Biometrics WILEY

Cross-component registration for multivariate functional data, with application to growth curves

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

[1] Department of Statistics, University of California, Davis, California

[2] Department of Economics, Universität Bonn, Bonn, Germany

Correspondence
Cody Carroll, Department of Statistics, University of California, Davis, One Shields Ave, Davis, CA 95616.
Email: cjcarroll@ucdavis.edu

Abstract

Multivariate functional data are becoming ubiquitous with advances in 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 that 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.

KEYWORDS
component processes, functional data analysis, growth curves, multivariate functional data, shift registration, time warping

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, the scopes of which range from human growth to time courses of protein levels during metabolic processes (Dubin and Müller, 2005; Park and Ahn, 2017). With the increasing ubiquity of multivariate functional data, the study of how to treat such data has recently become a very active field, partic-

ularly 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). Common approaches for analyzing multivariate functional data have focused on dimension reduction via multivariate functional principal components (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 yet been considered for multivariate functional data, which allows one to take advantage of repeated observations of a multivariate process for a cohort of subjects.

Typically, for each subject in longitudinal studies one has measurements on a grid of time points, where recordings are possibly contaminated with measurement error. Often these measurements are multivariate, notably in growth studies (eg, Han *et al.*, 2018), which prompts consideration of functional methods that are geared toward repeatedly sampled multivariate functional data. The Zürich Longitudinal Growth Study 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.

The idea of warping across components is most pragmatic when the component processes of multivariate functional data exhibit similarity in their shapes. In the case of growth studies, each body part's component process follows the same general pattern: a period of rapid development during infancy which then slows to a roughly constant rate of growth until puberty, at which time the growth velocity peaks (ie, the pubertal growth spurt) before decreasing to zero as the subject 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 legs undergo their growth spurt earlier in life than arms do. It is an interesting biological question to search for a common process that ordinates the timings of growth spurts across body parts. Another situation where this phenomenon arises is in the above-mentioned recordings 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.

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). Our study uses the same data but instead emphasizes the analysis of 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 that 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. We emphasize that our approach, while motivated by growth data, is by no means limited to this application and can also be implemented for multivariate functional data that has neither a well-defined time origin nor an endpoint, as in the case of blood protein time courses (eg, Dubin and Müller, 2005).

The organization of this paper is as follows. Section 2 motivates and establishes a shift-warping model for the cross-component registration 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 that illustrates the stability of the method even in the presence of nuisance peaks and sizeable measurement error is explored in Section 6. 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 (Figure 1), which will be revisited in its entirety in Section 5. We consider pubertal growth, that is, 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 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 occurring across the four modalities. Note also that the time differences between modalities are relatively consistent across children, despite individual differences in the age of the pubertal onset. This is worth
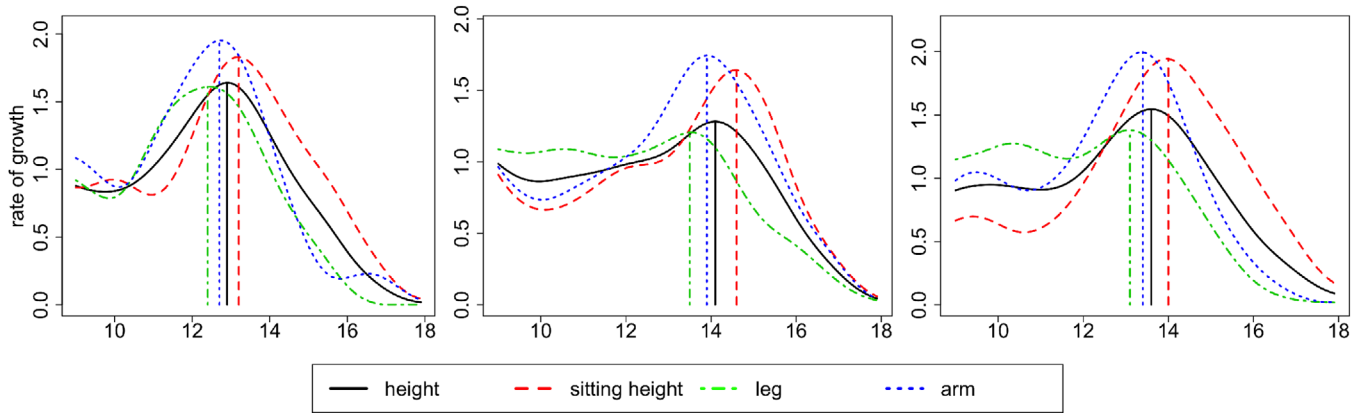
FIGURE 1 Three children's growth velocities for standing height (black, solid), sitting height (red, dashed), leg length (green, dot-dashed), and arm length (blue, dotted). Peak velocity positions are marked with vertical lines and can be used as rough markers of pubertal onset for each modality

highlighting for two reasons: (a) it motivates the estimation of a fixed population-wide set of shift parameters, and (b) 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. Subject-specific registration is a complex and varied field and we do not attempt to provide a comprehensive review here; for a recent overview of traditional warping methods, we refer less familiar readers to Marron *et al.* (2015) and Wang *et al.* (2016).

To register these curves across components, we propose a shift-warping model, which provides a simple and interpretable method for cross-component alignment of growth 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, for example, 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, $\{X_{i1}(t), \ldots, X_{ip}(t)\}^T$, $i = 1, \ldots, n$, which 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 jth component of the ith subject has the form

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

$$X_{ij}(t) = G(\boldsymbol{\xi}_{ij}, t - \theta_i - \theta_j) \equiv G^*(\boldsymbol{\xi}_{ij}^*, t - \theta_j),$$

$$\text{where } \boldsymbol{\xi}_{ij}^* = (\boldsymbol{\xi}_{ij}, \theta_i).$$ (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 $Z_i(s) := G^*(\boldsymbol{\xi}_i^*, s)$ we arrive at

$$X_{ij}(t) = Z_i(t - \theta_j), \quad j = 1, \ldots, p, \ i = 1, \ldots, n, \ t \in \mathcal{I},$$ (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 $j, l \in \{1, \dots, p\}$; to apply this argument will require some preprocessing in order to eliminate scale differences between the different components (see Section 5).

In the context of growth curves, subject-specific alignment based on nonparametric monotonic warping functions $h_i : \mathcal{I} \to \mathcal{I}$ has been studied extensively (Gasser *et al.*, 1990; Kneip and Gasser, 1995; Gervini and Gasser, 2004; Tang and Müller, 2008). 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 can be seen from (2) and (3) that in our context such functions h_i do not play any role and may simply be part of the parameter set $\boldsymbol{\xi}_i$. We therefore emphasize that in the nontraditional warping framework presented here, the pertinent issues are fundamentally different from those considered in the subject-specific warping framework discussed in the cited articles. In short, that it bypasses dealing with individual warping functions is a strength of our method and allows us to side-step the identifiability problems associated with subject-specific registration. A more detailed discussion of this matter in the context of the Zürich data can be found in the Supplement.

It is especially noteworthy that we obtain a $\sqrt{n}$-rate of convergence for 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, as these focus on individual 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. Of course, in some circumstances the model in (3) may just serve as a convenient approximation of a more nuanced warping relation between components. We discuss the potential for continuous analogues of cross-component shift-warping techniques in the Concluding Remarks.

A further distinction between cross-component warping as proposed here and the common subject-specific approach is that the latter 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, we argue that cross-component warping and the shift parameters $\theta_1, \dots, \theta_p$ provide insight into inter-component relationships and, when applicable, are 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 warping, which is also known as curve registration or alignment (Kneip and Gasser, 1992; Silverman, 1995; Ramsay and Silverman, 2005), as it aims at a situation where the component curves of a multivariate functional process are time-shifted versions of one another. A major difference is that instead of estimating n individual warping functions, which align curves across subjects 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. These component-wise shifts are then applied uniformly across all subjects to mutually align the component curves.

In the following, we write $(X_1, \dots, X_p)^T$ to represent the generic underlying multivariate process and $\{X_{i1}(t), \dots, X_{ip}(t)\}^T$, $i = 1, \dots, n$, for a sample of realizations of the functional vector. 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 independent and identically distributed 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 and Marron, 1990; Kneip and Gasser, 1995; Silverman, 1995).

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

minimizer of

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

with associated sample version

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

and sample-based shift parameter estimate

$$\hat{\tau} = \operatorname*{argmin}_{\tau} L_n(\tau), \quad (6)$$

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. To be precise, 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 (a) include the critical features of the sample curves, and (b) 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 that 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 Gasser (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 ith individual are contained entirely in $\mathcal{J}_i$. Then the overlapping interval $\mathcal{J}$ is defined as $\mathcal{J} = \cup_i \mathcal{J}_i$ and guarantees that all individuals' structural features are included. One can then either simply use this overlapping interval as the subinterval of integration, that is, 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 that 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 $\theta_1, \dots, \theta_p$ can be solved after the estimation of all of the pairwise shift parameters τ_{jk} for $1 \leq j < k \leq p$, as discussed in the following section.

4 | GENERAL CROSS-COMPONENT REGISTRATION

We now extend the methodology for bivariate XCR to the case of p-dimensional multivariate functional processes, aiming to align more than two component functions. Assume we observe p-variate functional data $\{X_{i1}(t), \dots, X_{ip}(t)\}^T$ for $i = 1, \dots, n$, now with $p > 2$. We search for a vector of global XC shifts, $\theta = (\theta_1, \dots, \theta_p)$, such that when each modality $X_j(t)$, $j = 1, \dots, p$, is shifted by θ_j, all p curves are aligned. Here it is useful to introduce the idea of an underlying *latent process*, which may be seen as the Z_i component in model (3).

To fix the idea, 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.

These problems can be overcome by assuming that each curve is a shifted version of an unobserved and unshifted latent component, visualized as the solid curve in Figure 2. The observed components are then time-shifted with respect to this latent component and the shifts are 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 shift estimates.
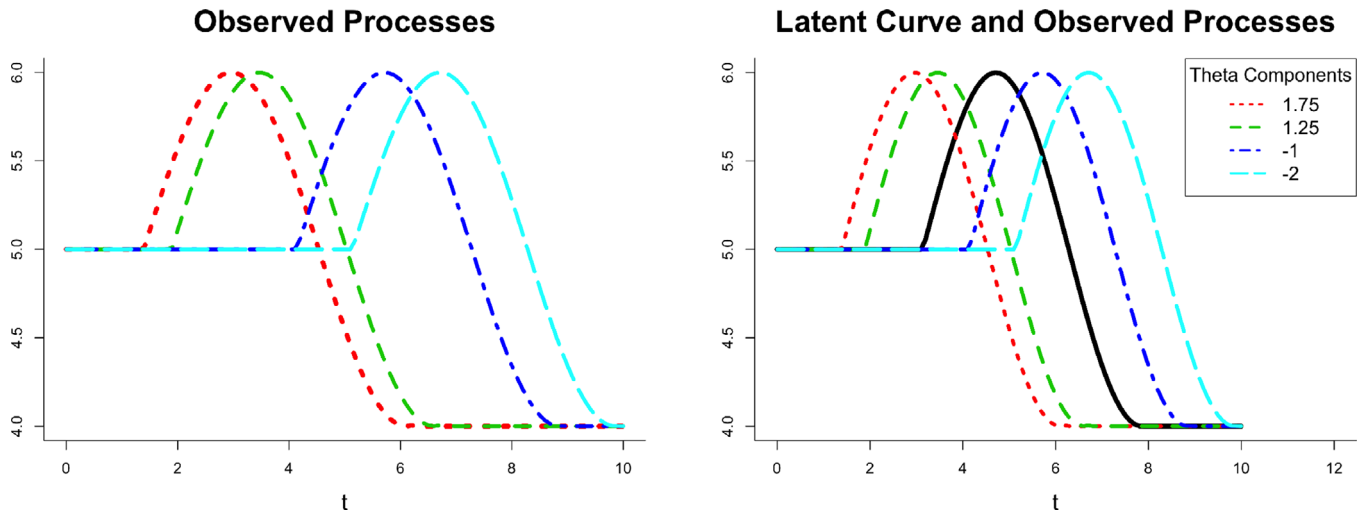
FIGURE 2 Observed components (dashed, left) and latent curve (solid, right) defined by identifiability constraint

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 Equation (7). Thus, after estimating bivariate XC shifts τ_{jk} between component functions, we can infer the global XC vector θ, and importantly, the linear map between the two is invariant with respect to the choice of the latent process.

More explicitly, the linear map L is given by

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

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

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

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

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

To implement this approach, we must first estimate the stacked vector of bivariate XC shifts, $\hat{\tau}^* = (\hat{\tau}^T, 0)^T = (\hat{\tau}_{12}, \hat{\tau}_{13}, \ldots, \hat{\tau}_{(p-1)p}, 0)^T$, leading to the model

$$\hat{\tau}^* = \mathbf{A}\theta + \varepsilon, \quad (9)$$

where ε is a vector of random noise with mean 0 and finite variance. Once the pairwise shifts $\hat{\tau}_{jk}$ are obtained, global shifts θ can be estimated as

$$\hat{\theta} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\hat{\tau}^* \quad (10)$$

by ordinary least squares. The p component curves will then be aligned (to the latent curve) once they are time-shifted with their respective estimated global XC shifts, that is, $X_{ij}(t + \hat{\theta}_j)$ for $j = 1, \ldots, p$.

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
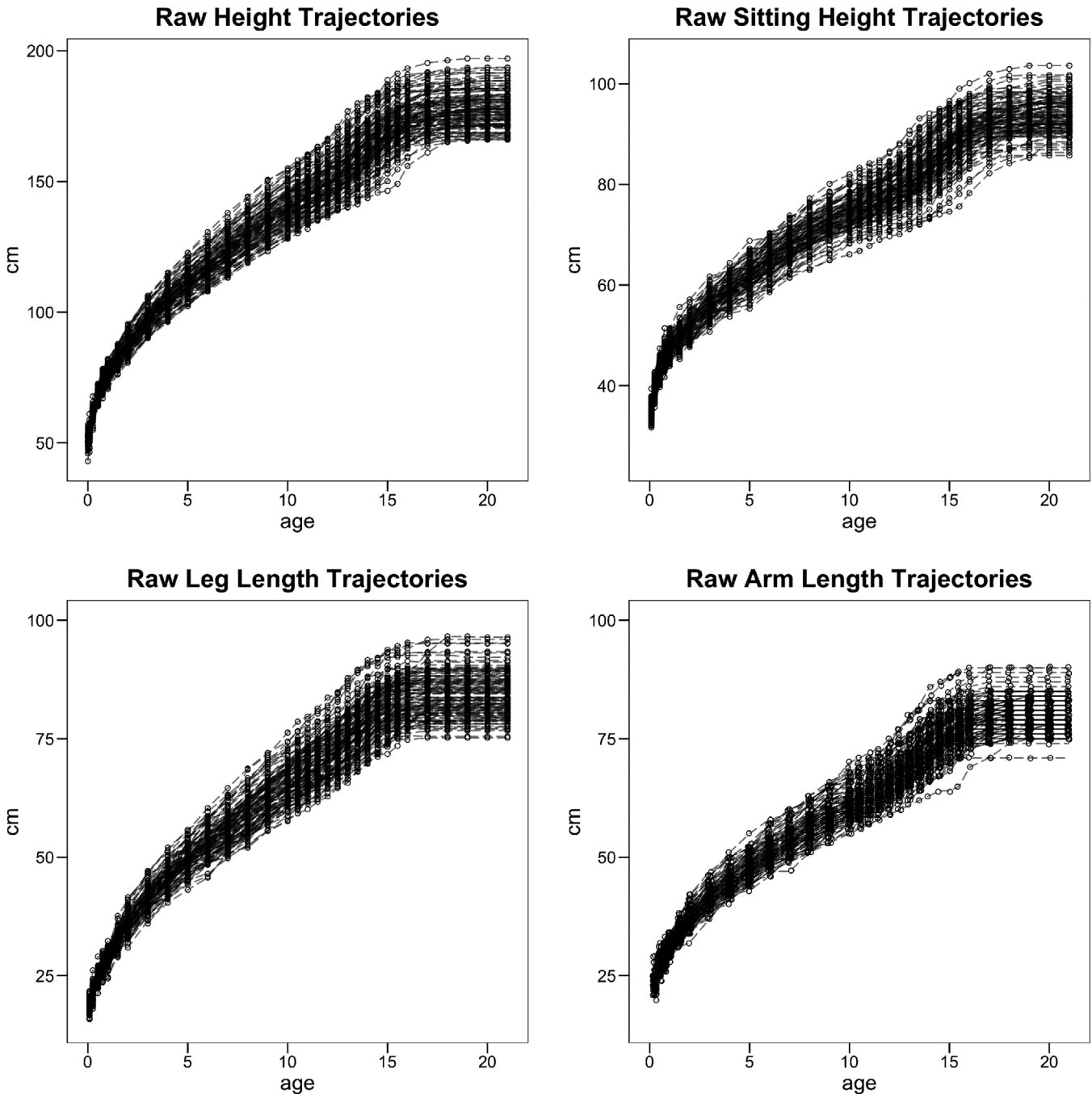
Raw Height Trajectories

Raw Sitting Height Trajectories

Raw Leg Length Trajectories

Raw Arm Length Trajectories

FIGURE 3 Raw growth trajectories for all Zürich boys

measurements stay constant in adulthood, as almost all subjects reach full maturation 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 3. 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, that is, the derivatives of the growth trajectories, via local weighted least squares using the package `fdapace` (Carroll *et al.*, 2020).
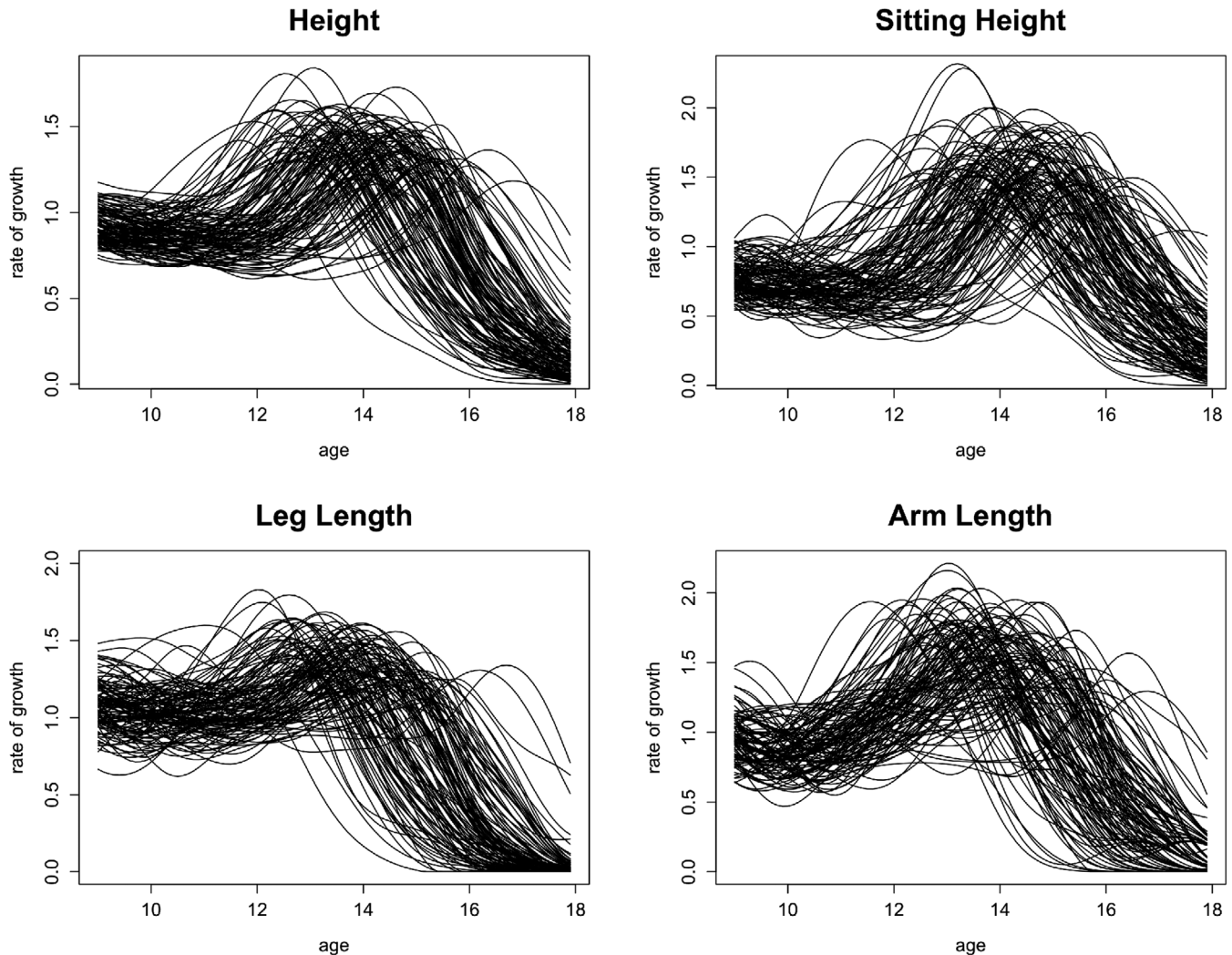
FIGURE 4 Scaled growth velocity curves for Zürich boys

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, for example, Sheehy *et al.* (1999)) and allow for the comparison of modalities that are on dissimilar scales. Figure 4 shows the rescaled derivative estimates for the four growth processes that we consider. After this preprocessing, we now have multivariate functional data with component functions such as those shown for the individuals in Figure 1. 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).

One can interpret these shift parameters in a pairwise manner. For example, legs tend to undergo their growth spurts roughly half a year before arms do

TABLE 1 Estimated global XC shifts for Zürich boys

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

Note. These estimates imply the following ordering of growth spurts: (1) leg, (2) height, (3) arm, (4) sitting height.

($\hat{\tau}_{34} = \hat{\theta}_3 - \hat{\theta}_4 \approx 0.5$) and sitting height trails roughly half a year behind standing height ($\hat{\tau}_{21} = \hat{\theta}_2 - \hat{\theta}_1 \approx -0.5$). 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).

We next investigate some individuals before and after component alignment for a demonstration of how XC alignment affects the curves. Figure 5 (top) shows three
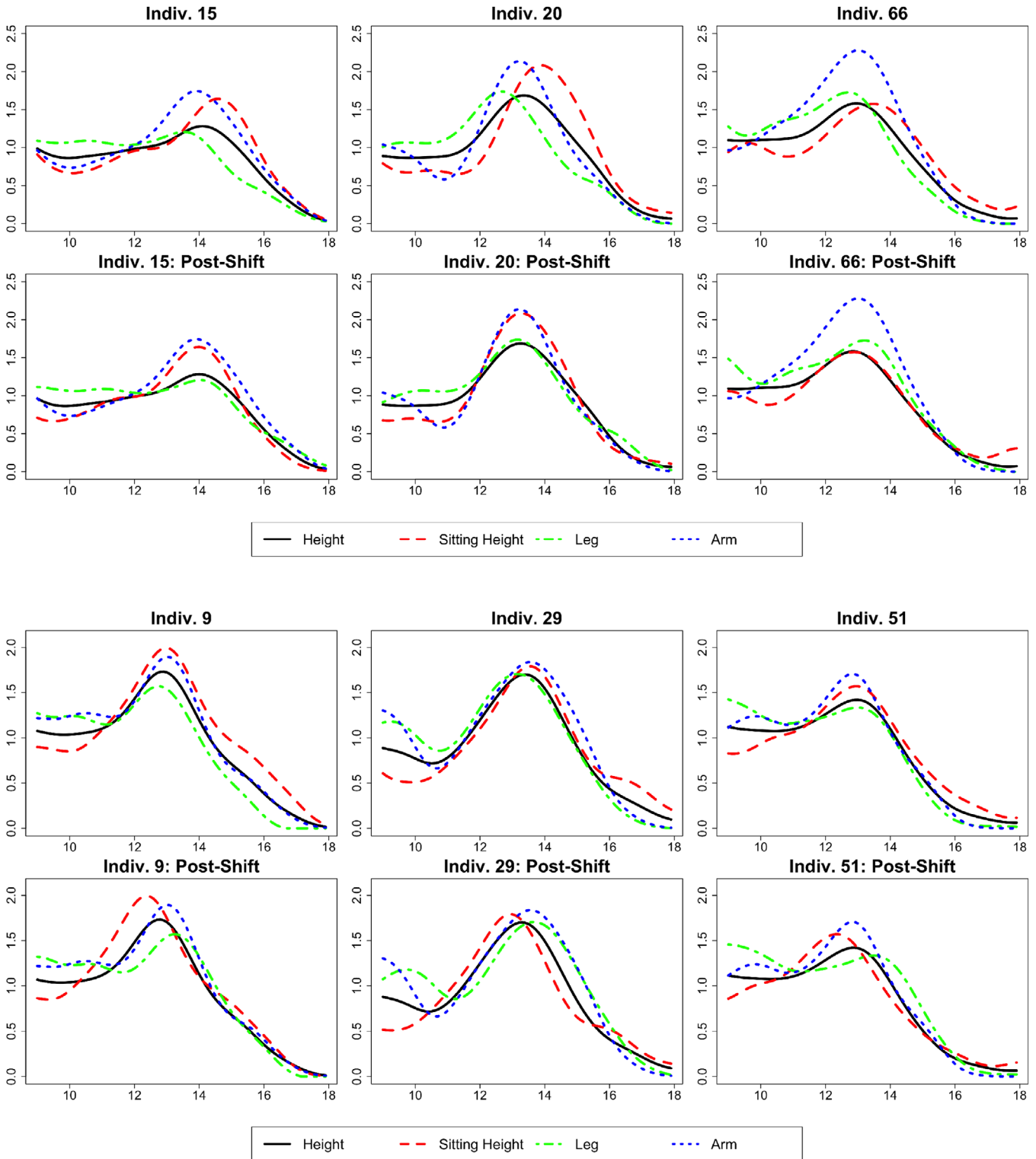
FIGURE 5 Well-aligned (top) and poorly aligned individuals (bottom) after component alignment. Growth modalities are standing height (black, solid), sitting height (red, dashed), leg length (green, dot-dashed), and arm length (blue, dotted)

individuals who are representative of the "average" ordering of growth spurts across modalities, whereas Figure 5 (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 5 (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.
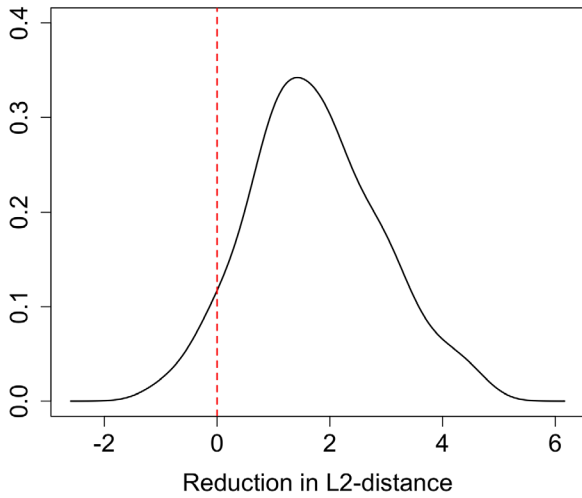
FIGURE 6 Kernel density estimate of the decrease in total $\mathcal{L}^2$-distance after performing XCR. The dashed line indicates no change

To illustrate this further, we use the total cross-component $\mathcal{L}^2$-distance (XD) for an individual as a function of θ,

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

noting that under a perfect model fit we would have $X_{ij}(t+\theta_j) = Z_i(t)$ for all $j = 1, \ldots, p$, and $\mathrm{XD}_i(\theta) = \sum_{j<k} \int_{\mathcal{I}} \{Z_i(t) - Z_i(t)\}^2 dt = 0$. Figure 6 displays the distribution of the difference in total cross-component $\mathcal{L}^2$-distance before and after shifting, for example, $\mathrm{XD}_i(\mathbf{0}) - \mathrm{XD}_i(\hat{\theta})$. Here it is noteworthy that implementing component alignment reduced total $\mathcal{L}^2$-distance in the sample by about 40%.

6 | SIMULATION STUDY

We demonstrate here the superior fit of curves aligned by XCR prior to analysis through FPCA. We use 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 and set $\theta = (-5, -2.5, 2.5, 5)$ and $\mathcal{I} = [10, 40]$. We contaminate the curves with functional noise, measurement error, and noisy shift parameters by generating contaminated component curves

$$X_{ij}(t_k) = X_j(t_k - \theta_j + \eta_{ij}) + \zeta_{ij} \sin\left(\frac{\pi t_k}{5}\right) + e_{ijk}, \quad (12)$$

where $\eta_{ij} \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma_\eta^2)$, $\zeta_{ij} \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma_\zeta^2)$, $e_{ijk} \overset{\text{iid}}{\sim} \mathcal{N}(0, 1)$, and k indexes the points on the data grid spanning $\mathcal{T}$ by increments of 0.5. Here the noise on the time domain

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

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

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 the component processes for all subjects as a noisy sample of Z and applying 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, as the cross-component warping masks the features of Z, and this is indeed what the following simulations show. A sample of $N = 100$ curves were fit via FPCA using the first two eigenfunctions, both with and without incorporating XCR. When incorporating XCR, curves were first generated and used to estimate XC shifts, whereupon components were shifted according to these estimates, followed by an 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 to their original domains through the estimated shifts. To quantify the advantage of incorporating XCR, we obtained the integrated mean squared error for both approaches. The benefit of including XCR for various noise scenarios was measured through the percentage 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 2 shows the average percentage 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 (ie, $\sigma_\eta^2 > 0.5$), the advantage of XCR starts to decrease. It conforms with expectations 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
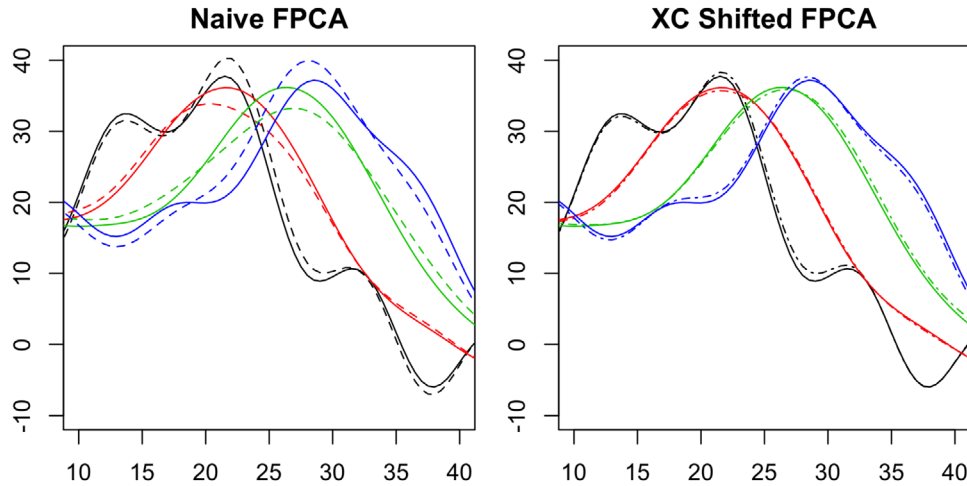
FIGURE 7 Differences in fit when using just naive FPCA (dashed, left) and FPCA and XCR together (dot-dashed, right), when $\sigma_\eta^2 = 0.25$, $\sigma_\zeta^2 = 25$. Solid lines represent the original data

inspection would not likely indicate the presence of any systematic cross-component warping.

A visual comparison of performance for the two approaches can be seen for an example set of curves in Figure 7. Unmodified FPCA is ill-suited to account for sources of horizontal variation, like shift warping, as its eigenfunctions and their scores are geared toward 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 7. By accounting for component warping with XCR, however, the burden of modeling time domain variation is lifted from FPCA, which can then focus on amplitude variability without the confounding phase noise. Another simulation study characterizing finite sample performance at more noise levels can be found in the Supplement.

7 | THEORETICAL RESULTS

For bivariate XCR, a key finding is that the centered process

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

converges weakly to a Gaussian limit process $Z(\tau)$, where L_n, Λ are as in (4), (5). The details of this result are shown in Lemma 1 of the Supplement. To show weak convergence of the pairwise estimate $\hat{\tau}$ as defined in (6), we require the following assumptions on Λ where d denotes $\mathcal{L}^1$-distance.

(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

$$\Lambda(\tau) - \Lambda(\tau_0) \geq Cd(\tau, \tau_0)^\beta.$$

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, for example, Petersen and Müller (2016). We also require the following assumptions for the observed random processes.

(A1) $X_j(t)$ is continuously twice differentiable for $j = 1, \dots, p$,
(A2) $E\{\int_I X_j^4(t)dt\} < \infty$, for $j = 1, \dots, p$,
(A3) $E\{\int_I X_j'^4(t)dt\} < \infty$, for $j = 1, \dots, p$.

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.

Theorem 1. *In the bivariate case, under assumptions (P1)-(P2), and (A1)-(A3), we have*

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

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

The proof is in the Supplement 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, that is, when $\beta = 2$, one obtains the parametric rate $n^{1/2}$.

Our main result for general XCR concerns the rate of convergence of the estimated global shift vector and its asymptotic distribution, as follows.

Theorem 2. *In the general case, 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

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

where $m_{\tau_0} = \{L_n(\tau_{12}), L_n(\tau_{13}), \dots, L_n(\tau_{(p-1)p})\}^T$ and V_{τ_0} is the Hessian of $\Lambda(\tau) = E(m_\tau)$ evaluated 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 scores can be used for subsequent downstream analysis. The identification and estimation of the underlying latent process may also lead to a more parsimonious representation and is of interest in itself.

There are limitations of the framework we have established here. Although the shift-warping model we develop in this paper is appropriate for certain applications such as the Zürich Longitudinal Growth Study, the cross-component warping phenomena need not be restricted to shifts in general and may emerge in the form of nonlinear 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 the warp has a simple structure, 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 complex 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 $\mathcal{L}^2$-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 that we have presented here. Although 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 XCR is left for future research. Other potential directions of interest concern alternative representations of the cross-component warping problem.

DATA AVAILABILITY STATEMENT
The data used in the application section are proprietary to the Zürich University Children's Hospital and therefore not shared.

ORCID
Cody Carroll https://orcid.org/0000-0003-3525-8653

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. *Electronic Journal of Statistics*, 38, 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*, 38, 211–220.

Carroll, C., Gajardo, A., Chen, Y., Dai, X., Fan, J., Hadjipantelis, P.Z., Han, K., Ji, H., Lin, S.-C., Dubey, P., Mueller, H.-G. and Wang, J.L. (2020) *fdapace: functional data analysis and empirical dynamics.* R package version 0.5.4.

Chiou, J.-M. (2012) Dynamical functional prediction and classification, with application to traffic flow prediction. *Annals of Applied Statistics*, 38, 1588–1614.

Chiou, J.-M., Chen, Y.-T. and Yang, Y.-F. (2014) Multivariate functional principal component analysis: a normalization approach. *Statistica Sinica*, 38, 1571–1596.

Chiou, J.-M., Yang, Y.-F. and Chen, Y.-T. (2016) Multivariate functional linear regression and prediction. *Journal of Multivariate Analysis*, 38, 301–312.

Claeskens, G., Hubert, M., Slaets, L. and Vakili, K. (2014) Multivariate functional halfspace depth. *Journal of the American Statistical Association*, 38, 411–423.

Dubin, J.A. and Müller, H.-G. (2005) Dynamical correlation for multivariate longitudinal data. *Journal of the American Statistical Association*, 38, 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*, 38, 93–100.

Gasser, T. and Kneip, A. (1995) Searching for structure in curve samples. *Journal of the American Statistical Association*, 38, 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*, 38, 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*, 38, 459–474.

Gasser, T., Köhler, W., Müller, H.-G., Kneip, A., Largo, R., Molinari, L. and Prader, A. (1984) Velocity and acceleration of height growth using kernel estimation. *Annals of Human Biology*, 38, 397–411.

Gasser, T., Müller, H.-G., Köhler, W., Molinari, L. and Prader, A. (1984) Nonparametric regression analysis of growth curves. *Annals of Statistics*, 38, 210–229.

Gervini, D. and Gasser, T. (2004) Self-modeling warping functions. *Journal of the Royal Statistical Society: Series B*, 38, 959–971.

Hall, P. and Horowitz, J.L. (2007) Methodology and convergence rates for functional linear regression. *Annals of Statistics*, 38, 70–91.

Han, K., Hadjipantelis, P.Z., Wang, J.-L., Kramer, M.S., Yang, S., Martin, R.M. and Müller, H.-G. (2018) Functional principal component analysis for identifying multivariate patterns and archetypes of growth, and their association with long-term cognitive development. *PLoS ONE*, 38, 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*, 38, 649–659.

Härdle, W. and Marron, J.S. (1990) Semiparametric comparison of regression curves. *Annals of Statistics*, 38, 63–89.

Jacques, J. and Preda, C. (2014) Model-based clustering for multivariate functional data. *Computational Statistics and Data Analysis*, 38, 92–106.

Jain, N.C. and Marcus, M.B. (1975) Central limit theorems for C(S)-valued random variables. *Journal of Functional Analysis*, 38, 216–231.

Kneip, A. and Gasser, T. (1992) Statistical tools to analyze data representing a sample of curves. *Annals of Statistics*, 38, 1266–1305.

Kneip, A. and Gasser, T. (1995) Model estimation in nonlinear regression under shape invariance. *Annals of Statistics*, 38, 551–570.

Marron, J.S., Ramsay, J.O., Sangalli, L.M. and Srivastava, A. (2015) Functional data analysis of amplitude and phase variation. *Statistical Science*, 38, 468–484.

Milani, S. (2000) Kinetic models for normal and impaired growth. *Annals of Human Biology*, 38, 1–18.

Park, J. and Ahn, J. (2017) Clustering multivariate functional data with phase variation. *Biometrics*, 38, 324–333.

Petersen, A. and Müller, H.-G. (2016) Fréchet integration and adaptive metric selection for interpretable covariances of multivariate functional data. *Biometrika*, 38, 103–120.

Preece, M. and Baines, M. (1978) A new family of mathematical models describing the human growth curve. *Annals of Human Biology*, 38, 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*, 38, 1811–1816.

Ramsay, J.O. and Silverman, B.W. (2005) *Functional Data Analysis*, 2nd edition. New York, NY: Springer.

Sangalli, L.M., Secchi, P., Vantini, S. and Vitelli, V. (2010) K-mean alignment for curve clustering. *Computational Statistics & Data Analysis*, 38, 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*, 38, 309–331.

Silverman, B.W. (1995) Incorporating parametric effects into functional principal components analysis. *Journal of the Royal Statistical Society: Series B*, 38, 673–689.

Srivastava, A. and Klassen, E.P. (2016) *Functional and Shape Data Analysis*. Berlin: Springer.

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.

Tang, R. and Müller, H.-G. (2008) Pairwise curve synchronization for functional data. *Biometrika*, 38, 875–889.

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

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

Wang, J.-L., Chiou, J.-M. and Müller, H.-G. (2016) Functional data analysis. *Annual Review of Statistics and Its Application*, 38, 257–295.

Zhou, L., Huang, J. and Carroll, R. (2008) Joint modelling of paired sparse functional data using principal components. *Biometrika*, 38, 601–619.

SUPPORTING INFORMATION

Supporting Web Appendices referenced in Sections 2, 5, and 6 are available with this paper at the Biometrics website on Wiley Online Library. A simulated data set and example code are also provided.

How to cite this article: Carroll C, Müller H-G, Kneip A. Cross-component registration for multivariate functional data, with application to growth curves. *Biometrics*. 2020;1–13. https://doi.org/10.1111/biom.13340