Elastic Shape Analysis of Movement Data

J.E. Borgert, Jan Hannig, J.D. Tucker, Liubov Arbeeva, Ashley N. Buck, Yvonne M. Golightly, Stephen P. Messier, Amanda E. Nelson & J.S. Marron

View supplementary material ↗

Accepted author version posted online: 20 Oct 2025.

Submit your article to this journal ↗

View related articles ↗

View Crossmark data ↗

Elastic Shape Analysis of Movement Data

J.E. Borgert[1,2,*], Jan Hannig[1], J.D. Tucker[3], Liubov Arbeeva[2], Ashley N. Buck[2,4,5], Yvonne M. Golightly[2,6,7], Stephen P. Messier[8], Amanda E. Nelson[2,7,9], J.S. Marron[1]

[1]Department of Statistics & Operations Research, University of North Carolina, Chapel Hill, NC

[2]Thurston Arthritis Research Center, University of North Carolina, Chapel Hill, NC

[3]Statistical Sciences, Sandia National Laboratories, Albuquerque, NM

[4]Human Movement Science Curriculum, University of North Carolina, Chapel Hill, NC

[5]Department of Exercise & Sport Science, University of North Carolina, Chapel Hill, NC

[6]College of Allied Health Professions, University of Nebraska Medical Center, Omaha, NE

[7]Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina, Chapel Hill, NC

[8]Department of Health & Exercise Science, Wake Forest University, Winston-Salem, NC

[9]Department of Medicine, University of North Carolina, Chapel Hill, NC

*elyseb@live.unc.edu

Abstract

Osteoarthritis (OA) is a highly prevalent degenerative joint disease, and the knee is the most commonly affected joint. Biomechanical factors, particularly forces exerted during walking, are often measured in modern studies of knee joint injury and OA, and understanding the relationship among biomechanics, clinical profiles, and OA has high clinical relevance. Biomechanical forces are typically represented as curves over time, but a standard practice in biomechanics research is to summarize these curves by a small number of discrete values (or *landmarks*). The objective of this work is to demonstrate the added value of analyzing full movement curves over conventional discrete summaries. We developed a shape-based representation of variation in full biomechanical curve data from the Intensive Diet and Exercise for Arthritis (IDEA) study (Messier et al., 2009, 2013), and demonstrated through nested model comparisons that our approach, compared to conventional discrete summaries, yields stronger associations with OA severity and OA-related clinical traits. Notably, our work is among the first to quantitatively evaluate the added value of analyzing full movement curves over conventional discrete summaries.

1 Introduction

Osteoarthritis (OA) is a prevalent degenerative joint disease characterized by cartilage loss, bone and soft tissue changes, joint pain, and diminished function. In the United States, OA affects at least 19% of adults aged 45 years and older, with the knee being the most commonly affected joint, accounting for more than 80% of the total burden of the disease (Dillon et al., 2006; Jordan et al., 2007; Lawrence et al., 2008; Vos et al., 2012). Knee OA is characterized by both the severity of radiographically-assessed damage and clinical symptoms, such as knee pain and function. Previous research, such as in Zhang and Jordan (2010), indicated that risk factors and disease progression may vary by clinical phenotype. Additionally, important work like Felson (2013) and Guilak (2011) identified biomechanical factors in the etiology and pathogenesis of knee OA.

Biomechanical variables, particularly forces exerted during walking, are often measured in modern studies of knee joint injury and OA, and understanding the relationship among biomechanics, clinical profiles, and OA has high clinical relevance. Biomechanical forces are typically represented as curves over time, but a standard practice in biomechanics research is to summarize these curves by a small number of discrete values. Such discrete summaries are called *landmarks* in the shape statistics terminology of Dryden and Mardia (2016). Analyses based on conventional discrete summaries, such as those by Sims et al. (2009) and Astephen et al. (2008), have identified differences between groups (e.g., sex differences) and discovered variations in gait patterns associated with knee OA-related outcomes. Recent work by Buck et al. (2024) evaluated the ability of various clinical traits and conventional discrete summaries of gait forces to predict early symptomatic knee OA. While simplifying the statistical methods required for analysis, relying on conventional discrete summaries risks overlooking information encoded in the complete range and patterns of movement data. Research by Muniz et al. (2006), Davis et al. (2019), Costello et al. (2021), Bjornsen et al. (2024), and others indicate the value of analyzing full movement curves. However, these studies are limited and did not formally compare analyses based on conventional discrete summaries. For harmonic analyses of full movement curves, see Trentadue and Schmitt (2024) and references therein.

The primary goal of this work is to demonstrate the added value of analyzing full movement curves over conventional discrete summaries. Our analysis of full movement curves follows an Object Oriented Data Analysis (OODA) approach. OODA, described in Marron and Dryden (2021), is a framework for analyzing complex data that emphasizes the careful selection of data objects for a given scientific question and the utilization of methods intrinsic to the data object space. This approach facilitates the consideration of full movement data curves as complex data objects in high-dimensional space. Furthermore, the richer information within these curves allows for potentially many different choices of both the *data object* of interest and the appropriate methodology for analysis.

We developed a shape-based representation of variation in full biomechanical curves using data from the Intensive Diet and Exercise for Arthritis (IDEA) study (Messier et al., 2009, 2013), and demonstrated through easily interpretable nested model comparisons that our approach, compared to conventional discrete summaries, yields stronger associations with OA severity and OA-related clinical traits. Notably, our work is among the first to

quantitatively evaluate the added value of analyzing full movement curves over conventional discrete summaries in OA research.

2 Curves as Data Objects

During gait data collection in the IDEA study, participants wore laboratory-provided cushioned shoes and walked at their preferred speed on a 22.5m walkway. Kinetic data, including Ground Reaction Force (GRF), were collected using an Advanced Medical Technologies, Inc. model OR-6-5-1 force plate (480 Hz) and filtered using a 4th order low-pass Butterworth filter with a cutoff frequency of 6 Hz. The GRF consists of three components: vertical GRF (vGRF), representing the force exerted downwards; anterior-posterior GRF (apGRF), the propulsive or braking force in the direction of walking; and medial-lateral GRF (mlGRF) in the third orthogonal direction. When representing GRF as a function of time, the horizontal axis reflects the percentage of *stance* (time interval of foot contact with the ground) over a gait cycle, and the vertical axis corresponds to body weight-normalized force values (measured in Newtons).

A walking trial was recorded as a successful observation when the participant's entire foot maintained contact with the force plate throughout the stance phase, and participants maintained their preferred walking speed within $\pm 3.5\%$. Walking speed was defined as the mean speed at which participants walked a 10m walkway at a self-selected pace over six practice trials. A photocell system registered speed and provided participants with real-time visual feedback for maintaining their walking speed during formal trials.

We studied the part of the IDEA data that contained measurements of each GRF component taken at a constant sampling rate over the duration of each step. For each IDEA participant, three trials per limb were collected. For some participants, one or more trials were missing from the data. In those cases, we used as many trials as were given in the data and did not impute. We considered the collection of all trials of both limbs from all participants.

The complete set of curves for each GRF component contained 2,686 curves from 454 participants. The top left panel of Figure 1 shows the collection of raw data vGRF curves colored by walking speed. There, the rainbow descends from fastest walking speed (red) to slowest (purple). Notice that many of those curves have consecutive starting and trailing zeros, which are outside the stance phase (i.e., after the foot has left contact with the force plate, but force data were still being collected) and hence do not correspond to a meaningful part of the measurements. To account for these spurious measurements, we take the beginning of the force curve as the zero value immediately preceding the first nonzero value, and similarly, the end of the curve as the zero value immediately following the last nonzero value. Using the relevant segment of each curve, we re-scaled the horizontal axis to the unit interval $[0,1]$ in order to establish a common time axis across all curves. Additionally, we applied linear interpolation to the force values, aligning them to an evenly spaced grid. Time normalization to the stance phase $(0\text{--}100\%)$ is a standard approach in biomechanics for analyzing GRF and other kinetic variables. The most commonly used method for time normalization is *linear length normalization*, which is a rescaling of the stance phases to a standard interval (Helwig et al., 2011), as was applied in our analysis. While this approach removes explicit information about stance duration, it retains timing differences in key joint-loading events, such as heel-strike and toe-off, which impact cartilage stress and are therefore relevant to understanding OA. In this work, we chose to focus on a careful analysis of shape,

and leave analyzing stance duration as a direction for future work. However, stance duration is strongly correlated with walking speed and distance (Hebenstreit et al., 2015), which were retained as clinical variables in our analysis.

The top right panel of Figure 1 presents the collection of vGRF curves shown in the first panel, following this processing. Those curves are also color-coded according to the participant's walking speed. Similar results were achieved using the same processing steps for the apGRF and mlGRF curves, and are shown in the bottom left and bottom right panels of Figure 1, respectively. For a more detailed view of GRF curve variations across walking speeds, additional plots showing subsets of the curves grouped by deciles of walking speed are provided in Section 1 of the supplementary material.

2.1 Conventional discrete summaries

Most relevant studies of GRF data, such as those in Messier et al. (1992), Hunt et al. (2006), Zeni Jr and Higginson (2009), and Wiik et al. (2017), rely on discrete summaries (also called landmarks) rather than analyzing the full curves. These discrete summaries typically correspond to critical subphases of stance during gait, such as the first and second vGRF peaks (heel-strike and toe-off) and positive and negative peaks of the apGRF curve (braking and propulsion). In contrast, mlGRF curves are highly variable and therefore more challenging to interpret and summarize (Costello et al., 2021). Consequently, GRF studies primarily analyze only the vGRF and apGRF directions.

Traditionally, the first vGRF peak has been of greater interest, as heel-strike is a significant factor in joint compressive loads and has been consistently linked to knee OA onset and progression. However, recent work by Buck et al. (2024) (and references therein) indicated that the valley (reflecting mid-stance) and the second peak of the vGRF curve are better predictors of knee OA-related symptoms.

In this analysis, we focus on three conventional discrete summaries: (1) the first vGRF peak, (2) the apGRF braking peak, and (3) the apGRF propulsion peak. Throughout the remainder of this paper, we refer to the 3-dimensional vector containing these landmark values as the *conventional discrete summaries*. While other landmark choices could have been analyzed, as discussed in the preceding paragraphs, these three were selected because they are the most commonly studied GRF features in biomechanical research on knee OA.

The first vGRF peak is defined in the references aforementioned as the maximum value within the first 50% (0-50%, heel-strike to midstance) of the stance phase (0-50%) (0-0.5 on the horizontal axis in Figure 1). Similarly, the apGRF braking peak is defined as the minimum value over the first 50% of the stance phase, and the apGRF propulsion peak is defined as the maximum value over the second 50% ($50-100\%$, mid-stance to toe-off) of the stance phase (0.5-1.0 on the horizontal axis in Figure 1).

3 Shape-based Functional Data Analysis

3.1 Elastic shape analysis

The goal of our analysis was to characterize patterns of variation in gait using information in the full force curves. Many datasets of curve data have variation that appears to be either

vertical or horizontal in nature. This variation is termed *amplitude* and *phase* variation, respectively. In this context, horizontal variation is viewed as a potentially important aspect of gait, while in other contexts it may represent temporal misalignment. The curves shown in Figure 1 exhibit interesting variation of both types. In particular, there is clear phase variation in the timing of the vGRF and mlGRF peaks, as well as the shift from posterior to anterior force in the apGRF curves.

Elastic warping of the time axis can provide aligned curves that better capture amplitude variation, as studied in Section 2.1 of Marron and Dryden (2021). This is important because poor alignment of curves due to phase variation can impact statistical methodology, potentially obscuring important geometric structure. As noted by Helwig et al. (2011), temporal alignment is a critical consideration in gait analysis, with various approaches applied in gait studies. That paper highlights that while linear length normalization, which we applied as a pre-processing step, removes differences in stance duration, it does not account for phase variation in the timing of key events (e.g., peaks and valleys). Other methods discussed by Helwig et al. (2011) require extensive manual tuning and may introduce distortions in curve shape. The results discussed in that paper support the need for an alignment method that captures phase variation while preserving curve shape and minimizes manual tuning.

Elastic warping involves a transformation of the time axis, which is described by a curve that can be usefully thought of as a stretching and compression of the horizontal axis. The functions are aligned by finding the *Karcher mean* (Tucker et al., 2013) which produces aligned functions and warping functions and will be defined later on. A warping function $\gamma(x):[0,1] \rightarrow [0,1]$ is strictly increasing, invertible, and *diffeomorphic*, meaning the function and its inverse are smooth. The collection of such warping functions serve as *phase* data objects.

Aligning points across functions is often referred to as *registration*. Many conventional Functional Data Analysis (FDA) techniques rely on the $\mathbb{L}^2$ norm, which simplifies computations into point-wise evaluations. While point-wise computations involve *vertical registration*, other methods focus on the *shape* of functions. L^2-based methods present well-known challenges, as detailed in Marron et al. (2015). *Elastic shape analysis*, as proposed in Srivastava et al. (2011); Tucker et al. (2013), uses the warp-invariant Fisher-Rao metric to overcome the limitations of conventional $\mathbb{L}^2$-based alignment techniques. This framework dates back to such seminal work as Younes (1998) and is the first to enable fully automatic (meaning that no manual tuning is needed) shape-based registration. For each curve, the elastic shape analysis method computes the warping function needed to align its peaks to a template mean curve, known as the Karcher mean. The Karcher mean is the curve that lies in the "center" of all the warped curves, meaning it minimizes the total distance (under the Fisher-Rao metric) between itself and all the aligned curves.

The key idea of this method is to define an equivalence relation between curves. Two curves $f_1(x)$ and $f_2(x)$ are called equivalent, $f_1 \sim f_2$, if there exists a warping function γ such that $f_1(\gamma(x)) = f_1 \circ \gamma(x) = f_2(x)$. Then, the set of all warps of a function f, given by

$$[f] = \{f \circ \gamma : \gamma \in \Gamma\},$$

is an equivalence class and defines the amplitude (called *shape* in Srivastava et al. (2011); Wu et al. (2024)) of *f*.

The Fisher-Rao metric defines a proper distance on the set of such equivalence classes. A natural framework for carrying out the computations required for curve alignment is achieved through a *Square Root Velocity Function* (SRVF) representation, which transforms the Fisher-Rao metric into the standard $\mathbb{L}^2$ metric. The Karcher mean equivalence class is defined using this distance, along with the warping functions needed to align individual functions to the Karcher mean template. The SRVF of a function $f \in \mathcal{F}$ is given by $q(t) = \text{sign}(\dot{f}(t))\sqrt{|\dot{f}(t)|}$. For any time warping of *f* by $\gamma \in \Gamma$, the SRVF of the warped function is given by $(q \circ \gamma)\sqrt{\dot{\gamma}}$, which we will denote by $q \star \gamma$ for convenience. Then, the warping functions needed to align a collection of curves $f_1, \ldots, f_n$ with corresponding SRVF representations, $q_1, \ldots, q_n$, to the Karcher mean μ_q^* are computed by solving the optimization problem:

$$\mu_q^* = \operatorname*{argmin}_{q \in \mathbb{L}^2} \sum_{i=1}^{n} \left(\inf_{\gamma \in \Gamma} \| q - q_i \star \gamma_i \|^2 \right).$$

The output of this optimization problem is the Karcher mean μ_q^* and the set of optimal warping functions $\{\gamma_i^*\}$. The optimization the problem is solved using *dynamic programming*: the time domain *[0, 1]* is discretized into *N* equally spaced bins, and each warping function γ_i is approximated by a piecewise linear mapping from $(0,0)$ to $(1,1)$. This discretization corresponds to a path through an $N \times N$ grid, where each step is required to have a positive slope to ensure invertibility.

In our analysis, we found substantial benefit in penalizing the amount of elasticity in the alignment of the curve. This is achieved by modifying the optimization problem to include a penalty term on the roughness of the γ_i as follows:

$$\mu_q^* = \operatorname*{argmin}_{q \in \mathbb{L}^2} \sum_{i=1}^{n} \left(\inf_{\gamma \in \Gamma} \| q - q_i \star \gamma_i \|^2 + \lambda \mathcal{R}(\gamma_i) \right)$$

The penalty $\mathcal{R}(\gamma_i)$, controlled by the constant $\lambda > 0$, imposes a constraint on γ_i. In our approach, we restrict the second derivative of γ_i, so that $\mathcal{R}(\gamma_i) = \int_0^1 \ddot{\gamma}_i(t)dt$. This places a restriction on the smoothness in γ_i and has the effect of keeping γ_i closer to the identity warp $\gamma_i(t) = t$, thus regulating the level of elasticity in the alignment. The case $\lambda = 0$ is referred to as *fully elastic alignment*, while $\lambda = \infty$ is the non-elastic case.

The penalty is computed along the grid path and incorporated into the total cost, which is then minimized using dynamic programming in the same manner as without the penalty. This approach is similar to the method described by Wu and Srivastava (2011), where further details on the algorithm are provided.

An interesting alternative to adding a smoothness penalty is a metric learning approach that considers the broader 1-parameter family of elastic metrics, which extend SRVFs as described in Bauer et al. (2024). This family allows for flexible control over warping by varying the transformation in the metric, thereby implicitly enforcing smoothness without requiring a penalty term on the γ_i.

For an intuitive overview of the elastic shape analysis procedure and a more detailed derivation of the Karcher mean, see Section 9.1.3 of Marron and Dryden (2021). A thorough comparison of functional data analysis with and without phase-amplitude separation is provided in Chapters 2.1, 5.4, and 9 of Marron and Dryden (2021), demonstrating its importance for data exhibiting both types of variation. Given these established results, the observed phase variation in the GRF data motivated our application of elastic shape analysis, from which we obtained a decomposition into amplitude and phase data objects. The following subsection details our implementation and selection of the penalty parameter λ.

3.1.1 Implementation and penalty parameter selection

The elastic shape analysis procedure was implemented via the FDASRSF Python package (Tucker et al., 2013). One possible approach to registering the GRF curve data is to apply the elastic shape alignment to each GRF component (vGRF, apGRF, mlGRF) separately. An analysis of amplitude and phase using this component-wise registration in the IDEA study data is detailed in Section 4.4 of Xiang (2023). However, that approach is less meaningful kinetically, as each component represents one direction of the same measured force. Instead, we adopted a more intuitive approach by treating the three components as a single multidimensional curve and applying elastic shape alignment to obtain a common set of warping functions. This approach allows us to focus on the phase aspects shared by all three components. Note that the Fisher-Rao mathematics extend to multi-dimensional functions (Srivastava and Klassen, 2016), where the SRVF for a vector-valued function $f(t)$ becomes:

$$q(t) = \frac{\dot{f}(t)}{\sqrt{\| \dot{f}(t) \|}}.$$

A subset of GRF curves exhibited atypical vGRF or apGRF components, which posed challenges to aligning these curves with the rest of the data. Those GRFs were atypical in the sense that the vertical component lacked the two-peak structure expected of normal gait and appeared closer to unimodal, and/or the anterior-posterior component was close to zero and relatively flat. Examples of these atypical cases are highlighted in Figure 2, with representative atypical cases colored by walking speed and other curves in gray. Each panel in this row corresponds to one component of the original (unaligned) GRF curves.

To address the lack of a common underlying structure, we registered the full dataset using the penalized elastic shape analysis procedure. By adjusting the elasticity parameter λ, we aligned the full set of GRF curves without distorting the shape of the atypical curves. To determine an appropriate λ, we computed warping functions iteratively over a grid of candidate values and visually examined the alignment results, with particular attention to atypical cases. We refined the grid in regions where λ values produced reasonable results, selecting values that avoided excessive smoothing of features in the atypical curves while still capturing meaningful phase variation. Over-alignment led to sharp corners forming a

staircase-like pattern in the warping functions, indicating drastic stretching and compression of curves with differing underlying structures, such as unimodal curves.

Based on this evaluation, we selected $\lambda = 2$ for mitigating the staircase effect without totally sacrificing alignment. To illustrate the alignment trade-off, Figure 3 compares results for three λ values: the fully elastic case $\lambda = 0$ in the top row, our selected $\lambda = 2$ in the middle row, and a less elastic case $\lambda = 4$ in the bottom row. Atypical cases are colored by speed, while other curves are shown in gray, as in Figure 2. The leftmost panels display the warping functions for each λ, where $\lambda = 0$ results in sharp staircase-like patterns indicative of over-alignment, while $\lambda = 4$ yields warping functions tightly clustered around the identity, suggesting insufficient alignment. The second through fourth columns show the aligned curves. Results for additional values of λ can be found in Figure 11 of the supplementary material.

While our approach treated alignment as agnostic to a specific performance criterion, a more formal selection method could be constructed depending on the analytical objective. For example, if the interest is in evaluating alignment quality in reference to a predictive or inferential task, λ could be optimized via cross-validation to maximize a relevant performance metric. However, in finite samples, cross-validation is prone to noisy selection due to its slow convergence to optimal results (see Hall and Marron (1987) in the context of bandwidth selection for kernel density estimation). Alternatively, if the goal is to estimate an underlying common signal, Kim et al. (2023) provides a scale-space approach for estimating both the shape of the unknown signal and the signal itself. Defining a notion of optimality depends on the specific context of the analysis, and the choice of λ for different analytical objectives is an open question for further research.

Next, we obtained amplitude objects by applying the (common) set of warping functions to each set of original GRF curves, which provides an intuitive representation of amplitude in each component. The amplitude objects obtained for different values of λ are shown in the second through fourth columns of Figure 3.

3.2 Modes of variation

In OODA terminology, a collection of members of the object space that summarize one component of variation and is in some sense one-dimensional is called a *mode of variation*. For example, in the vector matrix case, a mode of variation is a rank-one matrix. We can obtain modes of variation through *Principal Component Analysis* (PCA), where each object is considered as a point in high dimensional space (column vector). For an introduction to PCA, see Jolliffe (2002). Amplitude modes of variation (Tucker et al., 2013) were obtained for each direction of the ground reaction force (vGRF, apGRF, mlGRF) computing PCA on the set of 2,686 Fisher-Rao aligned curves, each corresponding to an individual gait observation. The sets of input curves are shown in the second, third, and fourth columns of the middle row of Figure 3.

Figure 4 shows the modes of variation of the amplitude objects of the vGRF, where the curves are colored according to walking speed. The first mode of variation, shown in the first panel of the second row of that figure, is associated with walking speed and reflects the contrast in peak heights and valley depths. Faster walkers (indicated in red in the rainbow color scheme) generally exhibit higher peaks and lower valleys, while slower walkers (purple

in the rainbow color scheme) have lower peaks and a shallower valley. The middle column of that figure displays the largest (dashed curve) and smallest (dotted curve) PC projections added back to the mean curve, which is shown as a solid black curve in each of the middle panels. The middle panel of the first mode shows that the slowest walkers (dotted curve) exhibit a vertical amplitude force that appears unimodal and does not exceed body weight (1 on the vertical axis), indicating that these walkers do not fully transfer their weight to the striking limb. This type of gait can be thought of as "shuffling." The second and third modes of variation are about the second and first peak, respectively. The middle panel in the third row shows the largest and smallest PC2 projections added back to the mean curve, distinguished with a dashed (largest) and dotted (smallest) line type. These extremes show that variation in this mode is mostly in the height of the second peak. In the panel below, the extremes of the third mode indicate phase variation in the first peak that is unique to the vertical component. The fourth mode of variation reflects the overall magnitude, particularly in the mid-stance phase. The second, third, and fourth PC projection extreme curves all suggest that some curves have a small third bump before the first peak. In gait analysis, this pattern is known as the *heel-strike transient* (HST), a rapid and transient rise in the vGRF immediately after ground contact. As discussed in Blackburn et al. (2016) and references therein, the presence and characterization (e.g. magnitude) of HST can indicate impulsive loading, which influences cartilage degradation and symptoms of OA. However, Blackburn et al. (2016) also noted that methods for identifying HST can be unreliable. The amplitude modes of variation we identified offer a potentially viable method for reliably identifying and understanding the HST.

Similar plots of the amplitude modes of variation of the components of apGRF and mlGRF are provided in Section 2 in the supplementary material.

Extracting phase modes of variation requires more careful consideration. Recall from Section 3.1 that the warping functions all have corresponding SRVFs that lie on the surface of a high-dimensional sphere in the function space. Thus, using PCA to identify phase modes of variation is essentially an approximation in the tangent space centered on the Karcher mean. In the case of warping functions, these SRVFs must also lie on the positive orthant Wu et al. (2024). It is demonstrated in Yu et al. (2017) that in cases of high variation, this tangent plane PCA may yield a distorted analysis, resulting in modes of variation that leave the positive orthant and consequently produce invalid warping functions. This phenomenon was also observed in our dataset. In such scenarios, a better decomposition of the variation can be achieved using the functional PCA methodology proposed by Yu et al. (2017), which is based on an improved PCA analogue for spheres known as Principal Nested Spheres (PNS) proposed by Jung et al. (2012). The PNS decomposition sequentially provides the best k-dimensional approximation U^k of the data for all $k = d-1, d-2, \ldots, 0$ such that

$$S^d \supset U_{d-1} \supset \ldots \supset U_1 \supset U_0.$$

For each k, the sphere U^k, called the k-dimensional *principal nested sphere*, is a submanifold of the higher dimensional principal nested spheres. The algorithm to find sample principal nested spheres is determined by iteratively minimizing an objective function to find the best-fitting subsphere, projecting the data to the lower dimensional sphere, and mapping to the original space through a relevant transformation. The signed residuals, defined as the signed length of the minimal geodesic joining the (projected and transformed) data points to the

subsphere, serve as analogs of principal component scores. Chapter 8 of Marron and Dryden (2021) provides further review of PNS and other geodesic-based methods.

We applied the PNS-based functional PCA methodology to the set of (common) warping functions to obtain phase modes of variation. We found that the great sphere decomposition from PNS yielded the most interpretable phase modes of variation because of weak interpretability of small sphere variation. Figure 5 depicts an intuitively useful notion of phase variation represented by warpings of the Karcher mean of the vGRF curves. The warping functions used to create these visualizations were generated by taking the inverse of the phase PNS projections added to the 45-degree line (identity warp). In each panel of the figure, the curves are colored based on the PNS scores for the corresponding mode, with cyan indicating the lowest scores and magenta indicating the highest. It is important to note that the curves are plotted in the order of the corresponding score, as over-plotting is an issue. The first mode (first panel) shows an overall shift in timing, with most apparent differences in the timing of the first peak (maximum heel-strike force) and valley. The second mode in the next panel appears to explain variability in the closeness of the peaks: the cyan curves are the curves with peaks closer together and the magenta curves have peaks farther apart. The third mode represents an overall phase shift (left vs. right) and seems to suggest that curves having a small third bump before the first peak correspond with earlier timing (cyan curves), especially an earlier second peak. The fourth mode appears to highlight variability in the timing of the second peak, independent of the rest of the curve.

4 Comparison to Conventional Discrete Summaries

We investigated the added value of analyzing patterns across the entire movement curve, rather than relying on conventional discrete summaries of GRF curves. For this purpose, we compared how strongly full-curve modes versus conventional discrete summaries were associated with OA-related clinical traits using a nested regression framework, detailed in Section 4.2. While some of the traits we considered are clinically meaningful to predict from gait biomechanics, the aim of this analysis was to quantify and compare the strength of associations between OA-related traits and gait features derived from full-curve analysis versus conventional discrete summaries. In this section, we outline the sets of gait features that served as independent variables in our models, and in the following section, we introduce the OA-related clinical traits.

To create independent variables derived from our full-curve analysis, we combined scores for 16 distinct modes of amplitude variation from three types of curve data objects (vGRFs, apGRFs, and mlGRFs) and phase variation. Each mode is represented by either a set of amplitude PC scores or PNS phase scores. Below are listed the 16 sets of scores that together formed our full-curve independent variables:

• PC1-PC4 scores of the vGRFs amplitude data objects (studied in Figure 4);

• PC1-PC4 scores of the apGRFs amplitude data objects (studied in Figure 12 in the supplementary material);

• PC1-PC4 scores of the mlGRFs amplitude data objects (studied in Figure 13 in the supplementary material);

• Great sphere PNS1-PNS4 scores of the common phase data object (studied in Figure 5).

We developed a set of independent variables based on conventional discrete summaries of GRF curves found in the literature, including the first peak of the vGRF curve (maximum over $0-50\%$ of the stance), the minimum value of the apGRF curve (minimum over $0-100\%$ of the stance) and the maximum value of the apGRF curve (maximum over $0-100\%$ of stance).

4.1 OA Clinical Traits

The IDEA study defined several OA disease outcomes and symptoms of interest (Messier et al., 2009, 2013). These included mechanistic outcomes: knee joint compressive force and inflammatory biomarkers (interleukin-6 [IL-6] and C-reactive protein [CRP]); and clinical outcomes: self-reported pain and function, mobility, and health-related quality of life. Increased knee joint compressive force is known to contribute to cartilage stress and degeneration and has been associated with patterns in gait biomechanics (D'Lima et al., 2012). Elevated IL-6 and CRP levels are linked to chronic inflammation and have been associated with knee OA (Messier et al., 2009). Pain and function were measured using the Western Ontario and McMaster Universities Arthritis Index (WOMAC) (Alexandersen et al., 2014). Mobility was assessed using walking speed and distance walked in a 6-minute trial, while health-related quality of life was evaluated using the SF-36 Physical and Mental Component Scales. These IDEA study outcomes were included as dependent variables in our regression analyses.

Although not specified as outcomes in the IDEA study, we analyzed additional biomedical measures that have been examined in other OA research. Notably, radiographic OA severity is evaluated using joint space width (JSW) and Kellgren–Lawrence grade (KLG), and both are key metrics for diagnosing and monitoring disease progression. Prior studies have reported associations between gait features and variability in these structural measures (e.g. Kwon et al. (2019); Jansen et al. (2024)), making them particularly relevant to this analysis. OA is also associated with elevated fall risk, and existing studies have examined fall risk as an outcome in patients with knee OA (for example, Rosadi et al. (2022)). Furthermore, anthropometric characteristics (e.g., body weight) and sociodemographic traits (e.g., age, sex, race) are important factors in OA risk, symptoms, and treatment (see, for example, Sims et al. (2009); Allen et al. (2022); Chang et al. (2024)). These additional measures, together with the IDEA study outcomes, are summarized in Table 1 of the supplementary material and are collectively referred to as *clinical traits* throughout the paper. For limb-level measurements (such as joint compressive force, KLG, and JSW), we used the patient-level clinical aggregates of these measures reported in the IDEA study for our analysis (Messier et al., 2009, 2013).

Since the IDEA study was an 18-month clinical trial of interventions in patients with advanced knee OA, traits were collected at different intervals throughout the study. For the purpose of this analysis, focusing on baseline values provides the most consistent basis for comparison and may better reflect OA variability in the dataset. For example, although change in JSW is a key marker of progression, prior studies suggest that knees at similar OA stages progress at comparable rates (Benichou et al., 2010), making baseline JSW a more meaningful indicator of disease severity for the IDEA participants. Additionally, using baseline values avoids the need to account for treatment effects, which are not of interest in this analysis.

Figure 15 in the supplementary material shows a heat map of missing values in the baseline data, with blue lines indicating missing entries for each trait; note that JSW has a relatively high rate of missing data. Baseline JSW measurements were missing for 126 of the 454 IDEA participants. Although the reasons for these missing values were not reported in the study, comparison of the other clinical traits between participants with and without baseline JSW showed no apparent sampling bias. This left 328 participants with JSW data for our analysis. Due to the large proportion of missing values, analysis involving JSW was restricted to the complete subset, with gait variables subset accordingly. Missing values for other traits, which were relatively few, were imputed using the mean of each trait (or the mean rounded to the nearest integer for integer-valued data). Importantly, there were no missing values for traits that may be particularly sensitive to imputation, such as categorical traits like sex and race, and low-resolution ordinal measures like KLG. The treatment of different types of dependent variables (count, binary, ordinal, continuous) for modeling purposes is discussed in detail in the following section.

4.2 Nested Model Comparison

We assessed the added value of full-curve gait analysis over conventional discrete summaries by comparing nested regression models. Specifically, we evaluated whether full-curve models provided additional explanatory value beyond conventional discrete summaries using a bootstrap-based likelihood ratio test (LRT) approach. Since we treat curves as data objects, all 2,686 gait curves were used as individual observations, except when modeling JSW, which was analyzed on the subset described earlier. Each trait was treated as a dependent variable and a full model fit on the combined set of gait features (full-curve modes plus conventional discrete summaries), while reduced models were fit using either the full-curve modes alone or the conventional discrete summaries alone.

The choice of regression model depended on the type of dependent variable: count, binary, ordinal, or continuous. Number of falls was the only count-valued trait and was modeled using Poisson regression. Sex and race were the only binary-valued traits and were fit using logistic regression. Ordinal logistic regression was used for fall-related traits, WOMAC pain and function, health-related quality of life measures, and KLG. KLG is a radiographic score of OA severity, while the other ordinal traits represent scores from Likert-type scales. All remaining traits were continuous and fit using linear regression. As detailed in the previous section, missing trait values (except for JSW) were imputed using the trait mean. For integer-valued traits, the mean was rounded to the nearest integer, whereas for continuous traits it was used directly.

To compare nested models, we computed likelihood ratio tests for two cases: (1) a full model containing both full-curve modes and conventional discrete summaries versus a reduced model with only the full-curve modes, and (2) the same full model versus a reduced model with only the conventional discrete summaries. The null hypothesis in each case is that the reduced model explains the data as well as the full model, while the alternative is that the additional independent variables in the full model significantly improve model fit. Failing to reject the null when the reduced model contains the full-curve modes suggests that discrete summaries add no explanatory value beyond the full-curve modes. Conversely, rejecting the null when the reduced model contains the discrete summaries is evidence that full-curve modes provide additional explanatory value not captured by discrete summaries. The same logic applies when the roles of the two variable sets are reversed. Note that for linear regression models, the likelihood ratio test is equivalent to the nested-model F-test.

Typically, a LRT statistic is compared to a chi-squared reference distribution, but this relies on an asymptotic result which assumes independent observations. This assumption is violated in our dataset due to a clear dependence between the multiple gait curves per patient. To account for this, we implemented a bootstrap procedure in which patients were resampled with replacement, and all gait curves associated with a selected patient were included in the resampled dataset. For each of 1,000 bootstrap resamples, the regression models were on the selected sample and the corresponding LRT statistic was computed. This resampling approach reflects the natural dependence structure in the data and generates an appropriate reference distribution for inference. All other bootstrap analyses in this paper also used 1,000 replications. The significance of the observed LRT statistic from the original data was then assessed by computing the proportion of bootstrap LRT statistics greater than or equal to the observed LRT statistic. When no bootstrap replications were larger than the original statistic, we defined the p-value to be $\dfrac{1}{2*1000} = 0.0005$.

Figure 6 presents a scatterplot of p-values from the LRTs of nested models, displayed on a logarithmic scale. The horizontal axis shows p-values for testing full-curve reduced models, while the vertical axis corresponds to p-values for conventional discrete summary reduced models. For readability, the axes are labeled on the original scale.

The green plus signs highlight traits for which the full-curve reduced model was not rejected, while the conventional discrete summary reduced model was rejected, meaning conventional discrete summaries provided no additional explanatory value. The corresponding traits include joint compressive force; both measures of radiographic OA severity (JSW and KLG); all of the anthropometric and mobility measures; and sex and race. Balance confidence, one of the fall-related traits, was also in this group, though close to the 0.05 threshold.

The gray circles represent cases in which neither reduced model was rejected, which account for roughly half of the traits. These include both inflammatory biomarkers; self-reported pain and function; health-related quality of life measures; all but one fall-related trait; and age. Except for the biomarkers and age, these traits are Likert-type scale scores, which are highly subjective and generally difficult to model. Although inflammatory biomarkers and age are important in OA, there is no known direct relationship with gait biomechanics. Therefore, it is unsurprising that more complex gait variables do not provide significant explanatory value beyond simple discrete summaries for the gray-circle traits.

No points fall to the left of the vertical red dashed line, meaning there were no instances in which the full-curve reduced model was rejected while the conventional discrete summary reduced model was not. Although sex (Male) is near the 0.05 threshold for the full-curve reduced model, it is well below the threshold for the conventional discrete summary reduced model. Overall, these results show that a model based solely on full-curve modes is a consistently suitable approach.

5 Conclusion

This paper quantitatively demonstrates the extent to which complete GRF curves, compared with conventional discrete summaries, capture information relevant to disease severity and clinical profiles of OA, demonstrating the added value of full-curve analysis. We apply a straightforward nested model comparison to highlight this difference. Furthermore, our

shape-based approach illustrates an intuitive representation of full movement curves that is applicable in broader analyses and reveals insightful modes of variation. To our knowledge, this work is among the first to show that analysis of full movement curves yields stronger associations with OA outcomes and OA-related clinical traits than conventional discrete summaries.

SUPPLEMENTARY MATERIAL

Supplement to Elastic Shape Analysis of Movement Data: Additional figures and tables, with accompanying brief discussions. (pdf)

Funding

Conflict of Interest Statement

The authors report there are no competing interests to declare.

Statement on Generative AI

ChatGPT 4.0 was used for minor coding assistance.

References

Alexandersen, P., Byrjalsen, I., Bay-Jensen, A.-C., Siebur, A.-S., Andersen, J., Riis, B., Christiansen, C. and Karsdal, M. (2014), 'Association between joint space width, Kellgren-Lawrence score, pain and progression in osteoarthritis subjects from two phase III studies–a clinical study reference database', *Osteoarthritis and Cartilage* **22**, S197.

Allen, K. D., Thoma, L. and Golightly, Y. (2022), 'Epidemiology of osteoarthritis', *Osteoarthritis and cartilage* **30**(2), 184–195.

Astephen, J. L., Deluzio, K. J., Caldwell, G. E. and Dunbar, M. J. (2008), 'Biomechanical changes at the hip, knee, and ankle joints during gait are associated with knee osteoarthritis severity', *Journal of orthopaedic research* **26**(3), 332–341.

Bauer, M., Charon, N., Klassen, E., Kurtek, S., Needham, T. and Pierron, T. (2024), 'Elastic metrics on spaces of euclidean curves: Theory and algorithms', *Journal of Nonlinear Science* **34**(3), 56.

Benichou, O., Hunter, D., Nelson, D., Guermazi, A., Kwoh, K., Myers, S. and Duryea, J. (2010), 'One-year change in radiographic joint space width in patients with unilateral joint space narrowing: data from the Osteoarthritis Initiative', *Arthritis care & research* **62**(7), 924–931.

Bjornsen, E., Berkoff, D., Blackburn, J. T., Davis-Wilson, H., Evans-Pickett, A., Franz, J. R., Harkey, M. S., Horton, W. Z., Lisee, C., Luc-Harkey, B. et al. (2024), 'Sustained limb-level loading: A ground reaction force phenotype common to individuals at high risk for and those with knee osteoarthritis', *Arthritis & Rheumatology* .

Blackburn, J. T., Pietrosimone, B. G., Harkey, M. S., Luc, B. A. and Pamukoff, D. N. (2016), 'Comparison of three methods for identifying the heelstrike transient during walking gait', *Medical Engineering & Physics* **38**(6), 581–585.

Buck, A. N., Lisee, C. M., Bjornsen, E. S., Schwartz, T. A., Spang, J. T., Franz, J. R., Blackburn, J. T. and Pietrosimone, B. G. (2024), 'Biomechanical threshold values for identifying clinically significant knee-related symptoms six months following anterior cruciate ligament reconstruction', *Journal of Athletic Training* .

Chang, E., Charlton, J. M., Krowchuk, N., Hatfield, G. and Hunt, M. (2024), 'Sex and disease severity differences in gait biomechanics in older adults with and without knee osteoarthritis', *Osteoarthritis and Cartilage* **32**, S157.

Costello, K. E., Felson, D. T., Neogi, T., Segal, N. A., Lewis, C. E., Gross, K. D., Nevitt, M. C., Lewis, C. L. and Kumar, D. (2021), 'Ground reaction force patterns in knees with and without radiographic osteoarthritis and pain: descriptive analyses of a large cohort (the multicenter osteoarthritis study)', *Osteoarthritis and cartilage* **29**(8), 1138–1146.

Davis, H. C., Luc-Harkey, B. A., Seeley, M. K., Blackburn, J. T. and Pietrosimone, B. (2019), 'Sagittal plane walking biomechanics in individuals with knee osteoarthritis after quadriceps strengthening', *Osteoarthritis and cartilage* **27**(5), 771–780.

Dillon, C. F., Rasch, E. K., Gu, Q. and Hirsch, R. (2006), 'Prevalence of knee osteoarthritis in the United States: arthritis data from the Third National Health and Nutrition Examination Survey 1991-94.', *The Journal of rheumatology* **33**(11), 2271–2279.

D'Lima, D. D., Fregly, B. J., Patil, S., Steklov, N. and Colwell Jr, C. W. (2012), 'Knee joint forces: prediction, measurement, and significance', *Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine* **226**(2), 95–102.

Dryden, I. L. and Mardia, K. V. (2016), *Statistical shape analysis: with applications in R*, Vol. 995, John Wiley & Sons.

Felson, D. T. (2013), 'Osteoarthritis as a disease of mechanics', *Osteoarthritis and cartilage* **21**(1), 10–15.

Guilak, F. (2011), 'Biomechanical factors in osteoarthritis', *Best practice & research Clinical rheumatology* **25**(6), 815–823.

Hall, P. and Marron, J. (1987), 'On the amount of noise inherent in bandwidth selection for a kernel density estimator', *The Annals of Statistics* pp. 163–181.

Hebenstreit, F., Leibold, A., Krinner, S., Welsch, G., Lochmann, M. and Eskofier, B. M. (2015), 'Effect of walking speed on gait sub phase durations', *Human movement science* **43**, 118–124.

Helwig, N. E., Hong, S., Hsiao-Wecksler, E. T. and Polk, J. D. (2011), 'Methods to temporally align gait cycle data', *Journal of biomechanics* **44**(3), 561–566.

Hunt, M. A., Birmingham, T. B., Giffin, J. R. and Jenkyn, T. R. (2006), 'Associations among knee adduction moment, frontal plane ground reaction force, and lever arm during walking in patients with knee osteoarthritis', *Journal of biomechanics* **39**(12), 2213–2220.

Jansen, M., Hodgins, D., Mastbergen, S. C., Kloppenburg, M., Blanco, F., Haugen, I. K., Berenbaum, F., Eckstein, F., Roemer, F. W. and Wirth, W. (2024), 'Can gait patterns be explained by joint structure in people with and without radiographic knee osteoarthritis? Data from the IMI-APPROACH cohort', *Skeletal Radiology* **53**(11), 2409–2416.

Jolliffe, I. T. (2002), *Principal component analysis*, Springer Series in Statistics, second edn, Springer-Verlag, New York.

Jordan, J. M., Helmick, C. G., Renner, J. B., Luta, G., Dragomir, A. D., Woodard, J., Fang, F., Schwartz, T. A., Abbate, L. M., Callahan, L. F. et al. (2007), 'Prevalence of knee symptoms and radiographic and symptomatic knee osteoarthritis in African Americans and Caucasians: the Johnston County Osteoarthritis Project.', *The Journal of rheumatology* **34**(1), 172–180.

Jung, S., Dryden, I. L. and Marron, J. S. (2012), 'Analysis of principal nested spheres', *Biometrika* **99**(3), 551–568.
https://doi.org/10.1093/biomet/ass022

Kim, W. M., Dasgupta, S. and Srivastava, A. (2023), 'Peak-persistence diagrams for estimating shapes and functions from noisy data', *arXiv preprint arXiv:2305.04826* .

Kwon, S., Ro, D., Song, M., Han, H.-S., Lee, M. and Kim, H. (2019), 'Identifying key gait features associated with the radiological grade of knee osteoarthritis', *Osteoarthritis and cartilage* **27**(12), 1755–1760.

Lawrence, R. C., Felson, D. T., Helmick, C. G., Arnold, L. M., Choi, H., Deyo, R. A., Gabriel, S., Hirsch, R., Hochberg, M. C., Hunder, G. G. et al. (2008), 'Estimates of the prevalence of arthritis and other rheumatic conditions in the United States: Part II', *Arthritis & Rheumatism* **58**(1), 26–35.

Marron, J. and Dryden, I. L. (2021), *Objected Oriented Data Analysis*, Chapman and Hall/CRC.

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

Messier, S., Legault, C., Mihalko, S., Miller, G., Loeser, R., DeVita, P., Lyles, M., Eckstein, F., Hunter, D., Williamson, J. and Nicklas, B. (2009), 'The intensive diet and exercise for arthritis (idea) trial: Design and rationale', *BMC musculoskeletal disorders* **10**, 93.

Messier, S., Mihalko, S., Legault, C., Miller, G., Nicklas, B., DeVita, P., Beavers, D., Hunter, D., Lyles, M., Eckstein, F., Williamson, J., Carr, J., Guermazi, A. and Loeser, R. (2013), 'Effects of intensive diet and exercise on knee joint loads, inflammation, and clinical outcomes among overweight and obese adults with knee osteoarthritis the idea randomized clinical trial', *JAMA : the journal of the American Medical Association* **310**, 1263–73.

Messier, S. P., Loeser, R. F., Hoover, J. L., Semble, E. L. and Wise, C. M. (1992), 'Osteoarthritis of the knee: effects on gait, strength, and flexibility', *Archives of physical medicine and rehabilitation* **73**(1), 29–36.

Muniz, A. M. S., Manfio, E. F., Andrade, M. C. and Nadal, J. (2006), Principal component analysis of vertical ground reaction force: A powerful method to discriminate normal and abnormal gait and assess treatment, *in* '2006 International Conference of the IEEE Engineering in Medicine and Biology Society', pp. 2683–2686.

Rosadi, R., Jankaew, A., Wu, P.-T., Kuo, L.-C. and Lin, C.-F. (2022), 'Factors associated with falls in patients with knee osteoarthritis: A cross-sectional study', *Medicine* **101**(48), e32146.

Sims, E. L., Carland, J. M., Keefe, F. J., Kraus, V. B., Guilak, F. and Schmitt, D. (2009), 'Sex differences in biomechanics associated with knee osteoarthritis', *Journal of women & aging* **21**(3), 159–170.

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

Srivastava, A., Wu, W., Kurtek, S., Klassen, E. and Marron, J. S. (2011), 'Registration of Functional Data Using Fisher-Rao Metric', *arXiv preprint arXiv: 1103.3817* .

Trentadue, T. P. and Schmitt, D. (2024), 'Fourier analysis of the vertical ground reaction force during walking: Applications for quantifying differences in gait strategies', *Journal of Applied Biomechanics* **1**(aop), 1–9.

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.

Vos, T., Flaxman, A. D., Naghavi, M., Lozano, R., Michaud, C., Ezzati, M., Shibuya, K., Salomon, J. A., Abdalla, S., Aboyans, V. et al. (2012), 'Years lived with disability (YLDs) for 1160 sequelae of 289 diseases and injuries 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010', *Lancet (London, England)* **380**(9859), 2163.

Wiik, A. V., Aqil, A., Brevadt, M., Jones, G. and Cobb, J. (2017), 'Abnormal ground reaction forces lead to a general decline in gait speed in knee osteoarthritis patients', *World journal of orthopedics* **8**(4), 322.

Wu, W. and Srivastava, A. (2011), 'An information-geometric framework for statistical inferences in the neural spike train space', *Journal of Computational Neuroscience* **31**, 725–748.

Wu, Y., Huang, C. and Srivastava, A. (2024), 'Shape-based functional data analysis', *TEST* **33**(1), 1–47.

Xiang, S. (2023), Binary Expansion Testing and Gait Force Analysis, PhD thesis, University of North Carolina at Chapel Hill.

Younes, L. (1998), 'Computable elastic distances between shapes', *SIAM Journal on Applied Mathematics* **58**(2), 565–586.

Yu, Q., Lu, X. and Marron, J. (2017), 'Principal nested spheres for time-warped functional data analysis', *Journal of Computational and Graphical Statistics* **26**(1), 144–151.

Zeni Jr, J. A. and Higginson, J. S. (2009), 'Differences in gait parameters between healthy subjects and persons with moderate and severe knee osteoarthritis: a result of altered walking speed?', *Clinical biomechanics* **24**(4), 372–378.

Zhang, Y. and Jordan, J. M. (2010), 'Epidemiology of osteoarthritis', *Clinics in geriatric medicine* **26**(3), 355–369.
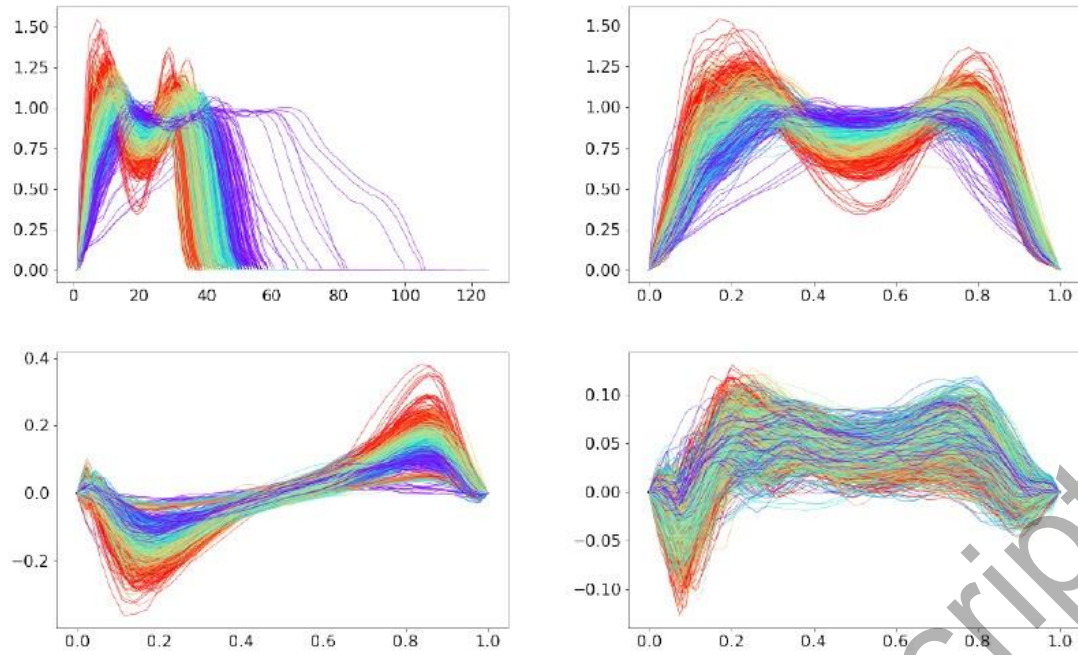
Figure 1: Top left panel: Raw vertical ground reaction force (vGRF) curves before processing, colored by walking speed. Top right: The same curves after re-scaling and interpolation of the time axis . The common rainbow color palette descends from fastest walking speed (red) to slowest (purple). Bottom row panels show apGRF curves (left) and mlGRF curves (right) after the same processing of the time axis. Note that the vertical axis (measured in N/kg, i.e. percentage of body weight) of each panel is scaled to the data it displays.
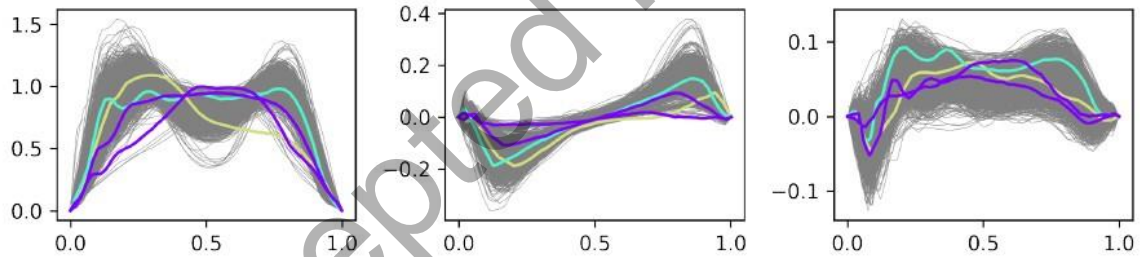
Figure 2: Left to right panels: vGRF, apGRF, and mlGRF curves. Examples of atypical GRF curves are highlighted, with representative atypical cases colored by walking speed and other curves in gray. These atypical curves lack the expected two-peak vertical structure or have near-zero, flat anterior-posterior components.
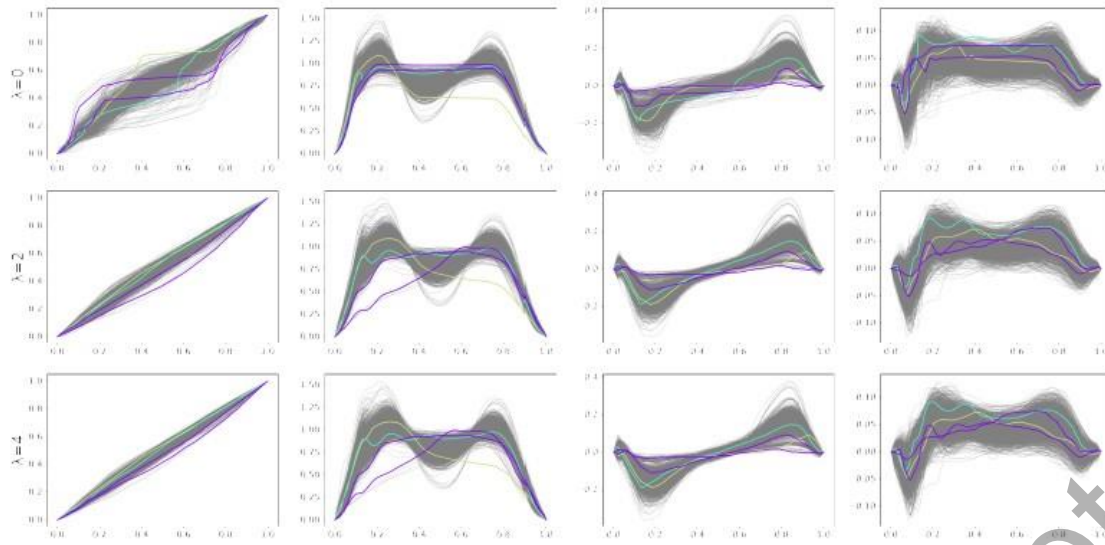
Figure 3: Alignment results for three elasticity parameter values: fully elastic ($\lambda = 0$), the selected $\lambda = 2$, and a more rigid case ($\lambda = 4$). Atypical cases are colored by speed, while other curves are shown in gray. The leftmost panels display the warping functions, and the second through fourth columns show the aligned curves. The staircasing effect in the fully elastic case ($\lambda = 0$) indicates over-alignment, while $\lambda = 4$ produces warping functions tightly close to identity, suggesting insufficient alignment. The selected $\lambda = 2$ balances these effects.
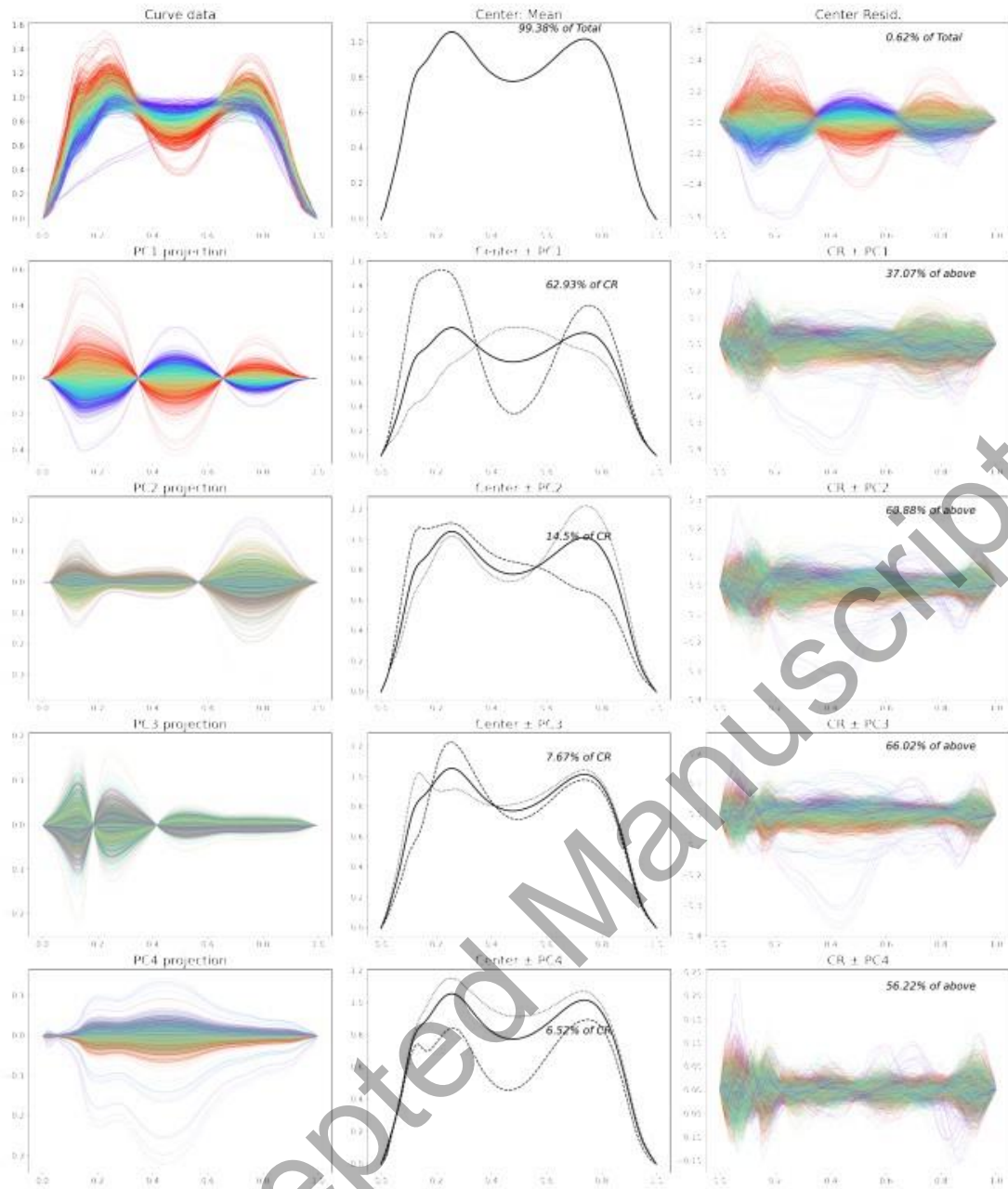
Figure 4: Modes of variation of the vGRF amplitude objects, shown in the top left panel. The first mode of amplitude variation, showing magnitude of the peaks and valley, is displayed in the second row, first panel. The panels in the middle column show the largest and smallest PC projections added back to the mean curve, while the solid black curve corresponds to the mean curve. The second and third modes of amplitude variation (third and fourth row, respectively) explain variation within each peak. The fourth mode indicates variation in overall magnitude, particularly of the valley, as seen in the middle panel of the last row.
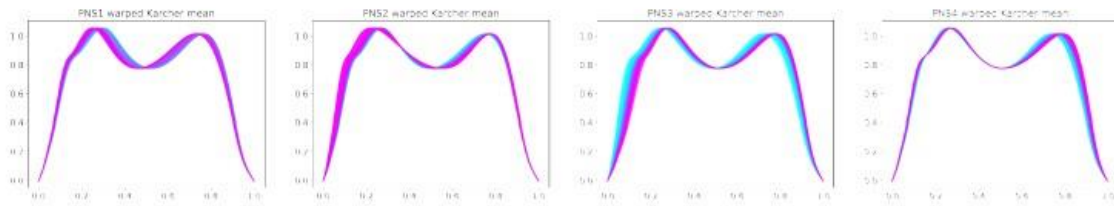
Figure 5: Visual representation of phase variation using warpings of the Karcher mean of the vGRF curves. Warping functions were generated by taking the inverse of the phase PNS projections added to the 45-degree line. Each panel shows curves colored by corresponding PNS scores, with cyan indicating the lowest scores and magenta indicating the highest. Curves are plotted in the order of the corresponding score to avoid over-plotting. The first panel (first mode) shows an overall shift in timing, with most apparent differences in the timing of the first peak (maximum heel-strike force) and valley. The second panel (second mode) shows variability in the closeness of the peaks. The third panel (third mode) represents an overall phase shift, suggesting earlier timing for curves with a small third bump before the first peak (cyan). The fourth panel (fourth mode) emphasizes variability in the timing of the second peak, independent of the rest of the curve.
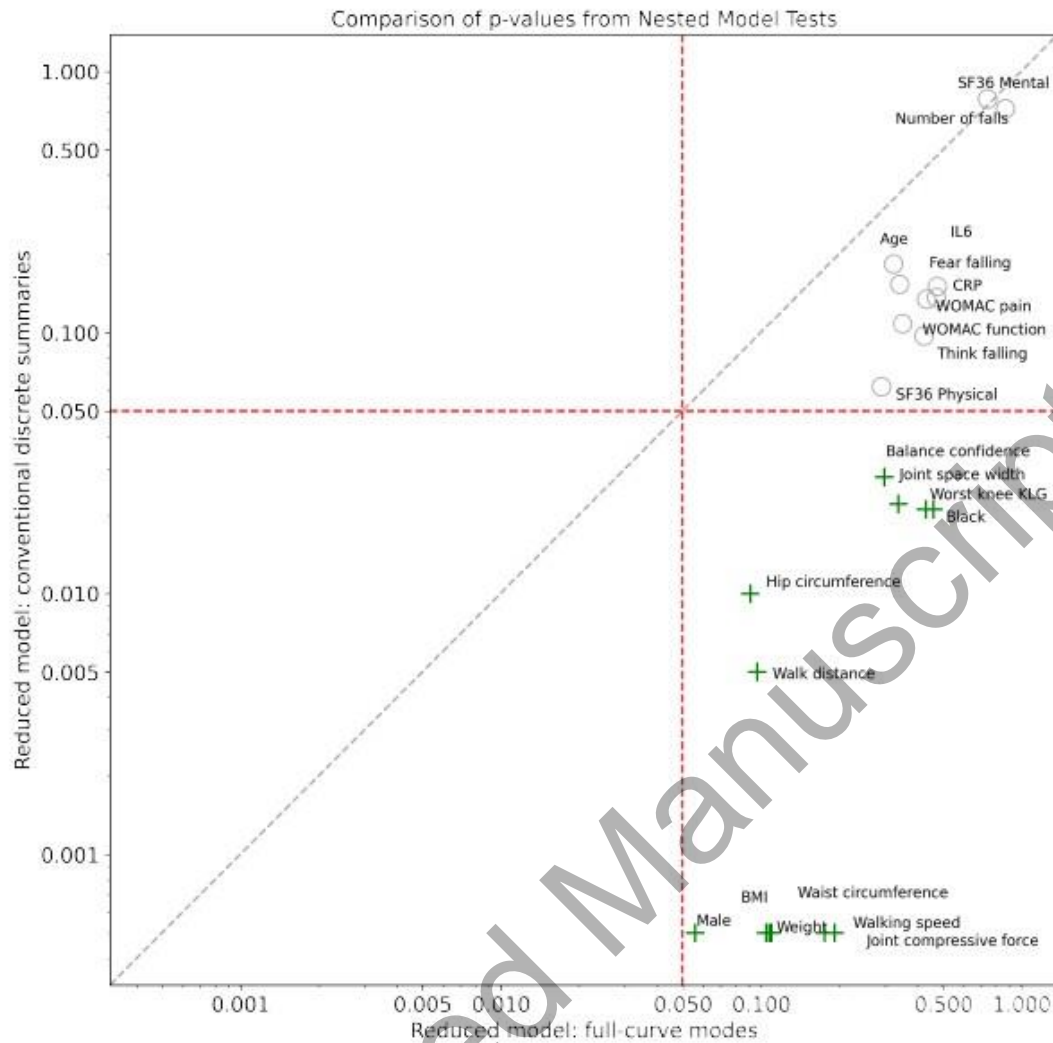
Figure 6: Scatter plot of bootstrap p-values resulting from LRTs of nested models, shown on a logarithmic scale. The horizontal axis represents the p-values for testing full-curve reduced models, and the vertical axis represents those for testing conventional discrete summary reduced models. For readability, the axes are labeled on the original scale. Gray circles represent traits where neither reduced model was rejected at this significance level, indicating that neither variable set adds substantial explanatory value beyond the other. Green plus signs represent traits where the full-curve reduced model was not rejected, but the conventional discrete summary reduced model was rejected, highlighting the added value of the full-curve approach. Note that the full-curve reduced model was never rejected and is consistently a suitable approach.