of integration $\mathcal{I} = [10, 40]$ is marked by dashed vertical lines. *This figure appears in color in the electronic version of this article, and any mention of color refers to that version.*

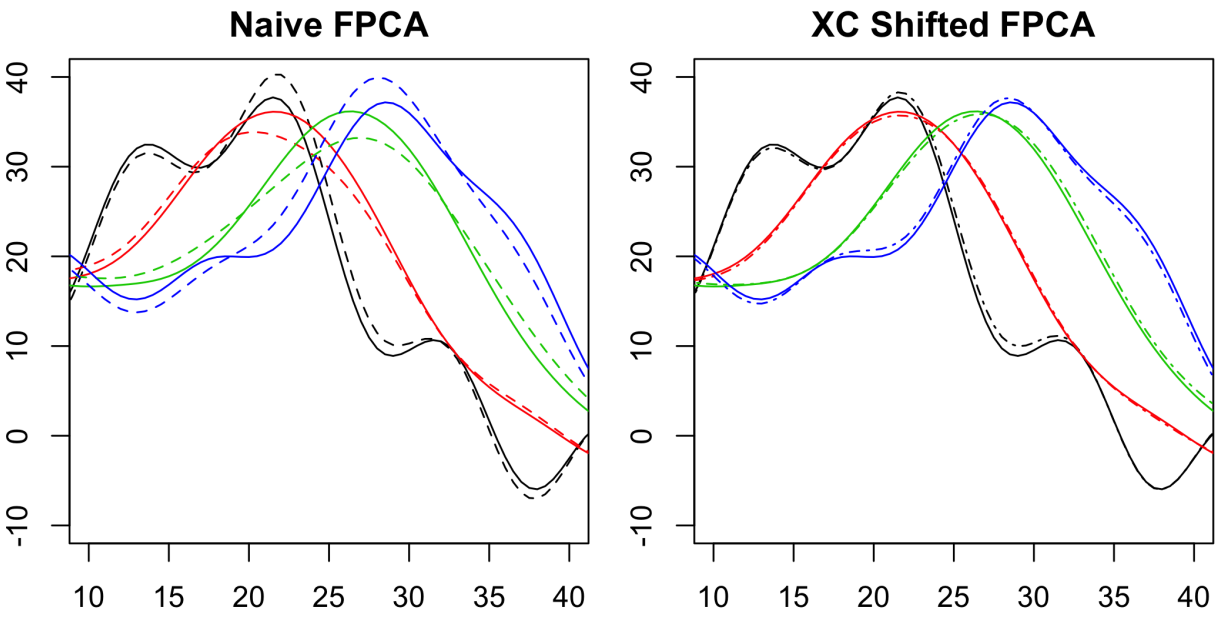

Figure 10. Differences in fit when using just naive FPCA (dashed, left) and FPCA and XCR together (dot-dashed, right), when $\sigma_\eta^2 = 0.25$, $\sigma_\zeta^2 = 25$. Solid lines represent the original data. *This figure appears in color in the electronic version of this article, and any mention of color refers to that version.*

Component	Modality	Estimate
θ_1	Height	-0.0875
θ_2	Sitting Height	-0.5850
θ_3	Leg Length	0.5825
θ_4	Arm Length	0.0900

Table 1

Estimated global XC shifts for Zürich boys. These estimates imply the following ordering of growth spurts: (1) leg, (2) height, (3) arm, (4) sitting height.

Noise Level	$\sigma_e^2 = 0$	$\sigma_e^2 = 9$	$\sigma_e^2 = 25$	$\sigma_e^2 = 64$	$\sigma_e^2 = 100$	component
$\sigma_\zeta^2 = 0$	0.0000	0.0010	0.0025	0.0036	0.0061	θ_1
	0.0000	0.0010	0.0038	0.0045	0.0066	θ_2
	0.0000	0.0012	0.0030	0.0040	0.0064	θ_3
	0.0000	0.0016	0.0026	0.0035	0.0060	θ_4
$\sigma_\zeta^2 = 9$	0.0004	0.0015	0.0034	0.0068	0.0051	θ_1
	0.0005	0.0016	0.0043	0.0072	0.0082	θ_2
	0.0009	0.0019	0.0038	0.0112	0.0062	θ_3
	0.0008	0.0019	0.0039	0.0109	0.0088	θ_4
$\sigma_\zeta^2 = 25$	0.0033	0.0035	0.0079	0.0037	0.0068	θ_1
	0.0042	0.0057	0.0070	0.0044	0.0069	θ_2
	0.0038	0.0059	0.0063	0.0055	0.0072	θ_3
	0.0036	0.0065	0.0087	0.0052	0.0076	θ_4
$\sigma_\zeta^2 = 64$	0.0018	0.0032	0.0035	0.0049	0.0091	θ_1
	0.0029	0.0037	0.0042	0.0055	0.0093	θ_2
	0.0031	0.0031	0.0053	0.0056	0.0085	θ_3
	0.0028	0.0032	0.0037	0.0054	0.0089	θ_4
$\sigma_\zeta^2 = 100$	0.0056	0.0069	0.0049	0.0122	0.0098	θ_1
	0.0080	0.0071	0.0094	0.0130	0.0125	θ_2
	0.0092	0.0081	0.0083	0.0121	0.0108	θ_3
	0.0052	0.0063	0.0100	0.0106	0.0123	θ_4

Table 2
MSEs of XC shift estimates under various noise settings. Estimates remain unbiased throughout.

Noise Level	$\sigma_\eta^2 = 0.1$	$\sigma_\eta^2 = 0.25$	$\sigma_\eta^2 = 0.5$	$\sigma_\eta^2 = 1$	$\sigma_\eta^2 = 2$
$\sigma_\zeta^2 = 25$	48.38	50.92	51.39	44.11	16.17
$\sigma_\zeta^2 = 64$	48.17	50.81	51.46	43.75	16.64
$\sigma_\zeta^2 = 100$	48.06	50.72	51.37	43.87	16.42

Table 3

Average percent decrease in IMSE after implementing XCR at various levels of contamination on the time (σ_η^2) and functional (σ_ζ^2) domains.