Cross-Component Registration for Multivariate Functional Data with Application to Longitudinal Growth Curves

Cody Carroll¹, Hans-Georg Müller¹, and Alois Kneip²
¹ Department of Statistics, University of California, Davis
² Department of Economics, Universität Bonn

September 2018

ABSTRACT

Multivariate functional data are becoming ubiquitous with the advance of modern technology. Multivariate functional data are substantially more complex than univariate functional data. In particular, we study a novel model for multivariate functional data where the component processes exhibit mutual time warping. That is, the component processes exhibit a similar shape but are subject to time warping across their 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 existing registration methods in several major ways. Namely, instead of focusing on individual-specific registration, we focus on registering across components on a population-wide level. By doing so our proposed estimates for these shifts enjoy parametric rates of convergence and often have intuitive physical interpretations. We exemplify these interpretations by applying our methodology to the Zürich Longitudinal Growth data. We also demonstrate the conditions under which our methodology works via simulation.

KEY WORDS: Functional data analysis, multivariate functional data, component processes, time warping, functional registration, shift-warp models, longitudinal study, growth curves.

¹Research supported by NSF Grant DMS-1712864.
1. INTRODUCTION

Multivariate functional data are often encountered in biological or chemical processes which are continuously measured for a group of subjects or observational units and also in the social sciences. Such processes arise in many longitudinal studies, ranging from traffic monitoring to measurements of protein levels during metabolic processes (Chiou 2012; Rubin 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).

However, directly applying MFPCA to multivariate functional data often yields lack-luster results if the curves are subject to warping distortions which are an important feature of some classical functional data such as multivariate growth curves. In particular, 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 time warping can distort principal components and inflate data variance (Marron et al. 2015). However, if handled properly, these mutual warpings may be used to discover underlying time shifts across the component curves, which in context may have intuitive physical interpretations.

The general idea of time relations between component processes has been considered in time series analysis, for example in the notion of Granger causality (Granger 1969).
However, the situation is quite different for functional data where one has repeated observations of the multivariate process for many subjects and can take full advantage of the entire sample. In addition, one may have measurements per subject on a grid of time points possibly contaminated with measurement errors. This situation, as found in many longitudinal studies with multivariate measurements across the sciences, naturally suggests to develop functional methods which are geared towards repeatedly sampled multivariate functional data. The analysis of the Zürich Longitudinal Growth study data motivates to model multivariate functional data by allowing the components to be mutually time-shifted against each other. This then leads to interesting biological interpretations.

In many cases, the component processes of multivariate functional data exhibit a 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. 1984b). 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 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). Surprisingly, models for multivariate functional data that include component time warping and specifically
component time shifts have not yet been studied.

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 our emphasis is not clustering but 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. We develop a simple approach where we assume that there are relative time shifts between the component processes that establish time relations between the components. Combining information about the relative shifts between pairs of components then informs the full system of relative timings across body parts. For this task, we develop a new functional registration approach.

Existing methods for aligning curves in a traditional warping context provide subject-specific warping functions and face several issues, including the unidentifiability of the warping functions, unintuitive or complicated interpretations of such functions, and a lack of satisfactory theoretical results. Conceptual problems of standard alignment, for example, are discussed in (Kneip and Ramsay 2008). In the non-traditional warping framework that we consider here, the identifiability problem can be handled with a simple and straightforward identifiability condition. It is especially noteworthy that we obtain a $\sqrt{n}$-rate of the estimated time shifts to their targets under mild regularity conditions. It seems that such fast convergence has not been observed before in warping models, due to the fact that previous warping approaches were focusing on subject-specific rather than component-specific warping, where one has $n$ subjects but only $p$ components, irrespective of the sample size $n$.

The organization of this paper is as follows. In Section 2 we introduce the proposed
cross-component model and present our theoretical results. An application to human growth curves and simulations are discussed in Sections 3 and 4, respectively.

2. BIVARIATE CROSS-COMPONENT REGISTRATION

2.1 Framework and Estimation

We introduce here the idea of registering different component times across modalities, which we call Cross-Component Registration (XCR). Note that XCR is very different from traditional curve warping, also known as curve registration or alignment (Ramsay and Silverman 2005; Wang et al. 2016; Kneip and Gasser 1992; Gasser and Kneip 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 a comprehensive overview of traditional warping methods; this is a varied field where one encounters complications rather quickly (Gasser et al. 1990; Gervini and Gasser 2004; Liu and Yang 2009; Tang and Müller 2008; Marron et al. 2015). The major difference is that instead of estimating $n$ subject-specific warping functions, which align univariate curves across individuals and is the goal of traditional curve warping methods, our novel approach targets a $p$-vector of shift parameters for the case of $p$-dimensional functional data. These component-wise shifts are then applied in the same way across all subjects to mutually align the component curves.

To illustrate, we write $(X_1(t),\ldots,X_p(t))^T$ to represent the generic underlying multivariate process and for sampled realizations of the process we write $(X_{i1}(t),\ldots,X_{ip}(t))^T$ for $i=1,\ldots,n$ subjects. In this subsection we first consider the case of multivariate functional data with $p=2$ component curves to introduce the main ideas, and will then discuss the extension to $p>2$. To fix the ideas, 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 \((X_1(t), X_2(t))\). An intuitive measure of alignedness for curves with similar shape is their \(L^2\)-distance. Naturally, if two curves are aligned, their \(L^2\)-distance is minimized. We then aim at minimizing the \(L^2\)-distance on a domain \(T\) (in our example, \(T = [0, 10]\)) as our criterion for alignment, and the shift we seek is such that it minimizes this loss function.

Specifically, we aim for the optimal value of the parameter \(\tau\), the cross-component (XC) shift, which seeks to minimize the criterion

\[
\Lambda(\tau) = E \int_T (X_1(t) - X_2(t - \tau))^2 dt, \tag{1}
\]

with associated sample version

\[
L_n(\tau) = \frac{1}{n} \sum_{i=1}^{n} \int_T (X_{1i}(t) - X_{2i}(t - \tau))^2 dt. \tag{2}
\]

The sample-based shift parameter estimate is then naturally

\[
\hat{\tau} = \arg\min_{\tau} L_n(\tau), \tag{3}
\]

targeting \(\tau_0 = \arg\min_{\tau} \Lambda(\tau)\).

2.2 Theory for Bivariate Cross-Component Registration

A key fact 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 \(Z_n(\tau) \rightsquigarrow Z(\tau)\). To obtain this weak convergence, we require the following assumptions on \(\Lambda\):
(P1) For any $\varepsilon > 0$, $\inf_{\tau \in \mathbb{R}} \Lambda(\tau) < \Lambda(\tau_0)$.

(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) arises from empirical process theory and controls the behavior of $L_n - \Lambda$ near the minimum in order to obtain rates of convergence. We also make the additional assumptions on the observed random processes:

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

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

(A3) $E \left[ \int_T X_j^4(t) dt \right] < \infty$, for $j = 1, \ldots, p$,

These assumptions are standard in the literature, for example as in [Hall and Horowitz (2007)], and allow us to obtain asymptotic covariance matrices for our estimates and ensure finiteness of specific bounding constants which arise in the technical proof.

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

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

where $Z(\tau)$ is a Gaussian process with mean zero and covariance $G(\tau_1, \tau_2) = \int_T \int_T E((X_1(t) - X_2(t - \tau_1))^2(X_1(s) - X_2(s - \tau_2))^2) dt ds - \Lambda(\tau_1)\Lambda(\tau_2)$.

**Theorem 1.** Under the assumptions of Lemma 1, 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 $V = 4\int_T E[(X_1(t) - X_2(t - \tau_0))X_2'(t - \tau_0)]^2 dt / (\Lambda''(\tau_0))^2$ where $\Lambda(\tau) = E\int_T (X_1(t) - X_2(t - \tau))^2 dt$ as before.

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

### 3. GENERAL CROSS-COMPONENT REGISTRATION

We now extend the methodology of bivariate XCR to the case of $p$-variate multivariate functional processes where one aims at aligning 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, $\theta = (\theta_1, \ldots, \theta_p)$, such that when each modality $X_j(t), j = 1, \ldots, p$, is shifted by $\theta_j$, all $p$ curves are aligned. To do this, 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 1 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. 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 latent component curve that is visualized with a dashed line in Figure 1. The observed components are then time-shifted with respect to the latent component and the shifts will be subjected to the constraint that $\sum_{j=1}^p \theta_j = 0$, so that there is no net XC shift from the latent component curve. This assumption is necessary for the identifiability.
A key observation is that the bivariate XC shifts between pairs of component functions are linear combinations of the components of the global XC shifts between each component function and the latent process. Furthermore, the linear map $L$ that maps the vector of XC shifts of the component functions to the latent process is invertible. Thus, by assessing the bivariate XC shifts $\tau_{jk}$ between the component functions, we can infer the global XC vector $\theta$, 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:

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

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

$$\tau^* = L(\theta) = A\theta, \quad (4)$$

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 the latent process and $A$ is the matrix of the linear map, corresponding to the

Figure 1: Observed Curves (left) and Latent Curve (right)
contrasts in [I]. Note that $A$ is of dimension $(p(p - 1)/2 + 1) \times p$, and is always of full rank. Specifically, we find that

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

To implement this approach for functional data analysis, we must first estimate the stacked vector of bivariate XC shifts,

$$\hat{\tau}^* = (\hat{\tau}^T, 0)^T = (\hat{\tau}_{12}, \hat{\tau}_{13}, \ldots, \hat{\tau}_{(p-1)p}, 0)^T. \quad (5)$$

Accordingly, we have the model

$$\hat{\tau}^* = A\theta + \varepsilon$$

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

$$\hat{\theta} = (A' A)^{-1} A' \hat{\tau}^*. \quad (6)$$

by ordinary least squares. The $p$ component curves will be aligned (“sitting” on the latent curve) once they are time-shifted with their respective estimated global XC shifts, $\hat{\theta}$. 
Theorem 2. Under assumptions (P0)-(P2) and (A0)-(A2)

\[ \hat{\theta} - \theta_0 = O_p(n^{-1/2(\beta-1)}). \]

In particular, when \( \beta = 2 \), the sequence \( \sqrt{n}(\hat{\theta} - \theta_0) \) 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 \( m_{\tau_0} = [L_n(\tau_{12}), L_n(\tau_{13}), ..., L_n(\tau_{(p-1)p})]^T \) and \( V_{\tau_0} \) is the Hessian of \( \Lambda(\tau) = E(m_\tau) \) at \( \tau_0 \).

4. APPLICATION TO 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 include standing height, sitting height, arm length, and leg length, and these can be naturally viewed as multivariate functional data (Gasser et al. 1984a, 1989).

The raw trajectories of the \( p = 4 \) component processes for the 232 children measured are displayed in Figure 2, which also indicates the measurement grid. Growth spurts occur at different times for individuals, as evidenced by the crossing of trajectories around ages 10 through 15.

We are especially interested in the timing of pubertal growth spurts, which occur between ages 9 and 18 typically. A common way to study growth velocities is to examine the derivatives of the growth curves instead of the curves themselves (Gasser et al. 1984b). 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 (and common knowledge) indicates that there is a difference in the ways that boys and girls undergo puberty. Accordingly, for the subsequent analysis we separate boys and girls and for brevity display only the results for boys. We estimate the growth velocities,
Figure 2: Raw Trajectories for all 232 children

i.e., the derivatives of the growth trajectories, via local linear smoothing.

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. Figure 3 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 4.
Figure 3: Amplitude Scaled Growth Velocities for Boys

Figure 4: Growth component functions from the Zürich data for an individual
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:

| Component | Modality          | Estimate |
|-----------|-------------------|----------|
| $\theta_1$ | Height            | -0.250   |
| $\theta_2$ | Leg Length        | 0.498    |
| $\theta_3$ | Sitting Height    | -0.523   |
| $\theta_4$ | Arm Length        | 0.053    |

Table 1: Estimated global XC shifts for Zürich boys

We can interpret these shift parameters in a pairwise manner. For example, legs tend to undergo their growth spurts roughly .44 years before arms do ($\hat{\tau}_{24} = .44 = \hat{\theta}_2 - \hat{\theta}_4$) and sitting height trails roughly .27 years behind standing height ($\hat{\tau}_{31} = -.27 = \hat{\theta}_3 - \hat{\theta}_1$).

We next investigate some individuals before and after component alignment for a demonstration of how implementing the alignment affects the curves. Figure 5 (top) shows three 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, as it was common for most individuals to have reduced $L^2$-distance between the component curves after alignment. To illustrate this, we use the total cross-component $L^2$-distance for an individual as a function of $\theta$,

$$XD_i(\theta) = \sum_{j<k} \int_T (X_{ij}(t - \theta_j) - X_{ik}(t - \theta_k))^2 dt.$$  

Figure 6 displays the distribution of the difference in total cross-component $L^2$-distance before and after shifting, i.e., $XD_i(0) - XD_i(\hat{\theta})$. Implementing component alignment reduced total $L^2$-distance by about 40%.
Figure 5: Well-Aligned (top) vs. Poorly Aligned Individuals (bottom) after component alignment.
Figure 6: Kernel density estimate of the distribution of the reduction in total $\mathcal{L}^2$-distance, with 0 marked with a red dashed line.

5. SIMULATION STUDY

We illustrate the proposed estimation methods in several simulation settings for varying levels of noise, using the base curve $Z(t) = 20 - .5t - 30e^{-(t-25)^2/2}$ on $t \in \mathcal{I} = [0, 50]$ as the underlying process dictating the common shape of the component curves. We then generate a 4-dimensional process with components of the form $X_j(t) = Z(t - \theta_j)$ for $j = 1, \ldots, 4$ where $\theta = (-3, -1, 1, 3)$ on a data grid spanning $\mathcal{I}$ by increments of .5, which is displayed in Figure 7.

Figure 7: Base 4-dimensional process
Once we have this base 4–dimensional process, we generate $n = 100$ noise contaminated versions of it as $X_{ij}(t_k) = X_j(t_k) + e_{ijk}$, where $e_{ijk} \overset{iid}{\sim} N(0, \sigma^2)$ and $k$ indexes the points on the data grid described above. Finally, we smooth these observations using local linear smoothing with a final output grid spanning $\mathcal{I}$ by steps of size .01. A random sample of 4 such noisy processes can be seen in Figure 8, in which the subinterval $\mathcal{T} = [10, 40]$ is marked by dashed vertical lines and the dashed horizontal line marks 0 on the y-axis.

![Figure 8: Four randomly selected sampled multivariate simulated processes with noise level $\sigma^2=100$](image)

We evaluate $L^2$-distances on the subinterval $\mathcal{T} = [10, 40]$ and use XCR to estimate $\theta$. 
We implement this at various levels of $\sigma$ to illustrate the bias and variance decomposition and to explore the noise level at which the XCR method breaks down. Simulation results are displayed in Table 2.

Table 2: Bias and variance decomposition of XC estimates at different noise levels

| Noise Level | comp | bias$^2$ | Var | MSE |
|-------------|------|----------|-----|-----|
| $\sigma^2 = 100$ | $\theta_1$ | $2.63 \times 10^{-5}$ | 0.0028 | 0.003 |
| | $\theta_2$ | $3.03 \times 10^{-6}$ | 0.003 | 0.003 |
| | $\theta_3$ | $1.25 \times 10^{-5}$ | 0.003 | 0.003 |
| | $\theta_4$ | $2.25 \times 10^{-8}$ | 0.003 | 0.003 |
| $\sigma^2 = 500$ | $\theta_1$ | $7.83 \times 10^{-6}$ | 0.018 | 0.018 |
| | $\theta_2$ | $5.63 \times 10^{-7}$ | 0.017 | 0.017 |
| | $\theta_3$ | $4.66 \times 10^{-7}$ | 0.018 | 0.018 |
| | $\theta_4$ | $7.45 \times 10^{-6}$ | 0.016 | 0.016 |
| $\sigma^2 = 1000$ | $\theta_1$ | $3.91 \times 10^{-7}$ | 0.036 | 0.036 |
| | $\theta_2$ | $3.98 \times 10^{-5}$ | 0.035 | 0.035 |
| | $\theta_3$ | $3.27 \times 10^{-5}$ | 0.039 | 0.039 |
| | $\theta_4$ | $1.60 \times 10^{-4}$ | 0.040 | 0.040 |
| $\sigma^2 = 5000$ | $\theta_1$ | $3.74 \times 10^{-3}$ | 0.379 | 0.383 |
| | $\theta_2$ | $8.23 \times 10^{-5}$ | 0.454 | 0.454 |
| | $\theta_3$ | $3.20 \times 10^{-6}$ | 0.426 | 0.426 |
| | $\theta_4$ | $2.54 \times 10^{-3}$ | 0.461 | 0.463 |
| $\sigma^2 = 10000$ | $\theta_1$ | 0.116 | 1.080 | 1.196 |
| | $\theta_2$ | 0.026 | 1.075 | 1.101 |
| | $\theta_3$ | 0.027 | 1.142 | 1.169 |
| | $\theta_4$ | 0.115 | 1.127 | 1.242 |
| $\sigma^2 = 50000$ | $\theta_1$ | 4.443 | 2.927 | 7.370 |
| | $\theta_2$ | 0.514 | 2.913 | 3.427 |
| | $\theta_3$ | 0.627 | 3.070 | 3.697 |
| | $\theta_4$ | 4.132 | 3.269 | 7.401 |

Estimates remain practically unbiased until we increase the noise level to $\sigma^2 = 10,000$. At this level the variance contributes roughly 10 times more to MSE than bias squared. If we increase $\sigma^2$ to 50,000, the bias overtakes variance in the decomposition and the method breaks down. If we consider the “signal” in this simulation as the amplitude of the peak of our original base process then this simulation empirically suggests that XCR remains essentially unbiased in settings with signal-to-noise ratios as low as $30: \sqrt{5000} \approx .42$. This
may seem surprising, but one can understand this result by observing that in order to disrupt the signal beyond recognition, the noise level must be large enough to drastically distort all measurements in the signal. If the noise level is small enough so that only a few measurements are more seriously distorted, the smoothing ameliorates this noise and can recover the base shape well enough to estimate the XC shifts accurately. As a consequence, XCR breaks down only in the presence of large errors.

6. CONCLUDING REMARKS

Cross-component registration seeks to address a problem of mutual time warping that is specific to multivariate functional data and does not manifest itself in univariate functional data. By focusing on the warping across components, and not warping between individual subjects, we are able to estimate population-wide time shift parameters with fast rates of convergence, under suitable assumptions.

This approach leads to insights about the relative timings of the component processes. While the proposed method for component wise alignment is new, as far as we know, applications of XCR will be relevant for many longitudinal studies and multivariate functional data. We have demonstrated that XCR is useful for the analysis of growth data. We expect it to be applicable for many other kinds of data that can be viewed as samples of multivariate processes. For example, one could employ XCR to estimate the relative timings of protein activations during biological processes. Potential applications extend to the relative timings of economics indices, internet trends, and composite material breakdown in engineering, to name a few.
REFERENCES

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.

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,” *J. Multivar. Anal.*, 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.

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., and Prader, A.
(1984a), “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. (1984b), “Nonparametric Regression Analysis of Growth Curves,” *Annals of Statistics*, 12, 210–229.

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

Granger, C. W. (1969), “Investigating causal relations by econometric models and cross-spectral methods,” *Econometrica: Journal of the Econometric Society*, 424–438.

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

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.

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 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.

Park, J. and Ahn, J. (2017), “Clustering multivariate functional data with phase vari-
tion,” *Biometrics*, 73, 324–333.

Ramsay, J. O. and Silverman, B. W. (2005), *Functional Data Analysis*, Springer Series in Statistics, New York: Springer, 2nd ed.

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.

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.
APPENDIX: PROOFS

Proof of Lemma 1.

In the following we consider compact intervals $\mathcal{T}$ and $\mathcal{I}$ such that $\mathcal{T} \subset \mathcal{I}$ where the functions are observed on $\mathcal{I}$ and $\mathcal{T}$ is the subinterval on which we take $L^2$-distance of shifted curves. We also use $C$ to represent a generic constant. We first establish a Central Limit Theorem for $Z_n(\tau) = \sqrt{n}(L_n(\tau) - \Lambda(\tau))$ by applying Jain and Marcus (1975), so the proof reduces to verify the following conditions:

(i.) $E(f(Z_n(\tau))) = 0$ for all continuous linear functionals $f$.

(ii.) $\sup_{\tau} E(Z_n(\tau)) < \infty$

(iii.) There exists a non-negative random variable $M$ with finite variance such that

$$|Z_n(\tau_1, \omega) - Z_n(\tau_2, \omega)| \leq M|\tau_1 - \tau_2|$$

(iv.) On the compact interval $\mathcal{T}$, $\int_{\mathcal{T}} \sqrt{\log N(\varepsilon, \mathcal{T}, d)d\varepsilon} < \infty$, where $N(\varepsilon, \mathcal{T})$ is the covering number of $\mathcal{T}$ with balls of radius $\varepsilon$ under the norm $d(x, y) = |x - y|$.

Verifying these conditions,

(i.) follows as $E(f(Z_n(\tau))) = f(E(Z_n(\tau))) = f(0) = 0$.

(ii.) Since $Z_n(\tau)$ is centered,

$$E(Z_n^2(\tau)) = \text{Var}(\int_{\mathcal{T}} (X_{i_1}(t) - X_{i_2}(t - \tau))^2 dt)$$

$$\leq E \left[ \left( \int_{\mathcal{T}} (X_{i_1}(t) - X_{i_2}(t - \tau))^2 dt \right)^2 \right]. \quad (A.1)$$
We examine the argument inside the square in more detail:

\[
\int_\tau (X_{i1}(t) - X_{i2}(t - \tau))^2 dt \leq \left( \left( \int_\tau X_{i1}^2(t) dt \right)^{1/2} + \left( \int_\tau X_{i2}^2(t - \tau) dt \right)^{1/2} \right)^2 = (U + V)^2,
\]

\[
E(Z_n^2(\tau)) = E \left[ (U + V)^4 \right] \leq C \left( E(U^4) + E(V^4) \right).
\]

This is finite since

\[
E(U^4) \leq \int_\tau E(X_{i1}^4(t)) dt \leq E \left[ \int_\tau (X_{i1}^2(t)) dt \right]^2 < \infty,
\]

The same argument applies for \(E(V^4)\). Then,

\[
\sup_\tau E(Z_n^2(\tau)) \leq \sup_\tau C \left( \int_\tau E \sup_t (X_{i1}^4(t)) dt + \int_\tau E \sup_t (X_{i2}^4(t - \tau)) dt \right) < \infty,
\]

where the right hand terms are finite under (A.1).

(iii.) Consider differences of the \(Z_n\) with different arguments \(\tau\). Defining

\[
D_i(\tau) := \int_\tau (X_{i1}(t) - X_{i2}(t - \tau))^2 dt - E \left[ (X_{i1}(t) - X_{i2}(t - \tau))^2 \right] dt
\]

and applying the mean value theorem,

\[
|Z_n(\tau_1) - Z_n(\tau_2)| = \\
\sqrt{n} \left| \frac{1}{n} \sum_{i=1}^n \int_\tau (X_{i1}(t) - X_{i2}(t - \tau_1))^2 dt - E \int_\tau (X_{i1}(t) - X_{i2}(t - \tau_1))^2 dt \right. \\
- \left. \left( \frac{1}{n} \sum_{i=1}^n \int_\tau (X_{i1}(t) - X_{i2}(t - \tau_2))^2 dt - E \int_\tau (X_{i1}(t) - X_{i2}(t - \tau_2))^2 dt \right) \right|,
\]
\[ |Z_n(\tau_1) - Z_n(\tau_2)| \leq \frac{1}{\sqrt{n}} \sum_{i=1}^{n} |D'_i(\xi)| \cdot |\tau_1 - \tau_2| = |\tau_1 - \tau_2| T_n \]

with \( T_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} |D'_i(\xi)| \).

Now, due to the continuity of \( D \) and its derivatives in \( \tau \) and \( t \),

\[
\frac{\partial}{\partial \xi_i} D_i(\xi_i) = \int_{\mathcal{T}} \frac{\partial}{\partial \xi_i} [(X_{i1}(t) - X_{i2}(t - \xi_i))^2 - E [(X_{i1}(t) - X_{i2}(t - \xi_i))^2]] \, dt
\]

\[
= 2 \int_{\mathcal{T}} \left[ E [(X_{i1}(t) - X_{i2}(t - \xi_i)) X'_{i2}(t - \xi_i)] - (X_{i1}(t) - X_{i2}(t - \xi_i)) X'_{i2}(t - \xi_i) \right]
\]

\[
= 2 \int_{\mathcal{T}} E[B_i(\xi_i, t)] - B_i(\xi_i, t) dt
\]

where \( B_i(\xi_i, t) = (X_{i1}(t) - X_{i2}(t - \xi_i)) X'_{i2}(t - \xi_i) \).

Then,

\[
\frac{1}{4} E((D'_i(\xi_i))^2) = E \left[ \left( \int_{\mathcal{T}} E[B_i(\xi_i, t)] - B_i(\xi_i, t) dt \right)^2 \right]
\]

\[
= \int_{\mathcal{T}} \int_{\mathcal{T}} \text{Cov}(B_i(\xi_i, u), B_i(\xi_i, v)) dudv \leq \int_{\mathcal{T}} \text{Var}(B_i(\xi_i, t)) dt
\]

Next, we bound this variance,

\[
\text{Var}(B_i(\xi_i, t)) \leq E[B_i^2(\xi_i, t)] = E \left[ (X_{i1}(t) - X_{i2}(t - \xi_i))^2 (X'_{i2}(t - \xi_i))^2 \right]
\]

\[
\leq \left( C \left( E(X_{i1}^4(t)) + E(X_{i2}^4(t - \xi_i)) \right)^{1/2} E \left[ (X'_{i2}(t - \xi_i))^4 \right] \right)^{1/2}
\]
so that

\[ E(T_n^2) = \leq 4 \sup_t \int_T \text{Var}(B_i(\xi_i, t))dt \]

\[ \leq 4 \left( C \left( E \left( \int_T X_{11}^4(t) dt \right) + E \left( \int_T X_{12}^4(t - \xi_i) dt \right) \right) E \left[ \int_T \left( X_{12}'(t - \xi_i) \right)^4 dt \right] \right)^{1/2} \leq \infty \]

since each of the terms in the last line are finite by the integral conditions in (A2) and (A3).

(iv.) Is trivially satisfied.

With all 4 criterions checked, we apply the CLT of (Jain and Marcus 1975) to \( Z_n(\tau) \) and have the result.

Proof of Theorem 1. Our next result relies on Lemma 1 of Jain and Marcus (1975), which implies that

\[ Z_n(\tau) = \sqrt{n}(L_n(\tau) - \Lambda(\tau)) \overset{D}{\to} Z(\tau), \text{ whence} \]

\[ \sup_{\tau} |\sqrt{n}(L_n(\tau) - \Lambda(\tau))| = \sup_{\tau} |Z_n(\tau)| \overset{p}{\to} \sup_{\tau} |Z(\tau)| = O_p(1), \text{ and therefore} \]

\[ \sup_{\tau} |L_n(\tau) - \Lambda(\tau)| = o_p(1) \quad (1.1) \]

From (1.1) and Theorem 3.2.3 in Van der Vaart and Wellner (1996), we then have

\[ \hat{\tau}_n - \tau_0 = o_p(1). \]
For the next part of the proof we consider the process $V_n(\tau) := L_n(\tau) - \Lambda(\tau)$. Then,

$$|V_n(\tau) - V_n(\tau_0)| = |L_n(\tau) - \Lambda(\tau) - (L_n(\tau_0) - \Lambda(\tau_0))|$$

$$= |L_n(\tau) - L_n(\tau_0) - (\Lambda(\tau) - \Lambda(\tau_0))|$$

$$\leq \frac{1}{n} \sum_{i=1}^{n} |W_i(\tau) - EW_i(\tau)|$$

where $W_i(\tau) := \int_{T} (X_{i1}(t) - X_{i2}(t - \tau))^2 - (X_{i1}(t) - X_{i2}(t - \tau_0))^2 dt$. To control this uniformly over small $d(\tau, \tau_0)$, we define the function $g$ as $g(\tau) = V_n(\tau) = L_n(\tau) - \Lambda(\tau)$ and the function class $\mathcal{M}_\delta := \{g(\tau) - g(\tau_0) : d(\tau, \tau_0) < \delta\}$. An envelope function for $\mathcal{M}_\delta$ is $G(\delta) = 2|T|\delta$ and $E(G^2(\delta)) = G^2(\delta) = O(\delta^2)$. Define $B_\delta(\tau)$ to be the ball of radius $\delta$ centered at $\tau$ and define the entropy integral to be $J = \int_{1}^{0} \sqrt{1 + \log N(\delta, B_\delta(\tau), d)} d\delta$ so that $J = O(1)$. Then Theorems 2.7.11 and 2.14.2 of Van der Vaart and Wellner (1996) imply that for small enough $\delta$,

$$E\left(\sup_{\tau : d(\tau, \tau_0) < \delta} \left| \frac{1}{n} \sum_{i=1}^{n} W_i(\tau) - EW_i(\tau) \right| \right) \leq J E\left(G^2(\delta)\right)^{1/2} \sqrt{n} = O(\delta n^{-1/2})$$

which implies

$$E\left(\sup_{\tau : d(\tau, \tau_0) < \delta} |V_n(\tau) - V_n(\tau_0)| \right) = O(\delta n^{-1/2})$$

Finally, let $r_n = n^{\frac{\beta}{(\beta - 1)}}$ and

$$S_{j,n} = \{\tau : 2^{j-1} < r_n d(\tau, \tau_0)^{\beta/2} < 2^j\}.$$

Choose $\eta > 0$ to satisfy (P1) and set $\tilde{\eta} := \eta^{\beta/2}$. For any integer $N$,

$$P(\tau_n d(\tilde{\tau}_n, \tau_0)^{\beta/2} > 2^N) \leq P(\tau_n d(\tilde{\tau}_n, \tau_0) \geq \tilde{\eta}) + \sum_{j>N} P\left(\sup_{\tau \in S_{j,n}} |V_n(\tau) - V_n(\tau_0)| \geq C\frac{2^{2(j-1)}}{r_n^2} \right).$$
where $\hat{\tau}_n$ is defined as in equation (5). For each $j$ in the sum on the right hand side we have $d(\tau, \tau_0) \leq \left( \frac{2}{r_n} \right)^{2/\beta} \leq \eta$, so this sum is bounded by

$$\sum_{j>N} \frac{2^{2j(1-\beta)/\beta} \sqrt{n}}{r_n^{2(1-\beta)/\beta}} \leq \sum_{j>N} \left( \frac{1}{4^{(\beta-1)/\beta}} \right)^j.$$ 

Since $\beta > 1$, the last series converges and therefore the original probability can be made arbitrarily small by choosing a large enough $N$. This proves the desired result that $d(\hat{\tau}_n, \tau_0) = O_p(r_n^{-2/\beta}) = O_p(n^{-1/2(\beta-1)})$. Finally, asymptotic normality follows from Theorem 5.23 of van der Vaart (1998), where the Lipschitz condition and bounded second moment have been shown already in parts (ii.) and (iii.) of the proof of Lemma 1.

**Proof of Theorem 2**: For convergence of $\hat{\theta}$, we recall that $\hat{\theta} = (A^T A)^{-1} A^T \tau^*$,

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

where the first $p - 1$ rows contain all pairwise contrasts of the first column with the rest, the next $p - 2$ rows contain all pairwise contrasts of the second column with the rest, and so on, until the second to last row contains the final pairwise contrast. The last line contains
all 1’s to represent the constraint that \( \sum_{j=1}^{p} \theta_j = 1 \). We observe that \( A^T A = pI_p \), which can be verified with a simple matrix multiplication. It then immediately follows that \( \hat{\theta} = (A^T A)^{-1} A^T \hat{\tau}^* = \frac{1}{p} A^T \hat{\tau}^* = B \hat{\tau}^* \) and since the linear mapping induced by the matrix \( B = \frac{1}{p} A^T \) is continuous, we can apply the continuous mapping theorem and observe that

\[
\hat{\theta} - \theta_0 = o_p(1).
\]

The proof for the rate of convergence follows the same arguments as those for \( \hat{\tau}_n \). The asymptotic normality of \( \hat{\theta} \) when \( \beta = 2 \) requires a closer examination. When estimating \( \hat{\tau}^* \), we are really estimating several pairwise XC shifts at the same time. This is done by minimizing several different \( L^2 \) distances, which we stack in a vector, \( m_{\tau_0}(X_1, \ldots, X_p) \), defined as:

\[
m_{\tau_0} = m_{\tau_0}(X_1, \ldots, X_p) = \begin{bmatrix}
\frac{1}{n} \sum_{i=1}^{n} \int_T (X_{i1}(t) - X_{i2}(t - \tau_{12}))^2 dt \\
\frac{1}{n} \sum_{i=1}^{n} \int_T (X_{i1}(t) - X_{i3}(t - \tau_{13}))^2 dt \\
\vdots \\
\frac{1}{n} \sum_{i=1}^{n} \int_T (X_{i1}(t) - X_{ip}(t - \tau_{1p}))^2 dt \\
\frac{1}{n} \sum_{i=1}^{n} \int_T (X_{i2}(t) - X_{i3}(t - \tau_{23}))^2 dt \\
\vdots \\
\frac{1}{n} \sum_{i=1}^{n} \int_T (X_{i2}(t) - X_{ip}(t - \tau_{2p}))^2 dt \\
\vdots \\
\frac{1}{n} \sum_{i=1}^{n} \int_T (X_{ip}(t) - X_{ip}(t - \tau_{p-1}p))^2 dt 
\end{bmatrix}
\]

By theorem 5.23 of van der Vaart (1998), then, we have

\[
\sqrt{n}(\hat{\tau}_n - \tau_0) \rightsquigarrow \mathcal{N}(0, V^{-1}_0 E[\nabla m_{\tau_0} \nabla m_{\tau_0}'] V^{-1}_0)
\]
where \( \nabla \mu_{\tau_0} \) represents the gradient of \( m_\tau \) evaluated at \( \tau_0 \), i.e.,

\[
\nabla \mu_{\tau_0} = \begin{bmatrix}
\frac{2}{n} \sum_{i=1}^{n} \int_T (X_{i1}(t) - X_{i2}(t - \tau_{12})) X'_{i2}(t - \tau_{12}) dt \\
\frac{2}{n} \sum_{i=1}^{n} \int_T (X_{i1}(t) - X_{i3}(t - \tau_{13})) X'_{i3}(t - \tau_{13}) dt \\
\vdots \\
\frac{2}{n} \sum_{i=1}^{n} \int_T (X_{i1}(t) - X_{ip}(t - \tau_{1p})) X'_{ip}(t - \tau_{1p}) dt \\
\frac{2}{n} \sum_{i=1}^{n} \int_T (X_{i2}(t) - X_{i3}(t - \tau_{23})) X'_{i3}(t - \tau_{23}) dt \\
\vdots \\
\frac{2}{n} \sum_{i=1}^{n} \int_T (X_{i2}(t) - X_{ip}(t - \tau_{2p})) X'_{ip}(t - \tau_{2p}) dt \\
\vdots \\
\frac{2}{n} \sum_{i=1}^{n} \int_T (X_{i(p-1)}(t) - X_{ip}(t - \tau_{34})) X'_{ip}(t - \tau_{(p-1)p}) dt
\end{bmatrix}
\]

and \( V_{\tau_0} \) is the Hessian of \( \Lambda(\tau) = E(m_\tau) \) at \( \tau_0 \). Finally, we apply the linear transformation \( \hat{\theta} = \frac{1}{p} A^T \hat{\tau}^* \), where \( \hat{\tau}^* \) and \( \hat{\theta} \) are as in equations (5) and (6), respectively, to obtain the result:

\[
\sqrt{n}(\hat{\theta}_n - \theta_0) \overset{d}{\sim} \mathcal{N}(0, \Sigma_p)
\]

where \( \Sigma_p = \frac{1}{p^2} A^T V_{\tau_0}^{-1} E[\nabla \mu_{\tau_0} \nabla \mu_{\tau_0}^T] V_{\tau_0}^{-1} A \).