Template Priors in Bayesian Curve Registration

W. Zachary Horton
Department of Statistics
University of California, Santa Cruz

Garritt L. Page
Department of Statistics
Brigham Young University

C. Shane Reese
Department of Statistics
Brigham Young University

Lindsey K. Lepley
School of Kinesiology
University of Michigan

McKenzie White
School of Kinesiology
University of Michigan

Abstract

In experiments where observations on each experimental unit are functional in nature, it is often the case that, in addition to variability along the horizontal axis (height or amplitude variability), there are also lateral displacements/deformations in curves (referred to as phase variability). Unlike the former, the latter form of variability is often treated as a nuisance parameter when making inferences. Therefore, it is common in functional data analysis to reduce this variability by aligning curves through a process called curve registration. Often, expert knowledge regarding the location and time that certain curve features occur is available to guide the curve realignment. We propose a Bayesian model that permits incorporating this knowledge when registering curves using a Gaussian process prior formulation. This novel approach capitalizes on the interpolation property of predictive distributions from Gaussian processes while still preserving the flexibility found in modern registration techniques. We detail computational strategies and illustrate the utility of the method through a simulation study and an analysis of knee-power biomechanics. Supplementary materials for the article are available online.

1 Introduction

Biomechanics studies require subjects to negotiate movement tasks that highlight biomechanical systems of interest and suggest effective treatment strategies. Typically, interest lies in being able to determine differences in mechanics between injured and healthy subjects or between injured and non-injured limbs (i.e., inter-limb asymmetry). For example, White et al. (2019) uncover inter-limb knee asymmetries that are masked during clinical hop tests designed to evaluate athlete readiness to return to sport following injury. This was done by continuously monitoring joint angles and ground reaction forces generated throughout a movement to determine knee power curves or knee power profiles. The continuous monitoring produces functional data for each subject. Analyzing these functional curves and the extent to which they are affected by injury or insufficient rehabilitation after injury can reveal underlying asymmetries. Figure 1.1 displays the injured and healthy power profile curves for each of the 16 subjects in the White et al. (2019) study. Each

gray curve in Figure 1.1 represents the continuous power in watts (scaled by subject weight in kilograms) for one subject over the stance phase (time), normalized as a percent of stance phase.

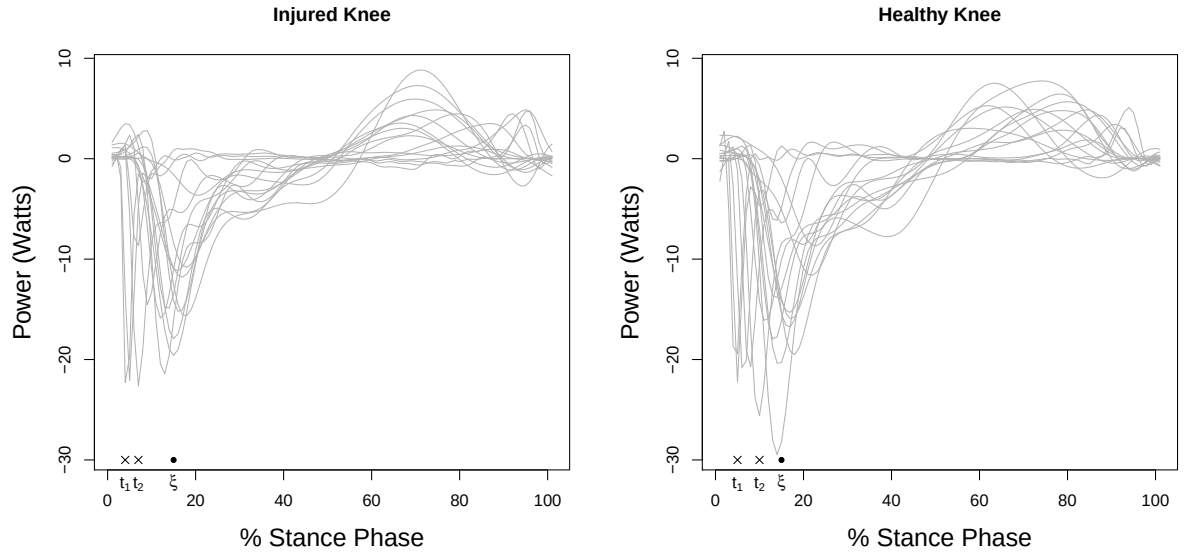

Figure 1.1: Knee power curves. The x-axis is the percent stance phase of the hop completed and the y-axis is the power measured in Watts.

Historically, rather than consider the entire power profile curve like those in Figure 1.1 when carrying out analyses, biomechanists reduce curves to one or two values of interest (e.g., value at maximum peak) and perform univariate statistical analyses such as analysis of variance or a t-test to compare groups such as healthy subjects and injured subjects. Recently, however, the biomechanics literature has migrated to analyses that consider the entire movement curve (e.g., Hopkins et al. 2013, Koshino et al. 2016, Seeley et al. 2017, Hopkins et al. 2019). Utilizing entire curves takes into account the functional nature of the data, but also introduces complications that are not present in a univariate analysis. One such complication commonly encountered in functional data analysis (FDA) is the existence of variability along the horizontal axis (phase variability or misalignment) in addition to that along the vertical axis (amplitude variability). It is common to accommodate the latter in a statistical analysis, while the former is often of little interest. To illustrate the two types of variability, consider Figure 1.2. The left plot in the figure demonstrates amplitude variation as curves follow the same shape, but vary on the height or depth at certain times, while the right plot displays phase variability as curves are offset. Notice that in the presence of phase variability, using a cross-sectional mean at each time to produce a "representative" curve is problematic as the resulting mean curve may not be representative of any of the observed curves, particularly at local extrema. Because of the misalignment, in practice it is common to remove phase variability through a procedure called *curve registration* or *curve alignment*.

Handling misaligned curves through registration or alignment has a rich literature. An early treatment of curve registration is Sakoe and Chiba (1978) who align two curves based on certain features in the context of speech recognition. Since then a substantial amount of effort has been dedicated to developing

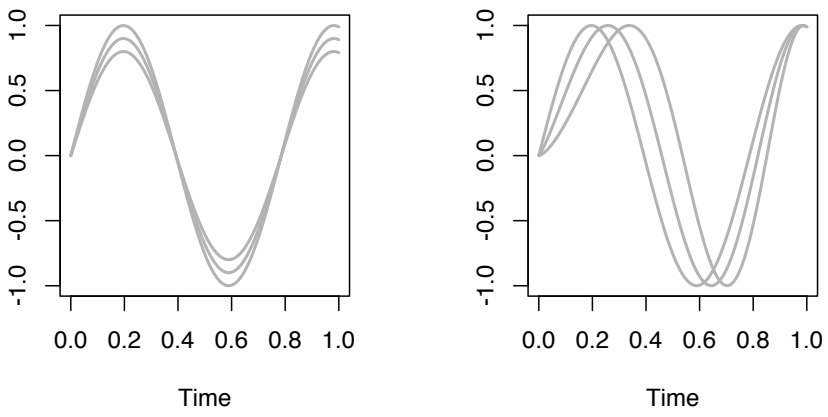

Figure 1.2: Example of amplitude variation (left) and phase variation (right).

methods that perform alignment by decomposing a curve into "amplitude" and "warping" functions where the former measures differences on the y-axis and the latter on the x-axis. Early attempts in the literature such as Ramsay and Dalzell (1991) and Lucero et al. (1997) seek to align curves for functional data analysis through two-step procedures in which a warping curve is estimated and then curves are registered before, and asynchronous to model fitting. More recent statistical work incorporates curve registration as an integral part of the model building process and we focus on these.

Landmark methods carry out registration by aligning or connecting curves based on particular curve features (see Ramsay and Silverman 2005, Wang et al. 2016 for an introduction). These methods are considered to be the gold standard assuming clearly identified landmarks exist on each curve (Wang et al. 2016). Kneip and Gasser (1992) are the first to consider this method and explore theoretical properties, while Kneip and Ramsay (2008) explore statistical properties of a warping function that incorporates landmarks. Andersen and Runger (2012) estimate landmark locations and used them as predictors and Srivastava et al. (2011) employ a geometric approach that matches landmarks using the Fisher-Rao metric. Others define landmarks as locations of significant zero-crossings of the continuous wavelet transform (Bigot 2006) or as curve gradients in addition to peaks and valleys (James 2007).

Absent clear landmarks, an alternative approach is to warp time so that curves are aligned to some target function. An early approach to align curves using a target function was to estimate an unknown global shift (Ramsay and Silverman, 2005, chapter 7 provides an overview). Other methods have been proposed, most of which are built on work by Ramsay and Li (1998). Earls and Hooker (2017) and Telesca and Inoue (2008) carry out smoothing and registration through a Bayesian hierarchical model where both warping function and observed function are modeled with B-splines. James (2007) shrinks functions towards a common target shape when landmarks are absent. Rakêt et al. (2014) simultaneously carry out registration and smoothing using random effects modeling. Recently, Cheng et al. (2016) develop a curve registration method based on statistical shape analysis and Lu et al. (2017) employ Gaussian processes to flexibly estimate the warping

between curves. These procedures are able to very smoothly characterize the warping functions, but as mentioned, landmark information is absent even when certain curve features have been identified by an expert and their location is scientifically relevant.

Our approach is motivated by the desire to smoothly estimate the warping function, but also ensure that the landmarks on subject-specific unwarped curves align on landmarks at pre-specified, known, time points as informed by expert knowledge. To highlight the novel aspect of our approach, in Figure 1.1 we draw attention to the times at which the minimum occurs for two subjects (t_1 and t_2) and the time at which expert knowledge expects the minimum to occur (ξ). (See Section 6 for the scientific rationale behind this.) The method developed by Telesca and Inoue (2008), called Bayesian Hierarchical Curve Registration (BHCR), would produce well-registered curves but the two minimums would not be guaranteed to align. Landmark registration methods would align the two curves so that the two minimums would occur at the same time, but the alignment would be guided through a rigid linear landmark interpolation (LLI). In addition, neither method would "anchor" the aligned curves at the correct location (ξ). The contribution of this paper can be thought of as a compromise between the BHCR and LLI methods and will be able to anchor observed landmarks at the times experts expect landmarks to occur. This last trait is one aspect that separates our approach to the that found in Bharath and Kurtek (2020) who produce warping curves by partitioning the time domain into disjoint sets defined by landmarks and in each segment estimate the curve. Thus, the main goal of this paper is to develop a hierarchical registration model that includes a landmark informed prior that makes use of available expert information in a very flexible way.

The rest of the article is organized as follows. We begin by reviewing the basics of registration and warping functions in Section 2.1 and Bayesian FDA in Section 2.2. We present our template prior formulation in Section 3 and associated model fitting details in Section 4. In Section 5 we compare methods using simulation studies. Finally, we perform data analysis on the knee power data set in Section 6 and discuss implications in Section 7.

2 Background and Preliminaries

In this section, we provide a brief overview of warping functions and the model we employ in our Bayesian FDA.

2.1 Warping Functions

Functional data are often measured over time, and consequently, curve registration is also known as time warping. The goal of registration is to map the original time t to a new time $t^\star$ for all t in the time domain and is typically carried out by way of a warping function $\omega(t) = t^\star$. Figure 2.1 provides a simple example of a warping function and illustrates how transforming time impacts the original curve. Note that the original curve (indicated by a dotted line on the left) has a maximum at $t = 0.5$ while the warped function attains its maximum at $t^\star = 0.3$. A visual inspection of the warping function $\omega(t)$ verifies that $\omega(0.5) = 0.3$.

The baseline warping function $\omega(t) = t$ corresponds to a line with slope one and intercept zero, also

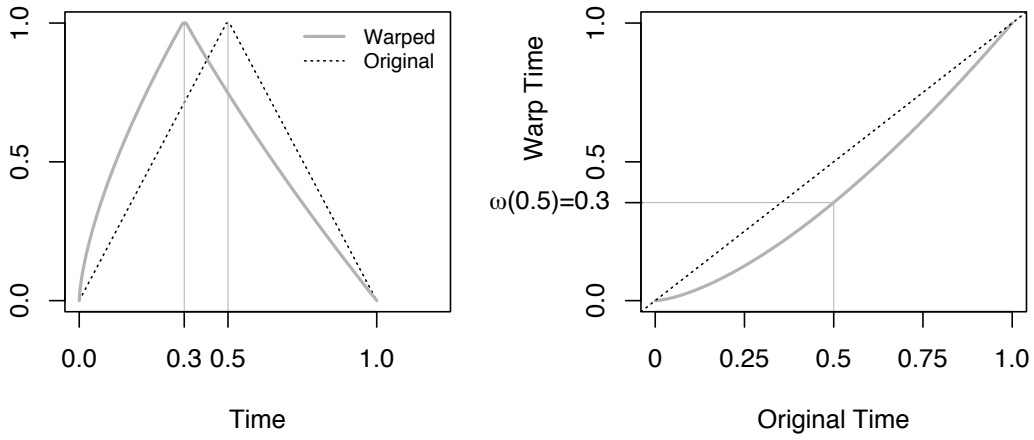

Figure 2.1: Example warping function $\omega(t)$ is depicted on the right. The resulting effect on an observed function is displayed on the left.

called the unit line, as this warping function does not alter time. Regions of a warping function that fall below this line (as in figure 2.1) lead to a compression of time. Similarly, time expansion occurs when the warping function raises above the unit line.

2.1.1 Warping Function Constraints

Not all real valued functions are permissible warping functions. Chapter 7 of Ramsay and Silverman (2005) and James (2007) discusses elements of valid warping functions. Telesca and Inoue (2008) further restrict attention to warping functions that comply with the following:

- $\lim_{t \to t_1} \omega(t) = \omega(t_1)$ (Continuous)
- $t_1 < t_2$ implies $\omega(t_1) \leq \omega(t_2)$ (Monotone increasing)
- $\omega(0) = 0$ (Curves preserve origination time)
- $\omega(1) = 1$ (Curves preserve termination time).

Continuity ensures that registration does not create gaps in time, monotonicity prevents partial time reversal, and end-point constraints force curves to preserve their boundary times after registration. In theory, the time domain can be any closed interval (i.e., $\mathcal{T} = [t_0, T_0]$), but it is common to normalize time so that $\mathcal{T} = [0, 1]$. Without loss of generality we formally adopt these constraints in the modeling that follows.

2.2 Functional Data Model with Warping

Here we detail a commonly used Bayesian functional data model when registration is carried out simultaneously with curve fitting (see Telesca and Inoue 2008). This model will form the basis of our approach.

For notational simplicity, we detail our approach assuming a balanced design. That is, measurements for each subject are taken at the same time points $\boldsymbol{t} = (t_1, \ldots, t_n)$. However, we make this assumption for expository simplicity and the balanced design assumption can be relaxed in practice. Now, let $y_i(t)$ denote the measured response from individual i at time point t and assume that

$$y_i(t) = a_i + c_i f(\omega_i(t)) + \epsilon_i(t), \text{ for i} = 1, \ldots, m, \tag{2.1}$$

where $f(\cdot)$ represents an unknown "global" measurement curve, $\omega_i(\cdot)$ is the ith subject's unknown warping function, $\epsilon_i(t) \sim N(0, \sigma_\epsilon^2)$ is an error term with $N(v, s^2)$ denoting a normal distribution with mean v and variance s^2, and a_i and c_i are shift and scale parameters that account for amplitude differences between the individual warpings. Model (2.1) is employed in Telesca and Inoue (2008). Since a common shape function $f(\cdot)$ is assumed for each subject, it is equivalent to modeling observed curves as warped variants of a "central" curve. To facilitate the interpretability and identifiability of a_i and c_i we employ $\sum_{i=1}^m a_i = 0$ and $\sum_{i=1}^m c_i = m$ as identifiability constraints (see Telesca 2015).

Because the unknown object $f(\cdot)$ lives in an infinite dimensional space, it is common to reduce dimensionality by representing $f(\cdot)$ as a linear combination of p basis functions (e.g., B-spline basis). As a result, $f(\cdot)$ can be expressed as

$$f(w_i(t)) = \sum_{\ell=1}^{p} \beta_\ell H_\ell(w_i(t), \boldsymbol{\zeta}), \tag{2.2}$$

where $H_\ell(w_i(t), \boldsymbol{\zeta})$ denotes the ℓ-th B-spline basis function evaluated at warped time $\omega_i(t)$ for knots $\boldsymbol{\zeta}$, and β_ℓ the ℓth B-spline basis coefficient. The knot values span the range between the first and last measurement times, regardless of subject. In matrix notation, model (2.2) can be re-expressed as

$$\boldsymbol{y}_i = a_i \mathbf{1} + c_i \boldsymbol{H}(\omega_i(\boldsymbol{t}))\boldsymbol{\beta} + \boldsymbol{\epsilon}_i, \tag{2.3}$$

where $\boldsymbol{y}_i = (y_i(t_1), \ldots, y_i(t_n))'$ and $\omega_i(\boldsymbol{t}) = (\omega_i(t_1), \ldots, \omega_i(t_n))'$ denote the $n \times 1$ vectors of measured response and warped times for the ith subject, $\mathbf{1}$ is a $n \times 1$ vector of ones, $\boldsymbol{H}(\omega_i(\boldsymbol{t}))$ is the $n \times p$ B-spline design matrix, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)'$ is the $p \times 1$ vector of the B-spline basis coefficients, and $\boldsymbol{\epsilon}_i = (\epsilon_i(t_1), \ldots, \epsilon_i(t_n)) \sim N_n(0, \sigma_\epsilon^2 \boldsymbol{I})$. We note that p (the length of $\boldsymbol{\beta}$) depends on the dimension of $\boldsymbol{\zeta}$ and the degree of the B-spline basis (which is cubic in what follows).

To finish the model construction, a model for $\omega_i(\cdot)$ needs to be specified. Older methods would estimate a warping function before modeling, so it would simply be substituted in (2.3). More modern methods simultaneously fit $\omega_i(\cdot)$ and $f(\cdot)$. The curve shape variability in Figure 1.1 would make a parametric model difficult to estimate. Telesca and Inoue (2008) consider a nonparametric model by employing P-splines with a small number of knots to model $\omega_i(\cdot)$. However, when performing registration, expert information is often available regarding the location on the curve of prominent features (i.e., landmarks) and the approximate time at which they occur. One could consider this information (which usually comes as a vector of feature locations) as forming a "template" feature set which helps guide registration. Ideally, a registration method

would be able to incorporate this information so that features of registered curves are aligned at the prespecified feature locations. The methods that model $\omega_i(\cdot)$ flexibly (such as BHCR) are unable to utilize expert input explicitly in the model, while methods that incorporate these landmarks are often not sufficiently flexible. Our contribution to the curve registration literature is to develop a method that balances these two objectives; specificity and flexibility.

3 Template Prior Registration Model

Landmarks, in principal, can be any prominent curve feature or simply a location on the curve that is of interest to a researcher. One common landmark in the types of curves we consider are points whose derivatives are zero (i.e., critical points). These types of landmarks are considered in the simulation study of Section 5. In our application, in addition to these types of landmarks, we consider curve roots or zero-crossings as well. critical points and zero-crossing both have biomechanical connections. More details are provided in Section 6. Landmarks need not be defined using a mathematical characteristic such as a mode or a root. Rather, they can correspond to any point of interest which can be associated with a specific time or location on a curve. That said, having a mathematical definition, is more convenient when manual landmark identification is tedious or infeasible as algorithms can automate the process. Although landmarks may be specified to have occurred at any time in the time domain, it is computationally convenient to select observed times as landmark locations. In cases where landmark definitions are not met exactly at observed times, selecting the nearest point is reasonable.

In order to ensure that the warping functions we consider are informed by a "template" landmark set, $\omega_i(\cdot)$ must align the occurrence of K landmarks to times specified by an expert. To this end, let $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_K)$ denote the prespecified (i.e., known) times at which K landmarks occur and $\mathcal{L}_i$ the collection of time indices at which the K landmarks occurred for the ith subject. (Note that both $\boldsymbol{\xi}$ and $\mathcal{L}_i$ are known in our application. In Section 7 we discuss how our approach can be employed when one or both is unknown.) Thus, the observed times for the ith subject can be partitioned such that $\boldsymbol{t} = (\boldsymbol{t}_{\mathcal{L}_i}, \boldsymbol{t}_{\mathcal{L}_i^c})$ where $\boldsymbol{t}_{\mathcal{L}_i}$ denotes the collection of times where the K landmarks occurred for the ith subject and $\boldsymbol{t}_{\mathcal{L}_i^c}$ denotes the ith subject's remaining $n-K$ time points. With this in mind, the additional constraint taken on by the warping functions we consider is $\omega_i(\boldsymbol{t}_{\mathcal{L}_i}) = \boldsymbol{\xi}$ for all i, resulting in the following class of warping functions

$$\{\omega : [0,1] \to [0,1] \mid \omega(0) = 0,\ \omega(1) = 1,\ 0 \leq \frac{d\omega}{dt} < \infty,\ \omega(\boldsymbol{t}_{\mathcal{L}}) = \boldsymbol{\xi}\}. \tag{3.1}$$

To incorporate the template feature constraints and preserve the desire for smooth warping curves, we assign to $\omega_i(\cdot)$ a Gaussian process. The predictive distribution based on a Gaussian process will be informed by landmark time occurrences so that the support of the resulting posterior distribution for the warping functions is restricted to a class of functions that pass through $\boldsymbol{\xi}$ (more details are provided in Section 3.1). To ensure that the support of the posterior distribution is also restricted to monotone functions, we employ the methodology found in Lin and Dunson (2014) who advocate projecting the posterior distribution of $\omega_i(\cdot)$ based on a Gaussian process onto the space of monotone functions. The result of the projection is an induced

probability measure on the space of monotone functions upon which inferences are based. As mentioned in Lin and Dunson (2014), this approach results in a posterior distribution for $\omega_i(\cdot)$ in an empirical Bayesian sense, but not a fully Bayesian sense.

As mentioned, we propose modeling $\omega_i(\cdot)$ with a Gaussian process. The Gaussian process is centered on the identity function, which corresponds to no warping, by setting the mean function equal to $m(\boldsymbol{t}) = \boldsymbol{t}$. The covariance function we employ is the squared exponential so that

$$cov(\omega_i(t), \omega_i(t')) = K(t, t'; \alpha, \lambda^2, \eta) = \lambda^2 \left(\exp\left(\frac{-\alpha(t-t')^2}{2} \right) + \eta I(t = t') \right). \tag{3.2}$$

Here α is the inverse range parameter that determines the smoothness of the curve, λ^2 is the multiplicative variance term, $\lambda^2\eta$ is the nugget effect where $I(\cdot)$ denotes an indicator function, and $K(\cdot)$ is the squared exponential covariance function. The *a priori* assumption of no warping is consistent with the prior mean in the method of Telesca and Inoue (2008). In addition to numerical stability, including a nugget effect in the Gaussian process specification accommodates any uncertainty an expert may have regarding the specific time at which a particular landmark is expected to occur. In the event that expert opinion is precise, one could simply set $\eta = 0$ (although this may introduce numerical issues, see Section 3.2).

3.1 Joint Data Model

We have that $\omega_i(\boldsymbol{t}_{\mathcal{L}_i})$ is observed and equal to $\boldsymbol{\xi}$ for each $i = 1 \ldots, m$. This added information should be considered when constructing a data model and thus the data model in (2.3) needs to be augmented by $\omega_i(\boldsymbol{t}_{\mathcal{L}_i})$. Our approach to building the joint model is based on the entire $\omega_i(\boldsymbol{t}) = (\omega_i(\boldsymbol{t}_{\mathcal{L}_i}), \omega_i(\boldsymbol{t}_{\mathcal{L}_i^c}))$ where we treat $\omega_i(\boldsymbol{t}_{\mathcal{L}_i^c})$ as "missing" and use a latent variable missing data model (Daniels and Hogan 2008). As a result, for the ith subject the full joint data model based on (2.3) and a Gaussian process for $\omega_i(\cdot)$ is

$$\begin{aligned} p(\boldsymbol{y}_i, \omega_i(\boldsymbol{t}_{\mathcal{L}_i}), \omega_i(\boldsymbol{t}_{\mathcal{L}_i^c}) \mid a_i, c_i, \boldsymbol{\beta}, \sigma^2, \alpha, \lambda^2, \eta) &= p(\boldsymbol{y}_i \mid \omega_i(\boldsymbol{t}_{\mathcal{L}_i}), \omega_i(\boldsymbol{t}_{\mathcal{L}_i^c}), a_i, c_i, \boldsymbol{\beta}, \sigma^2) \times & (3.3) \\ & p(\omega_i(\boldsymbol{t}_{\mathcal{L}_i}) \mid \alpha, \lambda^2, \eta) \times & (3.4) \\ & p(\omega_i(\boldsymbol{t}_{\mathcal{L}_i^c}) \mid \omega_i(\boldsymbol{t}_{\mathcal{L}_i}), \alpha, \lambda^2, \eta). & (3.5) \end{aligned}$$

In this joint data model (3.3) and (3.4) together form the likelihood (both are functions of observed quantities) and (3.5) is the latent variable missing model. Note that Gaussian error in (2.3) and modeling the $\omega_i(\cdot)$ with a Gaussian process make it so that (3.3) - (3.4) are Gaussian densities. In particular, note that (3.5) is the predictive distribution based on a Gaussian process which will produce warping curves that pass through $\omega_i(\boldsymbol{t}_{\mathcal{L}_i}) = \boldsymbol{\xi}$. In Section A.1 of the online supplementary material we provide the exact forms of (3.3) - (3.5). To ensure that the warping curve is monotone (i.e., $0 \leq \frac{d\omega_i}{dt} < \infty$), $\omega_i(\cdot)$ is projected into a monotone functional space by finding

$$\omega_i^* = \underset{\omega' \in \mathcal{M}}{\arg\min} \int_{\mathcal{T}} [\omega(t) - \omega'(t)]^2 dt, \tag{3.6}$$

where $\mathcal{M}$ denotes the space of monotone functions (see Lin and Dunson 2014 for more details).

3.2 Prior Distributions

To finish our model formulation, prior distributions for $(\{a_i\}, \{c_i\}, \boldsymbol{\beta}, \sigma_\epsilon^2, \alpha, \lambda^2)$ need to be assigned. Recall that $\boldsymbol{\beta}$ corresponds to a vector of B-spline coefficients whose dimension depends on the number of knots. Decisions associated with number and location of knots in $\boldsymbol{\zeta}$ (see equation (2.2)) are important as they affect the resulting curve fit. To facilitate this process, we select a large number of evenly spaced knots and employ P-splines developed by Lang and Brezger (2004). This produces the following hierarchical prior for $\boldsymbol{\beta}$

$$\begin{aligned} \boldsymbol{\beta}|\tau^2 &\sim N_p(\mathbf{0}, \tau^2\mathbf{P}^{-1}) \\ \tau^2 &\sim InverseGamma(a_t = 0.1, b_t = 0.1), \end{aligned} \tag{3.7}$$

where $\mathbf{P}$ is a $p \times p$ first order penalty precision matrix and τ^2 is a "smoothing" parameter. To ensure that $\mathbf{P}$ is full rank, Telesca and Inoue (2008) change the value of $\boldsymbol{P}$ in the first row and column from one to two, a procedure we adopt. For completeness, the $\mathbf{P}$ matrix we employ is provided Section A.2 in the online supplementary material.

For the amplitude shift and scale we use $a_i \sim N(0, \sigma_a^2)$ and $c_i \sim N(1, \sigma_c^2)$ for $i = 1, \ldots, m$ as they preserve constraints in prior expectation. We also assign commonly used $InverseGamma(0.1, 0.1)$ priors to σ_a^2 and σ_c^2. For the smoothing parameter, α, employing a diffuse prior is not recommended (Banerjee et al. 2015). Thus we use the following fairly informative prior for $\alpha \sim Uniform(a_a = 0, b_a = n)$ where $b_a = n$ is selected to impose a minimum amount of significant correlation (recall α is the inverse range). Values much more than n are nearly unidentifiable and lead to numerical instability in the MCMC algorithm employed. We set $\lambda^2 \sim InverseGamma(a_l = 0.1, b_l = 0.1)$ which is a commonly used diffuse priors and $\sigma_\epsilon^2 \sim InverseGamma(a_s, b_s = 0.1)$ where a_s is a user-supplied value that is selected on a case-by-case basis depending on the amount of measurement error that is present in the measuring process. In all our simulations and analysis we employ $a_s = m \times n$. The final step in model specification is selecting a value for η. For the analyses considered hereafter, we use $\eta = 0.005$ which reflects precision of information regarding landmark locations. We note that smaller values do not appear to provide much benefit, but frequently result in computational instability.

4 Model Fitting Details

The joint posterior distribution of model parameters given data is not analytically tractable. Thus, we resort to sampling from the posterior using a fairly straight-forward Gibbs sampler. The full conditionals for $\{a_i\}$, $\{c_i\}$, $\boldsymbol{\beta}$, σ_ϵ^2, τ^2, σ_a^2, σ_c^2, and λ^2 are of recognizable form while those for α and $\omega_i(\cdot)$ are not. Gibbs steps are used to update conditionally conjugate parameters. The full conditionals of these parameters follow well known arguments and are provided in Section A.1 of the online supplementary material. To update α, we employ a random walk Metropolis step and a normal distribution to generate candidate values. The more novel aspect of the MCMC algorithm we construct is associated with updating $\omega_i(\cdot)$ for each subject. We use a random walk Metropolis step with a suitably selected multivariate normal as a candidate generating distribution. The candidate generating distribution is constructed so that landmark constraints

are preserved. As a result all accepted candidate warping functions pass through landmarks. If a draw from the full conditional of $\omega_i(\boldsymbol{t}) = (\omega_i(\boldsymbol{t}_{\mathcal{L}_i}), \omega_i(\boldsymbol{t}_{\mathcal{L}_i^c}))$ is not monotone, we employ the algorithm described in Lin and Dunson (2014) to project the warping curve into the monotone functional space determined by (3.6). This results in a draw from the empirical posterior distribution whose support is the class of functions defined in (3.1). The exact form of the proposal density employed and more specific details of the model fitting algorithm are provided in Section A.1 of the online supplementary material.

4.1 Registering Curves

There are two approaches available to register curves once the model given by (2.3) is fit. The first, which is very natural from a Bayesian perspective, is to consider the posterior mean of $a_i\mathbf{1} + c_i\boldsymbol{H}(\boldsymbol{t})\boldsymbol{\beta}$ (denoted by $\hat{a}_i\mathbf{1} + \hat{c}_i\boldsymbol{H}(\boldsymbol{t})\hat{\boldsymbol{\beta}}$). The absence of $\omega_i(\cdot)$ is key. Recall that the model attempts to fit observed curves as warped versions of a "central" curve. This quantity represents a scaled version of the central curve, in which phase variation is removed. Note that these registered curves will be identical from subject-to-subject except for differences in a_i and c_i. This approach of alignment does not preserve the underlying shape of the measured warped curve.

The second approach is to "unwarp" the observed curves using the posterior mean of $\omega_i(\cdot)$ (denoted by $\hat{\omega}_i(\boldsymbol{t})$). This is done by fitting a linear interpolation function to the pairs of points defined by $\hat{\omega}_i(\boldsymbol{t})$ and $\boldsymbol{y}_i$. The result of the linear interpolation at points $\boldsymbol{t}$ yields an aligned curve which we denote using $y(\hat{\omega}^{-1}(t))$. This procedure of interpolation preserves the underlying shape of the originally observed curve. Because of this, we advocate using this approach to produce registered curves.

5 Simulation Study

As discussed previously, two common uses of registration are to accurately estimate a mean function and reduce phase variability. In this section we provide details associated with a numerical study that illustrates the performance of our registration approach in these areas. To this end, we generate synthetic functional realizations using two data generation scenarios, each based on a variety of mean curves and warping functions, and then compare how our method does at estimating the mean curve and at reducing phase variability relative to a number of other approaches.

5.1 Data Generation

We consider two simulation scenarios. Both consist of generating synthetic data sets under various mean and warping functions. Each data set generated consists of $m = 25$ curves comprised of $n = 31$ time points. In all data generating scenarios landmarks are defined as critical points (i.e., points at which the first derivative is zero) and $\boldsymbol{\xi}$ is comprised of critical point times from the true mean or "population" curve.

5.1.1 Data Generating Scenario 1

Our first generation scheme is concerned with removing phase variation. To this end, we consider using randomly generated curves which vary only in phase. That is, there is neither amplitude variability (i.e., $\sigma_a^2 = \sigma_c^2 = 0$) nor observation error (i.e., $\sigma_\epsilon^2 = 0$). In this setting, curves can be aligned exactly, which not only provides a means to compare models, but also a measure of overall alignment quality. A curve is generated using a three step procedure:

1. Select a true mean function and a baseline warping function.
2. Modify the warping function by adding smooth Gaussian process noise.
3. Warp the mean curve using the modified warping function.

The mean functions we consider for this data generating scenario are given in Figure 5.1 and are labeled means 1, 2, and 3 from left to right respectively. These functions are used to assess how landmark count affects registration results where mean 1 has two landmarks, mean 2 has three landmarks, and mean 3 has six landmarks. These mean curves are nonparametric and are produced by using the B-spline coefficients found in Tables S.1 and S.2 of the online supplementary material.

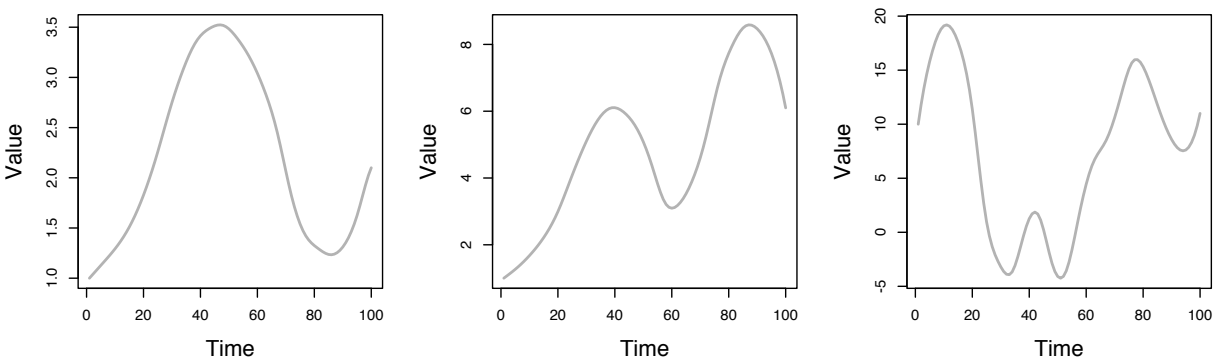

Figure 5.1: Mean functions for simulation study. From left to right, we label these curves as mean function 1, 2, and 3 respectively. The number of features (peaks and valleys) are two, three, and six respectively.

The baseline warping functions we consider for this data generating scenario are given in Figure 5.2 and are labeled warpings 1, 2, and 3 with warping 1 corresponding to the unit line (i.e., no warping) and warping 3 being the most severe. These are used to assess how increased systematic severity of misalignment affects registration results. Besides the unit line which corresponds to no systematic misalignment, these are circular arcs. See Section B.1 of the online supplementary material for details on how the warping curves were obtained.

We create phase variability by adding smooth Gaussian process noise to the warped mean curve. To do this we use a squared exponential covariance function with inverse range parameter of $\alpha = 5$, marginal variance of $\lambda^2 = 0.1$, and $\eta = 0$. The Gaussian process noise is generated independently for each subject.

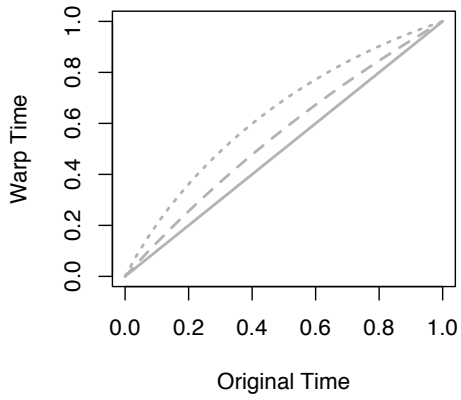

Figure 5.2: Warping functions for simulation study. We label these as warping function one (solid line), two (dashed line), and three (dotted line).

5.1.2 Data Generating Scenario 2

Our second generation scenario is concerned with curves which vary in both phase and amplitude but without observation error (i.e., $\sigma_\epsilon^2 = 0$). Additionally, the mean curves are not B-splines and the smooth noise added to the warping function is not from a Gaussian process (*i.e.* the functions are not from the model). The mean curves and baseline warping functions we consider are given in section B.1 of the online supplementary material. A curve is generated using a five step procedure:

1. Select a true mean function and baseline warping function.
2. Generate a vertical shift term a and a vertical scale term c from $N(0, 0.1^2)$ and $N(1, 0.1^2)$ respectively.
3. Modify the baseline warping function and the mean function by adding smooth noise.
4. Warp the modified mean curve using the modified warping function.
5. Scale the resulting curve by c and shift by a.

To generate smooth noise we use a B-spline with 4 basis functions and generate the coefficients independently from a $N(0, 0.05^2)$ distribution. We note that this noise generation procedure, as well as the mean curves and warping functions used for this study, are not devised from our proposed model.

In all data generating scenarios 100 synthetic datasets are created. An example dataset under each combination of mean curve and warping function is provided in Section B.1 of the online supplementary material.

5.2 Metrics To Compare Warped Curve Fits

To compare the performance of the procedures included in the simulation study, we use a variety of metrics. The first measures each procedure's ability to estimate the true underlying mean function $f(\cdot)$. It is inspired

by the squared-difference integral and is given by

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{j=1}^{n}[\{\mathbf{H}(t_j)\hat{\boldsymbol{\beta}} - f(t_j)\}^2]},$$

where $\hat{\boldsymbol{\beta}}$ is the posterior mean vector of B-spline coefficients and in an abuse of notation, $\mathbf{H}(t_j)$ denotes the $p \times 1$ vector of B-spline basis values evaluated at the jth time point. The second metric we consider produces an overall measure of goodness-of-fit. Letting $\hat{y}_i(t) = \hat{a}_i + \hat{c}_i\mathbf{H}(\hat{\omega}_i(t))\hat{\boldsymbol{\beta}}$ denote the posterior mean of the warped curve for subject i at time t, our goodness-of-fit metric is given by

$$\text{GOF} = \sqrt{\frac{1}{n}\frac{1}{m}\sum_{i=1}^{m}\sum_{j=1}^{n}[\{y_i(t_j) - \hat{y}_i(t_j)\}^2]}.$$

This metric measures how closed the fitted curves are to their corresponding observed curves. Lower values indicate better individual curve fits. The last metric we consider measures the variability between the n aligned curves. As mentioned in Section 4.1, $y_i(\hat{\omega}_i^{-1}(t))$ denotes the aligned observed curve for the ith subject where $\hat{\omega}_i(t)$ is the posterior mean of the warping function. The aligned curve variability metric is given by

$$\text{cVAR} = \sqrt{\frac{1}{n}\frac{1}{m}\sum_{i=1}^{m}\sum_{j=1}^{n}\left[\{y_i(\hat{\omega}_i^{-1}(t_j)) - \frac{1}{m}\sum_{k=1}^{m}y_k(\hat{\omega}_k^{-1}(t_j))\}^2\right]}.$$

We include the cVAR in our study as it seems desirable for an alignment procedure to produce aligned curves that display reduced variability, as this would result in more precise inference regarding a population function.

5.3 Competitors

As competitors in the simulation study we consider BHCR and linear landmark interpolation (LLI). Like BHCR, the data model of LLI is (2.3). However, unlike BHCR (and our method) $\omega_i(t)$ is a fixed, known piece-wise linear function that connects landmarks linearly. We include both methods as our approach can be thought of as a compromise between the two. In addition to BHCR, we consider two other nonlandmark-based approaches. The first is the elastic square-root slope (ESRS) approach (Tucker et al. 2013) which is carried out using the `time_warping` function found in the `fdasrvf` `R`-package (Tucker 2020). The second is a Bayesian square-root slope (BSRS) function approach (Cheng et al. 2016) which is carried out using the `function_group_warp_bayes` found in the same `R`-package. Default parameters settings were employed for both the ESRS and BSRS approaches.

5.4 Results of Simulation Study

To each synthetic data set created, we fit our model by collecting 1,000 MCMC iterates after discarding the first 5,000 as burn-in and thinning by 5. We use the same MCMC specifications to fit the BHCR and

LLI warping procedures. All computation associated with our method, BHCR, and LLI was carried using the R-package `warptk` which is available at `https://github.com/wzhorton/warptk`. For the procedures that can be found in the `fdasrvf` R-package, default settings were employed. For numerical stability we fix $\eta = 0.005$.

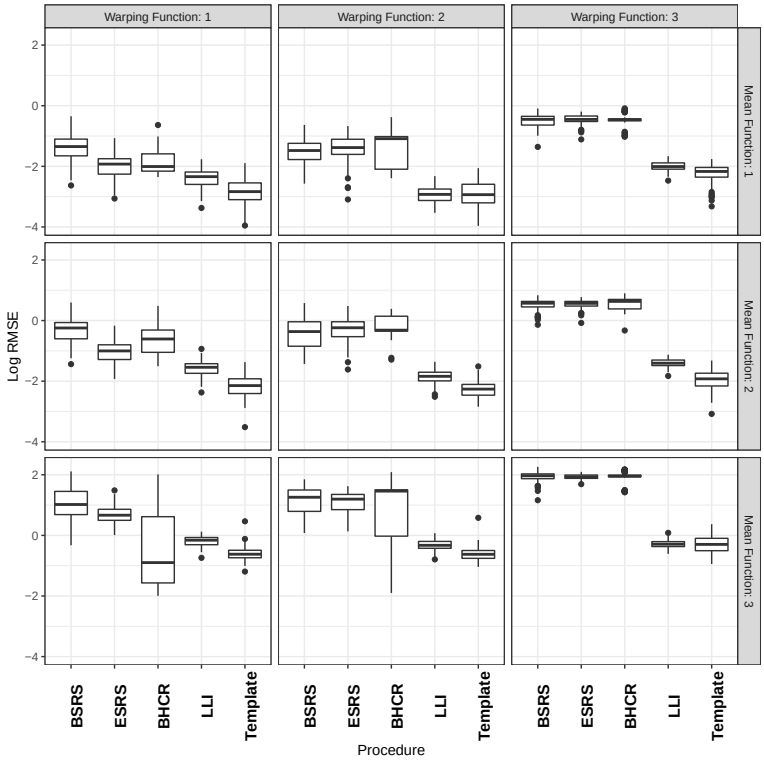

Figure 5.3: Boxplot grid showing Log RMSE values under combinations of warping functions (x-axis) and mean functions (y-axis) using the five considered methods for the first data generating scenario.

Here we focus on results under the first data generating scenario. The results under the second are provided in passing with more detail provided in the online supplementary material. Figures 5.3 - 5.5 shows boxplots of the log RMSE, log GOF, and log cVAR values. We also provide in the supplementary material Tables S.3 - S.5 that contain the average RMSE, GOF, and cVAR values across the 100 synthetic data sets for each data generating scenario.

From Figure 5.3 the LLI and our template prior approach perform the best at recovering the underlying mean curve based on RMSE for all scenarios except the most complex mean function with the identity warping function. Between LLI and our approach, the template prior method performs best for all warping and mean functions. The other three methods perform similarly and tend to degrade more as the warping becomes more extreme and the mean functions more complex relative to the landmark based methods. This highlights the benefit of including landmark information that not only aligns curves, but also anchors them at the correct time points. The non-landmark based methods suffer from an identifiability issue as they are not able to determine where the aligned curves should be "pegged". Figure S.3 of the online supplementary material displays the log RMSE values in the second data generating scenario. There LLI outperforms our

template prior approach based on RMSE. However, our template prior approach is still very competitive to the other methods and performs better at recovering the true mean curve when observed curves display the same type of phase variability (i.e., either time compression or expansion).

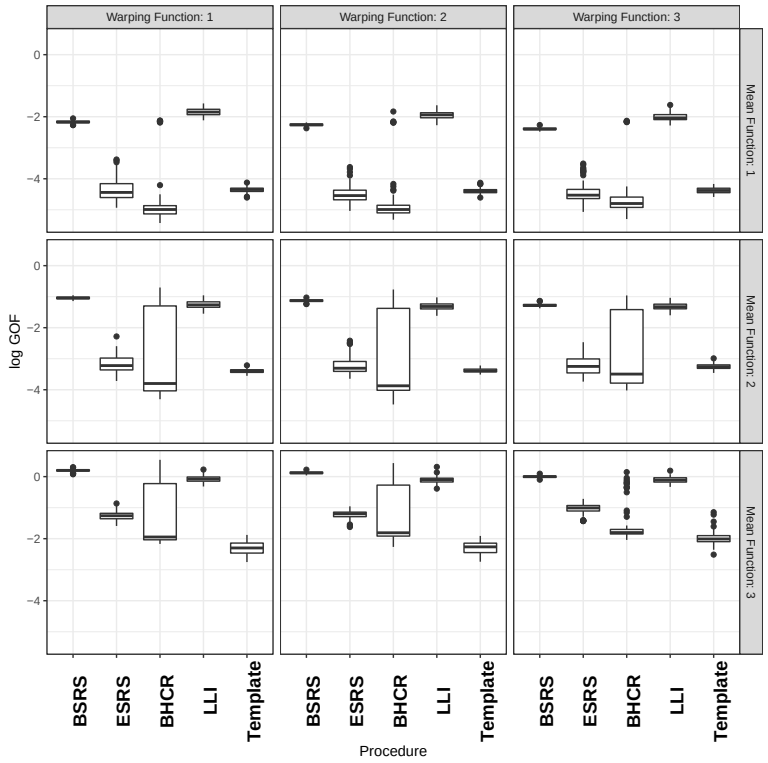

Figure 5.4: Boxplot grid showing Log GOF values under combinations of warping functions (x-axis) and mean functions (y-axis) using the five considered methods for the first data generating scenario.

Figure 5.4 displays the log Goodness-of-fit metric. Here, the results are more mixed. The ESRS and BHCR outperform our template prior approach in scenarios with simpler mean functions. Our approach performs best in the scenario with the more complex mean function. This is expected as more information is included in the landmark approaches. It seems that the type of warping doesn't impact much goodness-of-fit. Overall, the LLI approach performs the worst. Figure S.4 displays the same metric in the second data generating scenario. There ESRS method clearly performs best with our method outperforming the other approaches.

Figure 5.5 displays the results for the log cVAR metric which measures the variability between the aligned curves. Here BHCR seems to perform quite well with the simpler mean functions but our approach does better for the more complex mean functions. In addition, our approach seems to degrade less as the warping becomes more extreme. LLI in particular performs poorly while the BSRS and ESRS perform similarly. Figure S.5 of the online supplementary material displays the log cVAR metric for the second data generating scenario. There, all methods perform similarly with ours producing the smallest curve-to-curve variability for more complex mean functions.

In short, the take-home message of the simulation study is that our template prior approach seems to

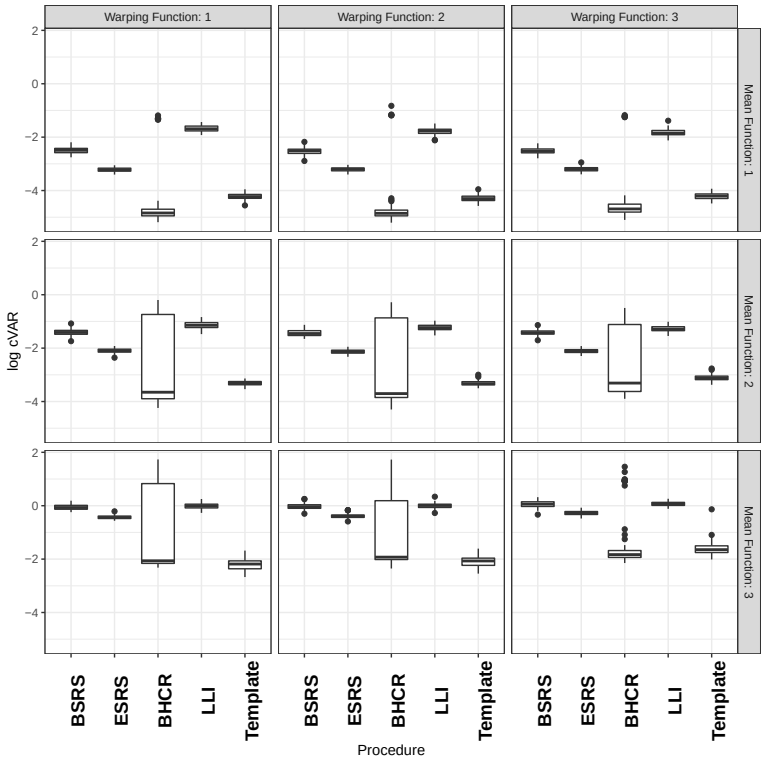

Figure 5.5: Boxplot grid showing Log cVAR values under combinations of warping functions (x-axis) and mean functions (y-axis) using the five considered methods for the first data generating scenario.

balance well the two main purposes of aligning curves in a statistical analysis; reducing variability between curves and bias when estimating a mean curve.

6 Case Study: Knee Power

We now return to the knee power profile data described in Section 1. As mentioned, the biomechanists are very interested in determining the difference in power produced by a healthy and reconstructed knee. We investigate the impact that a severe injury has on the knee's ability to produce power.

6.1 Registration of Knee-Power Data

Recall that White et al. (2019) seek to uncover inter-limb asymmetries after knee injury using clinical hop tests. To this end, the 16 subjects in their study have undergone anterior cruciate ligament reconstruction (ACLR) on one knee followed by a period of rehabilitation. Comparing power profiles between a subject's ACLR knee hop test to the corresponding healthy knee hop test may reveal hidden asymmetries that more common peak value analyses may not capture. White et al. (2019) administer the triple hop for distance task, which requires a subject to perform 3 consecutive maximal hops forwards on the same limb. Subjects were allowed two practice trials for each limb and data collection was concluded when three successful trials were attained for each limb. A trial is considered successful and recorded if the participant's full foot lands

on a force plate on the second of three hops and if they are able to maintain balance during the landing of the last hop. In this study, a 12 camera motion capture system synchronized with 2 force plates was used to record kinematic and kinetic data at 240 Hx and 1200 Hz respectively. Retroreflective marker trajectories were exported to Visual3D software where knee angles were determined using the default Cardan sequence method (X-Y-Z convention). Knee moments were subsequently determined using a standard inverse dynamics approach. Joint power was then computed as the product of joint moment and angular velocity. White et al. (2019) describe the full experimental setup in detail. The data are then time normalized to 100% of ground contact. This normalized axis is usually referred to as the stance phase amount or percent stance phase. Ground contact is defined as beginning with the heel strike (defined as the point when the vertical ground reaction force exceeds 10N) and ending with the take-off (defined as the point when the vertical ground reaction force is less than 10N). The measured outcome of each successful trial is the power profile of the knee during the second hop. This profile or curve measurement is a clinically meaningful outcome as it represents the capacity of the knee muscles to manage load.

For clarity, we briefly connect the power curves to observable physical movements. The initial power reading begins with the heel strike. The negative peak occurs near 15% stance phase and is the moment of maximal work done to counteract downward movement. Near 50% stance phase power reaches zero as the subject reaches maximum knee flexion and ceases vertical movement, marking the end of the landing phase and the beginning the take-off phase. Power peaks during take-off around 80% stance phase as the subject works to exert enough force to jump. The end of the curve occurs with take-off. As a result, the landmarks identified by the biomechanist are the global minimum, maximal, and root zero which are expected to occur at 15%, 50%, and 80% stance phase.

Using these landmarks, we fit our template prior model separately to power curves produced by ACLR and healthy knees employing the same estimation scheme as that of the simulation studies except for one small change. Here, the we lower the upper uniform bound on α's prior distribution from n to $\frac{n}{2}$. This produces smoothness in the warping functions that is needed to avoid over-fitting the complex shapes of the observed data. BHCR employs a similar tactic by using a small number of basis functions. The aligned curves for the ACLR knee and resulting from our method are displayed in Figure 6.1 while results for healthy knees are provided in Figure S.10 of the online supplementary material. In addition to the aligned curves, we include 95% point-wise posterior credible intervals of the mean curve. These are obtained by examining the posterior distribution of $\mathbf{H}(t)\boldsymbol{\beta}$ along a grid of t values. In addition to our method, we also provide in Figure 6.1 the aligned curves based on LLI and BHCR with their corresponding mean curve and 95% point-wise posterior credible intervals.

Note that for all three registration methods it is more reasonable to estimate the mean curve by using cross-sectional averages compared to warped curves in Figure 1.1. These results also highlight some differences found between registration methods. In particular, some BHCR registered curves are poorly aligned at the minimum and maximum and the LLI aligned curves awkwardly towards the end of the stance phase. Also consider the posterior mean warping functions given in Figure 6.2. Although BHCR sometimes fails to align features which exhibit strong phase variation, it produces very flexible warping functions compared to LLI registration which is much more rigid. We also see that template prior registration is able to simultaneously

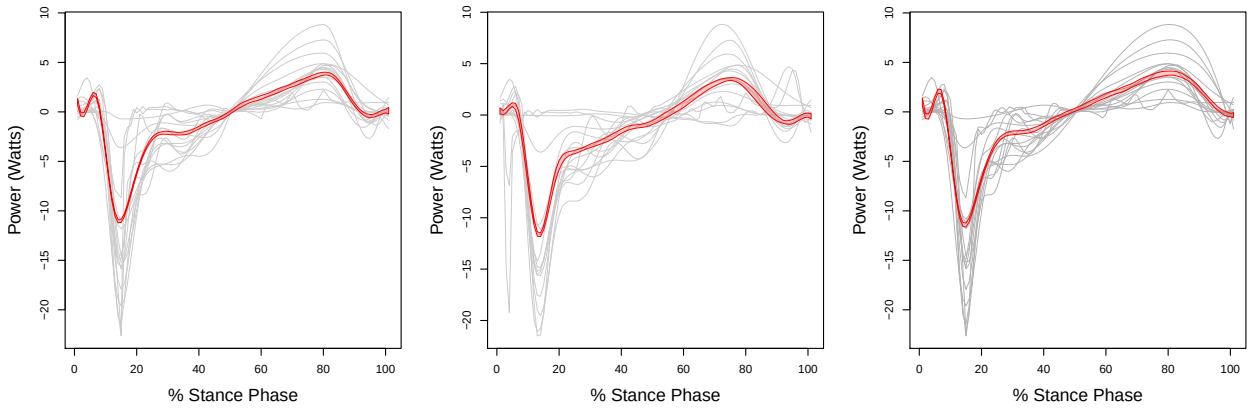

Figure 6.1: Registered ACLR knee power data with posterior mean curve bands. LLI left, BHCR middle, Template priors right.

align specified features while remaining quite flexible.

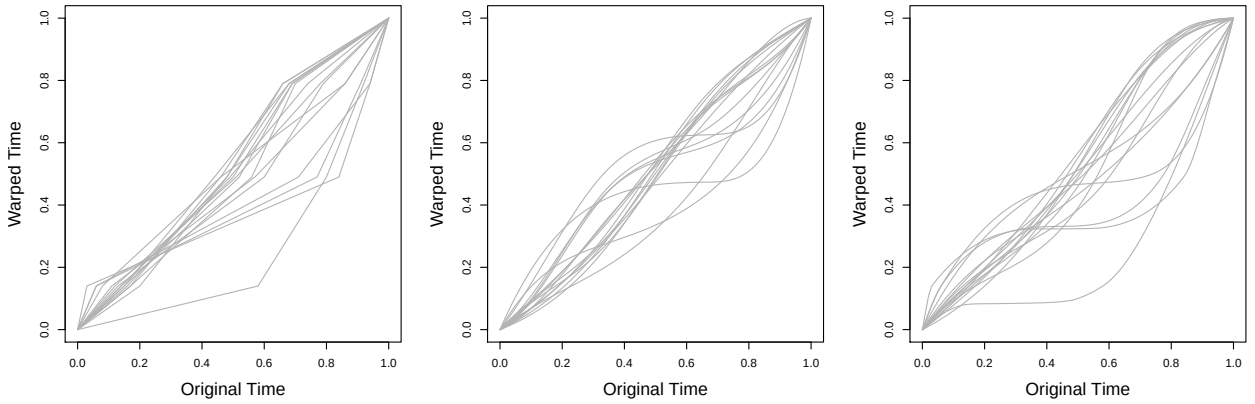

Figure 6.2: Estimated warping functions for the ACLR knee power data. LLI left, BHCR middle, Template priors right.

To further compare the three curve registration methods, we compute the cVAR metric described in the simulation study for all three alignments. Template prior registration has lowest value of 4.14 with LLI warping and BHCR having 4.28 and 6.24 respectively. For the healthy knee curves, our method performs even better with a cVAR value of 3.90 compared to 4.79 and 5.33 for the LLI and BHCR procedures respectively. See Figure S.10 of the online supplementary material to see the aligned healthy curves.

6.2 Difference Curve Analysis

Of specific interest to White et al. (2019) is understanding how power profiles differ between ACLR and healthy knees. To this end, in Figure 6.3 we provide each subjects difference curve (ACLR curve minus healthy curve). We also superimpose the cross-sectional mean difference curve.

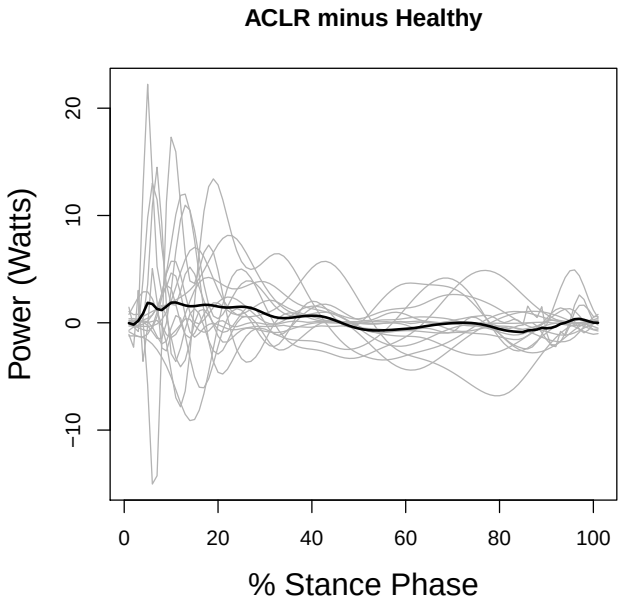

Figure 6.3: Knee power profile difference curves with cross-sectional mean given by a darker line.

At first glance, one might conclude that no meaningful difference between knee groups exists due to the flat mean difference curve. However, phase variation can diminish, mutate, or completely mask effects in differences. Thus, we construct difference curves based on the aligned ACLR and healthy curves obtained from the previous results, then estimate the average difference curve using the posterior distribution of $\mathbf{H}(t)\boldsymbol{\beta}_{\text{ACLR}} - \mathbf{H}(t)\boldsymbol{\beta}_{\text{Healthy}}$. These results are shown in Figure 6.4. Pronounced shapes in these difference curves are visible where they were masked previously due to phase variation. Notably, all three methods suggest a power increase - or rather a negative power decrease - near the first feature at 15% stance phase, which corresponds to injured knees producing less power during landing. The results showcase the usefulness of registration in unmasking potential patterns.

We note that the lack of variability at $t = 50\%$ in Figure 6.4 is expected. Recall that the central landmark we consider is the time at which the power curve crosses zero as a subject transitions from landing to launch. Our method (and LLI warping) ensures registered curves present this feature at $t = 50\%$, resulting in the pattern seen.

Figure 6.4 also highlights the advantage of landmark based alignment when analyzing differences. LLI warping and template priors produce flat mean difference curves which have the initial negative power decrease discussed. BHCR however introduces artificial patterns which do not exist in the data. Although phase variation has been removed from the ACLR and healthy groups separately, BHCR does not guarantee alignment between groups, thus producing artificial patterns in the difference curves. On the other hand, landmark methods do guarantee alignment between groups as long as the reference landmark locations are common, making them a convenient tool in these settings.

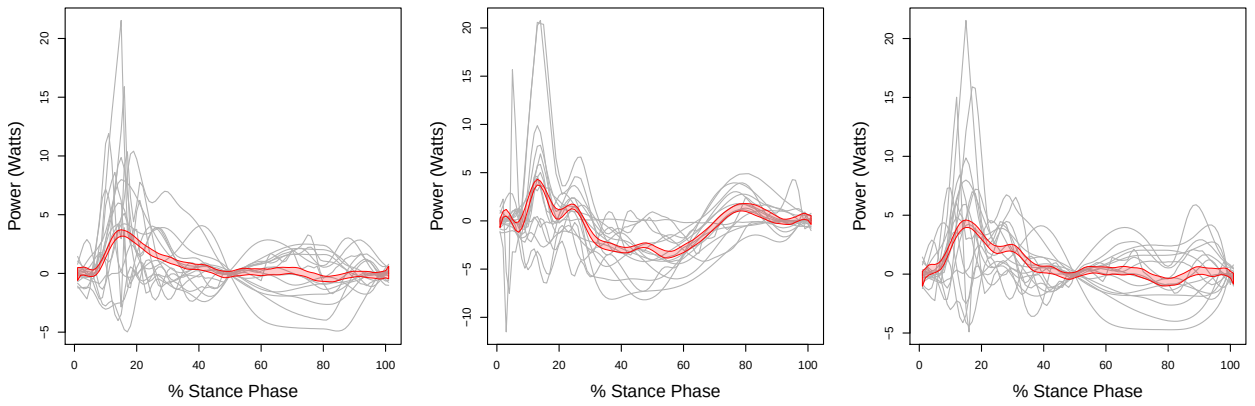

Figure 6.4: Aligned knee power profile difference curves with mean difference curve band given by a shaded region.

7 Discussion and Conclusion

We have developed a curve registration method that explicitly incorporates expert knowledge on the type and location of landmarks into a "template" prior. This method blends the reference feature mechanics of landmark matching with the flexibility of BHCR. Using simulation we show that template prior registration more accurately estimates underlying mean functions relative to non-landmark counterparts, but also produces aligned curves that are less variable compared to landmark based methods. We also demonstrate the utility of template prior registration on a knee power profile data set and show that it produced aligned curves that are less variable relative to other registration procedures.

While the development of our method was motivated by an exercise science study in which landmarks were clearly defined and experts had information about when they should occur *a priori*, our method can be effectively employed in other settings. For example, if landmarks are clearly defined but the expert is unsure when the landmarks should occur, then the time at which the landmarks occur is immaterial and as a result can be empirically assigned (e.g., through a mean occurrence time). In this setting "anchoring" the aligned curves at specific time points is not crucial and so inference for the average curve will not be impacted. We intuit that this would result in a method similar to that found in Bharath and Kurtek (2020). In situations where landmarks are not clearly defined our methodology can still be employed. In this setting $\mathcal{L}_i = \varnothing$ for all i and as a result $\boldsymbol{t} = \boldsymbol{t}_{\mathcal{L}_i^c}$ and the joint data model becomes

$$p(\boldsymbol{y}_i, \omega_i(\boldsymbol{t}) \mid a_i, c_i, \boldsymbol{\beta}, \sigma^2, \alpha, \lambda^2, \eta) = p(\boldsymbol{y}_i \mid \omega_i(\boldsymbol{t}), a_i, c_i, \boldsymbol{\beta}, \sigma^2)\; p(\omega_i(\boldsymbol{t}) \mid \alpha, \lambda^2, \eta). \tag{7.1}$$

We intuit in this setting that the warping functions estimated based on (7.1) will be similar to those of Telesca and Inoue (2008) and Lu et al. (2017).

Lastly, even though it is common in biomechanical applications to "normalize" curves so that each subject's measured curve begins and ends at the same time, it is conceptually straightforward to relax this

requirement in our methodology. This can be done by adopting the stochastic time model approach similar to that found in Telesca and Inoue (2008). As described there, we could consider a monotone map time-transformation from $\mathcal{T} = [t_0, T_0]$ to an interval $[t_0 + \delta_{i1}, T_0 + \delta_{i2}]$ where δ_{i1} and δ_{i2} are random. Pursuing this idea further is the topic of further research. In addition, addressing issues surrounding missing or extraneous landmarks, something that plagues most landmark methods, is a topic of future research.

Overall, we have shown that template prior registration is a flexible registration method, drawing on the benefits of a model-based approach, that also enables experts to incorporate their own subject matter expertise. The ability to utilize expert input along with modern techniques may be of great benefit to biomechanists and other practitioners.

8 Acknowledgments

The authors thank the editor, associate editor, and three anonymous referees for reviewing the article and providing valuable comments that greatly improved the quality of the paper.

9 Supplementary Material

The online supplementary materials for this article contains additional results from the simulation studies and the power knee application, as well as computer codes that were employed to run the simulation studies. The `R`-package `warptk`, available on the first author's Github page `https://github.com/wzhorton/warptk/`, contains codes for fitting the template prior method described in the paper.

References

Andersen, S. W. and Runger, G. C. (2012), "Automated feature extraction from profiles with application to a batch fermentation process," *Journal of the Royal Statistical Society: Series C*, 61, 327–344.

Banerjee, S., Carlin, B. P., and Gelfand, A. E. (2015), *Hierarchical Modeling and Analysis for Spatial Data*, Chapman and Hall/CRC, 2nd ed.

Bharath, K. and Kurtek, S. (2020), "Distribution on Warp Maps for Alignment of Open and Closed Curves," *Journal of the American Statistical Association*, 115, 1378–1392.

Bigot, J. (2006), "Landmark-Based Registration of Curves via the Continuous Wavelet Transform," *Journal of Computational and Graphical Statistics*, 15, 542–564.

Cheng, W., Dryden, I. L., and Huang, X. (2016), "Bayesian Registration of Functions and Curves," *Bayesian Analysis*, 11, 447–475.

Daniels, M. and Hogan, J. (2008), *Missing Data in Longitudinal Studies: Strategies for Bayesian Modeling and Sensitivity Analysis*, Chapman & Hall/CRC Monographs on Statistics & Applied Probability, CRC Press.

Earls, C. and Hooker, G. (2017), "Variational Bayes for Functional Data Registration, Smoothing, and Prediction," *Bayesian Analysis*, 12, 557–582.

Hopkins, J. T., Coglianese, M., Reese, S., and Seeley, M. K. (2013), "Alterations in Evertor/Invertor Muscle Activation and Cop Trajectory during a Forward Lunge in Participants with Functional Ankle Instability," *Clinical Research on Foot and Ankle*, 2, 122–127.

Hopkins, T. J., Son, J. S., Hyunsoo, K., Page, G. L., and Seeley, M. K. (2019), "Characterization of Multiple Movement Strategies in Participants with Chronic Ankle Instability," *Journal of Athletic Training*, 54, 698–707.

James, G. M. (2007), "Curve Alignment by Moments," *Annals of Applied Statistics*, 1, 480–501.

Kneip, A. and Gasser, T. (1992), "Statistical Tools to Analyze Data Representing a Sample of Curves," *The 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.

Koshino, Y., Ishida, T., Yamanaka, M., Ezawa, Y., Okunuki, T., Kobayashi, T., Samukawa, M., Saito, H., and Tohyama, H. (2016), "Kinematics and muscle activities of the lower limb during a side-cutting task in subjects with chronic ankle instability," *Knee Surgery, Sports Traumatology, Arthroscopy*, 24, 1071–1080.

Lang, S. and Brezger, A. (2004), "Bayesian P-Splines," *Journal of Computational and Graphical Statistics*, 13, 183–212.

Lin, L. and Dunson, D. B. (2014), "Bayesian Monotone Regression Using Gaussian Process Projection," *Biometrika*, 101, 303–317.

Lu, Y., Herbei, R., and Kurtek, S. (2017), "Bayesian Registration of Functions With a Gaussian Process Prior," *Journal of Computational and Graphical Statistics*, 26, 894–904.

Lucero, J. C., Munhall, K. G., Gracco, V. L., and Ramsay, J. O. (1997), "On the Registration of Time and the Patterning of Speech Movements," *Journal of Speech, Language, and Hearing Research*, 40, 1111–1117.

Rakêt, L. L., Sommer, S., and Markussen, B. (2014), "A Nonlinear Mixed-Effects Model for Simultaneous Smoothing and Registration of Functional Data," *Pattern Recognition Letters*, 38, 1 – 7.

Ramsay, J. O. and Dalzell, C. (1991), "Some Tools for Functional Data Analysis," *Journal of the Royal Statistical Society: Series B (Methodological)*, 53, 539–561.

Ramsay, J. O. and Li, X. (1998), "Curve Registration," *Journal of the Royal Statistical Society: Series B*, 60, 351–363.

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

Sakoe, H. and Chiba, S. (1978), "Dynamic Programming Algorithm Optimization for Spoken Word Recognition," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 26, 43–49.

Seeley, M. K., Son, S. J., Kim, H., and Hopkins, J. T. (2017), "Walking Mechanics for Patellofemoral Pain Subjects with Similar Self-reported Pain Levels Can Differ Based Upon Neuromuscular Activation," *Gait and Posture*, 53, 48–54.

Srivastava, A., Wu, W., Kurtek, S., Klassen, E., and Marron, J. S. (2011), "Registration of Functional Data Using Fisher-Rao Metric," *ArXiv e-prints*, 1103.3817.

Telesca, D. (2015), "Bayesian Analysis of Curves Shape Variation Through Registration and Regression," in *Nonparametric Bayesian Inference in Biostatistics*, eds. Mitra, R. and Müller, P., Cambridge: Springer International Publishing, pp. 287–310.

Telesca, D. and Inoue, L. Y. T. (2008), "Bayesian Hierarchical Curve Registration," *Journal of the American Statistical Association*, 103, 328–339.

Tucker, J. D. (2020), *fdasrvf: Elastic Functional Data Analysis*.

Tucker, J. D., Wu, W., and Srivastava, A. (2013), "Generative Models for Functional Data Using Phase and Amplitude Separation," *Computational Statistics and Data Analysis*, 61, 50–66.

Wang, J.-L., Chiou, J.-M., and Müller, H.-G. (2016), "Functional Data Analysis," *Annual Review of Statistics and Its Application*, 3, 257–295.

White, M. S., Burland, J. P., Davi, S. M., Lepley, A. S., and Lepley, L. K. (2019), "Hidden Asymmetries in ACLR Patients Who Pass Triple Hop Test Following ACLR," *ACL Research Retreat VIII*. Greensboro, North Carolina.