Estimating Knee Movement Patterns of Recreational Runners Across Training Sessions Using Multilevel Functional Regression Models

Marcos Matabuena[a1], Marta Karas[b1], Sherveen Riazati[c,d], Nick Caplan[d], Philip R. Hayes[d]

[a]Centro Singular de Investigación en Tecnologías Intelixentes, Universidad de Santiago de Compostela, Santiago de Compostela, Spain

[b]Department of Biostatistics, Harvard T.H. Chan School of Public Health, Harvard University, Boston, MA, USA

[c]Department of Kinesiology, San José State University, CA

[d]Department of Sport Exercise and Rehabilitation, Faculty of Health and Life Sciences, Northumbria University, Newcastle upon Tyne, UK

Corresponding author: MM (email: marcos.matabuena@usc.es)

[1]MM and MK joint first co-authorship.

ABSTRACT

Modern wearable monitors and laboratory equipment allow the recording of high-frequency data that can be used to quantify human movement. However, currently, data analysis approaches in these domains remain limited. This paper proposes a new framework to analyze biomechanical patterns in sport training data recorded across multiple training sessions using multilevel functional models. We apply the methods to subsecond-level data of knee location trajectories collected in 19 recreational runners during a medium-intensity continuous run (MICR) and a high-intensity interval training (HIIT) session, with multiple steps recorded in each participant-session. We estimate functional intra-class correlation coefficient to evaluate the reliability of recorded measurements across multiple sessions of the same training type. Furthermore, we obtained a vectorial representation of the three hierarchical levels of the data and visualize them in a low-dimensional space. Finally, we quantified the differences between genders and between two training types using

functional multilevel regression models that incorporate covariate information. We provide an overview of the relevant methods and make both data and the R code for all analyses freely available online on GitHub. Thus, this work can serve as a helpful reference for practitioners and guide for a broader audience of researchers interested in modeling repeated functional measures at different resolution levels in the context of biomechanics and sports science applications.

keywords: Biomechanics; Knee movement; Multilevel functional data analysis; Patterns; Subsecond-level data; Wearable sensors.

1 Introduction

Recent advances in technology have led to the ever-increasing popularity of wearable technology in health research [54]. Modern sensors can monitor an individual's motor activity with great accuracy and measure various physiological and biomechanical variables in a near-continuous manner. This provides an opportunity to have a detailed assessment of an athlete's physical capability and performance [32], and to schedule optimal interventions [29, 6]. Promising fields for implementing these novel strategies are sports training and biomechanics [21, 55].

1.1 Wearable Sensors: Opportunities in Sports and Biomechanics

Although we are in the early stages of this technological revolution, the first research papers are appearing that use through high-resolution data gathered with biosensors to answer unknown and complex questions about training load [7], daily biomechanical patterns [25], and injury prediction [4, 35]. Furthermore, sensor data may enable us to build predictive models that support decision-making and help optimize the performance [36, 19, 43]. For example, several recent works provide new epidemiological knowledge using biomechanical data of human locomotion [26, 62]. Other papers have tried to predict sports injuries [48] or other motor or neurological diseases prematurely [1], or even the impact of therapy together with their prognosis in the recovery phase after surgery [24].

Importantly, with the boom of wearable devices, their use is increasingly common among professional athletes and the general population, such as amateur runners. Thus, the remote control of athlete training and even monitoring their daily routine outside of sports activity is feasible and opens a broad spectrum of opportunities in biomechanics applications.

1.2 Quantifying Biomechanics Patterns in Walking and Running: Methodological Challenges

In both sports and general populations, abnormal movement patterns are synonymous with muscular and motor problems, risk of injury, or even the appearance of severe neurological diseases such as Parkinson's [38]. Therefore, characterizing movement patterns and detecting their abnormalities in biological activities such as walking and running, are essential.

The predominant data analysis practice in gait biomechanics is to summarize the curve recorded for each stride using several statistical metrics and apply standard multivariate techniques. However, this traditional approach yields a substantial loss of information given that gait data recorded is functional, such as a cycle of gait movement.

A more detailed and meaningful analysis can be attained by using a complete stride cycle with functional data analysis (FDA) techniques [16, 62]. In FDA applications for biomechanics, the general procedure is to normalize data curves collected at a fixed body location for each step into the $[0,1]$ interval, compute the mean of the multiple curves recorded, and create an average functional curve for analysis. However, this procedure can be suboptimal because the constructed mean representation ignores the individual variability between the distinct steps of the same individual – a crucial feature in evaluating the movement patterns in some settings. In addition, the mean curve statistic can be very sensitive to outliers that are frequently observed in biomechanical data. This is particularly true in measuring movements performed at high or low speed, where sensor and human variability often increase. Moreover, we often need to compare the effect of interventions along with the different training sessions on different days, and for this, we have several repeated measures per individual in different periods. In such cases, a more suitable approach might be to

employ multilevel functional data analysis (MFDA) models that allow accounting for a natural hierarchy in data [33]. For example, using MFDA models, we can estimate biomechanical patterns using data curves from multiple gait cycles, using data from a significant fraction of or even from a complete training session. MFDA methods also capture the variations in different periods at an intra-inter individual level and evaluate the changes produced along with relevant outcomes at different resolution levels among individuals.

Surprisingly, there is little use of FDA techniques within the applications literature, either in sports or other clinical areas [56]. This might be partially due to FDA being a relatively novel modeling approach and, consequently, a lack of broader knowledge about the value of using the FDA for biomedical and biomechanics data.

1.3 This Paper's Contributions

This work demonstrates statistical modeling using MFDA models to characterize knee biomechanical patterns along with two training type sessions. Specifically, we use subsecond-level data of a knee location, recorded in three dimensions, collected in 19 recreational runners during (a) medium-intensity continuous run (MICR); (b) high-intensity interval training (HIIT) sessions, with 20 steps recorded during each participant-session. Using multilevel functional models, we estimate functional intra-class correlation coefficients to evaluate the reliability of measurements across two separate HIIT sessions. We also computed the scores of different hierarchical levels of multilevel functional models to analyze variability patterns between individuals, runs, and strides, and to visually compare the scores by gender. We further quantify differences in knee position trajectories between genders and between the two intensities of exercise sessions (MICR vs. HIIT). Finally, we provide the overview of the relevant methods and make both data and the R code for all analyses freely available online on GitHub (martakarass/biomechanics-manuscript). For the biomechanical practitioner's audience, this paper provides a methodological guide and read-and-go R code examples to address questions similar to the following we tackle:

1. What is the reliability of functional running measurements in two independent HIIT running sessions?
2. What are the different modes of variability at different hierarchical levels in the data (e.g., individual level, session level, and a running stride level)?
3. What are the population-level differences in knee location trajectories between MICR and HIIT running sessions and between genders?

To date, several works have studied the etiology of running-related knee injuries in recreational runners, some even using three-dimensional time series analysis [37]. However, to the best of our knowledge, no previous studies have compared biomechanical changes during HIIT and MICR training, nor investigated the reliability of biomechanical measures at the knee in two or more training sessions.

2 Methods

2.1 Study Design and Population

Data used in this manuscript were collected in a study that recruited 20 participants to complete two types of energy expenditure-matched running sessions: a medium-intensity continuous run (MICR) and a high-intensity interval training (HIIT) session. Participant enrollment criteria and study design have previously been reported in detail in [47]. In short, 20 healthy, experienced runners (10 women and 10 men) were recruited.

For HIIT sessions, athletes ran 6 × 800 meters intervals at a pace of 1 km/h below their maximum aerobic speed with $1:1$ recovery time. For the MICR sessions, the athletes completed a continuous run halfway between the speeds at lactate threshold and the lactate turn point. The duration of the MICR session was individualized to yield the same estimated energy expenditure as the HIIT session. All sessions were conducted at the same time of the day to avoid diurnal variation. All sessions were performed in an environmentally controlled laboratory setting, with all the athletes using the same treadmill. Running kinematics used in this analysis were recorded at the start of the final minute of each run. Measurements were recorded in three dimensions with the Vicon Nexus motion analysis system (Vicon Motion Systems Ltd, Oxford, United Kingdom) at a frequency of 500 Hz. Data

recorded with the motion analysis system were further segmented to extract individual stance phases.

This work uses data from two independent HIIT sessions and one MICR session. Specifically, 20 cycles of the stride stance phase for 19 participants were analyzed. Data for one of the participants were excluded from the analysis set due to missing data in some parts of running strides. Our analysis focuses on subsecond-level recordings of knee segment trajectories within each of the three dimensions, *x*, *y*, and *z*.

2.2 Why use functional data analysis?

A core element of many functional data analysis methods is to approximate the vector observations recorded for each individual as a through a basis of functions. This conceptual leap from multivariate euclidean space to functional space presents several advantages compared to high-dimensional multivariate data analysis techniques.

- The information can be represented and summarized in low-dimensional spaces with the same or more accuracy than with multivariate analysis techniques. As a consequence, functional data analysis may lead to inferential methods with more powerful and robust hypothesis testing, and more accurate predictive models. Moreover, the computational efficiency of the algorithms may increase.
- The methods can also reduce noise and be more robust towards outlier data points. More advanced techniques exist to remove and filter the measurement error.
- Functional regression models can be evaluated at any point of the continuous domain and not only within the finite set of domain points where the data were recorded. This presents opportunity for better interpretability of the results where the observed data is a function, recorded on a discrete grid, defined on some continuous domain (e.g., gait cycle).
- Analyses can evaluate rates of change of the underlying function.

2.3 Multilevel Functional Data Analysis (MFDA) Models

This subsection provides foundations and then reviews the methods used in our statistical data analysis. We first review frameworks for standard functional principal component analysis (FPCA) and multilevel FPCA, then outline several general formulations of multilevel functional data analysis (MFDA) models, and discuss intra-class correlation coefficient and hypothesis testing between different levels. While provided for completeness, some technical parts of this presentation are kept in the Supplementary Material for the succinctness of the main text.

2.3.1 Functional Principal Component Analysis (FPCA)

Functional principal component analysis (FPCA) technique, an extension of multivariate principal components analysis, is widely used in FDA to describe the variability of a sample of curves when one curve per subject is available. In summary, FPCA decomposes the space of curves into principal directions of variation.

To describe FPCA framework, let $X(t), t \in [0,1]$, be a random function with mean $\mu(t) = E(X(t))$ and covariance function $\Sigma(t,s) = E\{(X(t) - \mu(t))(X(s) - \mu(s))\}$ for all t, $s \in [0,1]$. The heart of many FDA models is based on calculating modes of variability of the random function $X(t)$ based on the spectral decomposition of the covariance operator $\Sigma(\cdot,\cdot)$ in a set of eigenfunctions $\{e_i(\cdot)\}_{i=1}^{\infty}$ and eigenvalues $\{\lambda_i\}_{i=1}^{\infty}$, where $\lambda_1 \geq \lambda_2 \geq \cdots$. Specifically, from the decomposition of Karhunen-Loève we have

$$X(t) = \mu(t) + \sum_{k=1}^{\infty} c_k e_k(t), \qquad (1)$$

where $c_k = \int_0^1 (X(t) - \mu(t)) e_k(t) dt$ are uncorrelated random variables with mean zero and variance λ_k. These variables are usually known as scores or loading variables.

In the real-world FDA setting, we typically consider n realizations, generally independent, of the process $X(\cdot)$, denote $X^1(\cdot), \ldots, X^n(\cdot)$. Also, we typically only observe a sample of n vectors, assume each vector of length m, denote $X^1, \ldots, X^n$, sampled in a grid $\{0 \leq t_1 < \ldots < t_m \leq 1\}$, where $X_j^i = X^i(t_j)$ for all $i = 1, \ldots, n,\ j = 1, \ldots, m$.

A given functional sample can be used to estimate mean function μ and covariance function Σ,

$$\hat{\mu}(t_j) = \frac{1}{n}\sum_{i=1}^{n} X^i(t_j), \quad (2)$$

$$\hat{\Sigma}(t_j, t_k) = \frac{1}{n}\sum_{i=1}^{n} \left(X^i(t_j) - \hat{\mu}(t_j)\right)\left(X^i(t_k) - \hat{\mu}(t_k)\right), \quad (3)$$

for all $j, k = 1, \ldots, m$.

In many applications where observations are subject to a large measurement error, a smoothing step is taken in the above procedure to ensure the optimal performance of the empirical estimator $\hat{\Sigma}$. Three different smoothing strategies have been generally used in the literature [51, 9]: (1) smoothing of the original functional data; (2) introduction of a regularization term in the estimation of $\hat{\Sigma}$; and (3) direct application of a smoothing procedure in the raw estimation of $\hat{\Sigma}$.

Next, eigenvectors $\{\hat{e}^i\}_{i=1}^m$ and eigenvalues $\{\hat{\lambda}_i\}_{i=1}^m$ can be estimated from the empirical covariance function $\hat{\Sigma}$ via the spectral theory of linear algebra, similarly as in the context of classical PCA in multivariate statistics. Finally, K ($K < m$) eigenvectors $\{\hat{e}^i\}_{i=1}^K$ and eigenvalues $\{\hat{\lambda}_i\}_{i=1}^K$ can be selected and used to provide the following decomposition:

$$X^i(t_j) \approx \hat{\mu}(t_j) + \sum_{k=1}^{K} \hat{c}_k^i \hat{e}_j^k \quad (4)$$

for $i = 1, \ldots, n$, $j = 1, \ldots, m$, and $\hat{c}_k^i = \langle X^i - \hat{\mu}, \hat{e}^k \rangle$, where $\langle , \rangle$ denotes the usual scalar-product, and $\hat{e}_j^k$ is the j-th component of the eigenvector $\hat{e}^k$. In applications, a small $K < m$ is often sufficient to capture the important modes of variations in the elements of the random sample. More details of these procedures can be found in the reviews and general books of functional data analysis, where different estimation procedures of the number of components, K, are established [51, 27, 34].

2.3.2 Multilevel Functional Principal Component Analysis (MFPCA)

In the previous subsection (Sect. 2.3.1), we presented a FPCA procedure applicable when a random functions are measured once for *n* independent units (e.g., *n* individuals). In practice, it is common to have several repeated functional measurements for each individual in the data set. For example, in sports applications, functional data of an an athlete may be collected at multiple points in time (e.g., over the training session and/or when investigating the training load throughout a season). Such settings yield multiple functional observations per individual, and thus yield a correlation structure in data for whom the previously discussed FPCA procedure may be inadequate. To review the statistical framework for multilevel FPCA (MFPCA), we first expand the notation introduced in Sect. 2.3.1 while employing a specific setup of our biomechanics real data example.

Let $X^{i,j,k}(t), t \in [0,1]$, be a random function – a *k*-th stride for *j*-th race in the *i*-th individual, for $i = 1,\dots,n,\ j = 1,\dots,n_i,\ k = 1,\dots,K_{i,j}$; for simplicity, onward, assume that $K_{i,j} = M$ and $n_i = J$ for all $i = 1,\dots,n$. For illustration as is the model that we use in the study-case, we first consider the following three-way functional nested ANOVA model

$$X^{i,j,k}(t) = \mu(t) + Z^{i}(t) + W^{i,j}(t) + U^{i,j,k}(t) + \epsilon^{i,j,k}(t) \quad (5)$$

for $i = 1,\dots,n,\ j = 1,\dots,J,\ k = 1,\dots,M$. In Equation 1, $\mu(t)$ is the mean global, $W^{i,j}(t)$ is the residual subject- and race-specific deviation from the global mean, and $U^{i,j,k}(t)$ is the residual subject- race- and stride-specific deviation from the global mean. In this framework, $\mu(t)$ is treated as a fixed function, while $Z^{i}(t)$, $W^{i,j}(t)$, and $U^{i,j,k}(t)$ are treated as a random function of mean zero. Moreover, with the proposal of identification correctly the model, we assume that $Z^{i}(t), W^{i,j}(t)$, and $U^{i,j,k}(t)$, are random uncorrelated functions. In the literature, functions $Z^{i}(t)$'s, $W^{i,j}(t)$'s, $U^{i,j,k}(t)$'s are known as the 1-level, 2-level and 3-level functions, respectively. We note that we introduce for convention a additional random error $\epsilon^{i,j,k}(t) \sim N(0,\sigma^2)$ in terms to identify the model.

Again, the MFPCA framework relies on the Karhunen-Loève decomposition. For example, for the model defined in Equation 1, we have

$$X^{i,j,k}(t) = \mu(t) + \sum_{r=1}^{\infty} c_r^i e_r^{(1)}(t) + \sum_{r=1}^{\infty} d_r^{i,j} e_r^{(2)}(t) + \sum_{r=1}^{\infty} f_r^{i,j,k} e_r^{(3)}(t) \quad (6)$$

for $= 1, \ldots, n,\ j = 1, \ldots, J,\ k = 1, \ldots, M$. In the above, $\{e_r^{(1)}\}_{r=1}^{\infty}, \{e_r^{(2)}\}_{r=1}^{\infty}$, and $\{e_r^{(3)}\}_{r=1}^{\infty}$, are the eigenfunctions related to the random functions of levels 1, 2 and 3, respectively, while $\{c_r^i\}_{r=1}^{\infty}$, $\{d_r^{i,j}\}_{r=1}^{\infty}, \{f_r^{i,j,k}\}_{r=1}^{\infty}$ are the eigenvalues of levels 1, 2, and 3, respectively.

The eigenfunctions and eigenvalues of the model defined in Equation 2 can be derived from covariance funtions: $\Sigma_T(s,t) = Cov\left(X^{i,j,k}(s), X^{i,j,k}(t)\right)$ – "total" covariance function, $\Sigma_W(s,t) = Cov(W^{i,j}(s), W^{i,j}(t))$ – "between" covariance function (covariance between the units of the second level while keeping the effect of the first level fixed), $\Sigma_Z(s,t) = Cov(Z^i(s), Z^i(t))$ – "between" covariance function (covariance between the units of the first level). As explained in Appendix A in the Supplementary Material, a method of moments can be used to obtain these covariance operators.

2.3.3 MFDA Models: More General Formulations

Different levels of hierarchy may appear in real problems that can be nested (e.g., three-way functional ANOVA model presented in Sect. 2.3.2) or crossed. For example, following [53], Table 1 presents a number of possible data hierarchy scenarios corresponding model formulations.

The models specified in Table 1 share the same model formula structure: $X(t) = \mu(t) + (\sum \text{latent processes}) + \epsilon(t)$, where $\mu(t)$ is the mean curve or fixed effect and $\epsilon(t)$ is a white noise, $\epsilon(t) \sim N(0, \sigma^2)$ for all $t \in [0,1]$. The latent processes are assumed to be zero-mean and square-integrable so that they are identifiable, and the standard statistical assumptions for scalar outcomes can be generalized to functional data. In this way, the total variability of a functional outcome is decomposed into a sum of process-specific variations plus σ^2. For these models, the algorithm details of estimation procedure are provided in Appendix B in the Supplementary Material.

2.3.4 Intra-class correlation coefficient (ICC)

When several repeated measurements are collected from a subject over different days or other periods, it is often of interest to determine how much variability is

explained by the subjects' effect and how much by collecting measurements over different levels of the data hierarchy. This problem is known in the literature as estimating the coefficient of intra-class correlation (ICC) [39] that represents the variability arising from measuring a subject in conditions that are assumed to be standardized across different tests. The estimation of ICC is crucial, for example, in the field of clinical laboratory testing, where one often wants to use clinical variables that are not modified abruptly between days as a result of a device's measurement error or by intra-day variability of individuals [49]. In biomechanics and exercise sciences, the ICC estimation is critical in searching for objective criteria to assess performance and control the individual's degree of fatigue [58, 28]. For example, while a variable may have high variability, it may pose a helpful criterion for decision-making. In such cases, it is necessary to make several measurements to capture that variable accurately and ICC can allow us to quantify how many measures need to be made to capture the variable's distribution accurately.

To define ICC for a functional model in our setting, consider the $(N3)$ model (see Table 1) given by

$$X^{i,j,k}(t) = \mu(t) + Z^{i}(t) + W^{i,j}(t) + U^{i,j,k}(t) + \epsilon^{i,j,k}(t). \qquad (7)$$

For a fixed $t \in [0,1]$, by analogy with a univariate non-functional case, the proportion of the total variability explained by the subjects' effect at point *t* is given by

$$\rho(t) = \frac{Var\left(Z^{i}(t)\right)}{Var\left(Z^{i}(t) + W^{i,j}(t) + U^{i,j,k}(t) + \epsilon^{i,j,k}(t)\right)}, \qquad (8)$$

where $\rho(t)$ is the intra-class correlation coefficient at point *t*.

The ICC formula from Equation 3 can be generalized as a global measure following [52]. In particular, we divide the variability generated by the hierarchy associated with subjects by the sum of all variability sources:

$$\rho = \frac{\sum_{k=1}^{\infty} \lambda_k^{(1)}}{\sum_{k=1}^{\infty} \lambda_k^{(1)} + \sum_{k=1}^{\infty} \lambda_k^{(2)} + \sum_{k=1}^{\infty} \lambda_k^{(3)} + \sigma^2}.$$

2.3.5 MFDA Models incorporating covariate information

In many applications, together with a functional outcome variable $X^{i,j}(\cdot), t \in [0,1]$, for $i = 1,\dots,n$ and $j = 1,\dots,n_i$, we have additional dynamic or static information available about a subject, e.g. demographics or a clinical condition, that we may want to incorporate in the model as fixed and/or random effects. Consider a multilevel functional regression model that incorporates information about an individual's characteristics, given by

$$X^{i,j}(t) = \langle M^{i,j}, \beta(t) \rangle + \langle Z^{i,j}, u_i(t) \rangle + \epsilon^{i,j}(t), \quad (9)$$

for $i = 1,\dots,n$ and $j = 1,\dots,n_i$, where $Z^{i,j} \in \mathbb{R}^m$ denote a random variable measure collected in random design, $M_{i,j} \in \mathbb{R}^p$ denote fixed-effect terms in the model, $\beta : t \in [0,1] \to \beta(t) \in \mathbb{R}^p$ and $u_i : t \in [0,1] \to u_i(t) \in \mathbb{R}^m$ for $i = 1,\dots,n$, denote coefficient functions for the fixed and random effect terms.

Obtaining a global estimation of the Equation 4 model is challenging. Recent work by [12] proposes an efficient estimation strategy of such complex models using the following steps (see [12] for further details):

1. For each point t of the observed functional data grid, fit a separate point-wise generalized linear mixed model using standard multilevel software.
2. Smooth the model coefficients obtained at different points t with a linear smoother along the functional domain.
3. Obtain a global model inference with a joint confidence band using analytics approach for Gaussian data or using bootstrap for Gaussian/non-Gaussian data.

An important topic for practitioners is the connection between global p-value and pointwise confidence intervals [50, 45]. For example, following [50], for each covariate, we can define a natural global p-value as a minuscule level α in the joint confidence interval that does not contain zero. Although other alternatives exist in the literature [45], the [50] approach appears to us easier to apply in terms of interpretability.

2.4 Statistical Analysis

We applied the methods outlined above to our real data in the following steps. First, we smoothed the raw biomechanical functional profiles with a linear smoother to remove potential measurement error. Then, we visualized functional biomechanics profiles collected during two different HIIT sessions, and inspected the differences at an individual level. We also computed and visualized sample mean and standard deviation of functional profiles across genders and races.

We further fitted a three-level multilevel model without covariates (see model N3 in Table 1) to data collected during two different HIIT sessions to decompose the different modes of variability between the hierarchical levels and to obtain a vectorial representation of units that compose the model. With this decomposition, we visualized the scores from hierarchical levels 1,2, and 3, inspected scores differences by gender, computed cumulative variance explained by subsequent functional principal components at each hierarchical level, and visualized first two eigenfunctions at each hierarchical level.

Finally, to quantify the differences in biomechanical patterns between HITT and MICR session types and between the genders, we fitted a multilevel functional model to data collected during one HIIT session and one MICR session. In that model, we set knee location trajectory for each stride as a functional outcome, included a fixed effect for gender (coded 1 for male and 0 for female), a fixed effect for run session type (coded 1 for HIIT and 0 for MICR), and a subject-specific random intercept and random slope for the run type term. Formally, the fitted model formula is given by

$$
\begin{aligned}
Y^{i,j}(t) = {} & \beta_0(t) + [\mathrm{male}]_i \beta_1(t) + [\mathrm{HIIT}]_{i,j} \beta_2(t) \\
& + u_{1i}(t) + [\mathrm{HIIT}]_{i,j} u_{2i}(t) \\
& + \epsilon^{i,j}(t),
\end{aligned}
$$

where *i* denotes the participant's index, *j* denotes participant's functional observation index, $[\mathrm{male}]_i$ is sex indicator of *i*-th participant (equal to 1 for male and 0 for female), $[\mathrm{HIIT}]_{i,j}$ is session type indicator of *j*-th functional observation of *i*-th participant (equal to 1 for HIIT and 0 for MICR), $Y^{i,j}(t)$ is *j*-th functional observation of *i*-th

participant, $\beta_0(t), \beta_1(t), \beta_2(t)$ is fixed effect coefficient functions, $u_{1i}(t), u_{2i}(t)$ is random-effect coefficient functions, $\epsilon^{i,j}(t)$ is residual error process.

All analyses were performed using statistical software R. The data and the R code used can be found on GitHub (martakarass/biomechanics-manuscript).

3 Results

3.1 Descriptive analysis

Figure 1 shows knee location trajectories raw data collected from 20 running strides per individual during two HIIT sessions. Each plot corresponds to data of one participant. Measurements are shown with separate colors for data recorded from *x*, *y* and *z* dimensions of a 3D plane. By visual inspection, we can see there are individuals for which there is little difference in their biomechanical stride patterns within an axis between the two HIIT sessions (e.g., participant with ID 7). For others, noticeable differences are present, demonstrating higher intra-individual variability (e.g., participant with ID 18). In addition, the observed stride patterns exhibit a noticeable between-individual variability.

Figure 2 shows sample mean and ± 95% confidence intervals (CIs) of the mean computed point-wise from raw data collected from 20 running strides per individual during one HIIT and one MICR session. Sample statistics were computed separately for data collected from each measurement axis (*x*, *y*, *z*), in strata according to gender (female, male) and race type (MICR, HIIT). In this descriptive analysis, we observe significant differences between genders for each measurement axis data along whole functional domain (stride cycle). Differences between run types were signifcant at subsets of the functional domain for data from measurement axis *x* and *y*. However, this exploratory analysis does not account for repeated observations per individual, and does not provide a more holistic picture by simultaneously incorporating information from gender and race type. In addition, the functional nature of the recorded measurements is not exploited, leaving potential opportunities for increased inference efficiency unused. This is the rational for the use of the multilevel functional models.

3.2 Estimating scores and ICC with multilevel functional model (without covariates)

A three-level multilevel functional (N3) model was fitted for knee location data collected during two HIIT sessions, separately for data from *x*, *y* and *z* dimensions of a 3D plane. Figure 3 shows values of the first two scores along the three hierarchical levels considered in the models: level 1 – an individual (1 value per participant), level 2 – a run session (1 × 2 values per participant), level 3 – a running stride ($1 \times 2 \times 20$ values per participant). Table 2 summarizes data from Figure 3 by providing sample mean and 95% CIs of the mean for scores across strata by gender (female, male). Figure 4 shows cumulative variance explained by subsequent functional principal components of level 1, 2 and 3, calculated with N3 multilevel models. In addition, Figure C1 in Appendix C in the Supplementary Material shows the first two eigenfunctions associated with each of the three hierarchical levels of the N3 multilevel models.

In this analysis, the first eigenvalue alone captures more than 90% of variability at levels 1 and 2 for models for data collected at each of three axes (*x*, *y* and *z*); at hierarchical level 3, between two to three components are needed to cross the 90% threshold of variability explained.

From a visual inspection of Figure 3, we seem to observe distinct patterns in the distribution of scores between males (“M”) and females (“F”); for example, females’ scores are mostly centered at the top-left corner of the top-left plot in the Figure 3. Table 2 provides the numeric observations from figure 3. In general, they not we observe differences in point estimates of the mean of scores between genders. However, we are not statistically significant according to the 95% confidence intervals. Given the limited sample size, this marginal analysis requires more data to obtain enough power. A potential solution is to use a more integrated approach and to analyze all covariate information available in our setting simultaneously.

Using the variability decomposition derived from the model, we estimated the functional ICC to be 0.55, 0.54, and 0.61 for measurements from *x*, *y* and *z* dimensions of a 3D plane, respectively. The ICC values obtained indicate that, in

general, the reproducibility of measurements across the two HIIT tests was moderate.

3.3 Estimating the run type and gender effect with multilevel functional model (with covariates)

Figure 5 shows the results of estimating knee location trajectory with a multilevel functional model for each stride set as functional outcome, a fixed effect of gender (coded 1 for male and 0 for female), a fixed effect of run type (coded 1 for HIIT and 0 for MICR), and a subject-specific random intercept term. Smoothed coefficient estimates are denoted using blue dashed lines. Point-wise and joint 95% confidence intervals are shown as the dark and light gray shaded area, respectively.

Following the notion of a global p-value (see Sect. 2.3.5), the results indicate statistically significant (at $\alpha = 0.05$) time-varying effect of session type for the measurements from *x*, *y* and *z* dimensions of a 3D plane. The effect is present despite the two different session types were performed with the same energy expenditure. The corresponding confidence intervals are relatively narrow – likely due to the fact we included participant-specific random slope for run session type. For the gender differences, the effect is borderline-significant for measurements from *z* dimension of a 3D plane (and is not significant for the other dimensions of measurement).

Notably, our results show more differences in the extreme phases (beginning, end) of the biomechanic cycle than in the middle phase regarding uncertainty (gender effect) and value magnitude (race effect for data from measurement axis *x* and *z*). This can be as explained as follows: the middle of the stride is the ground contact phase, i.e., the foot is in contact with the ground, and it is, therefore, unlikely to differ between the two-run types. However, stride mechanics vary with running speed. As running speed increases, there is an increasing reliance on forces generated at the hip rather than the ankle [14]. The net result is an increased range of movement with a higher knee lift and longer back swing. The differences at the extremes of the stride found in this study can be hence explained by the difference in running speed between the two-run types.

4 Discussion

Knee injuries are one of the most frequent problems faced by recreational runners [57]. Therefore, an accurate characterization of the biomechanical changes that occur in typical training sessions can be critical in identifying the etiology of injuries [13] and developing predictive models to detect injury risk [10]. Here, we have illustrated how to use multilevel functional models to exploit functional information from running strides to: (i) examine the different modes of variability of data at different levels of hierarchical structure and obtain a specific vector-valued data representation; (ii) measure the reliability between two training sessions of the same type; (iii) analyze the biomechanical differences between HIIT vs. MICR race types and the gender effect, an unexplored research topic in biomechanics literature. In particular, we believe we are the first to employ functional multilevel models with covariate information in biomechanics literature, which we did to address the issue (iii). The data and R code to reproduce all presented results are publically available on GitHub (martakarass/biomechanics-manuscript).

The complete analysis of each cycle through functional analysis techniques that analyze the curve in its totality has led to more nuanced findings [13]. Traditional techniques that analyze either fixed angles, the average angle, the range of movement or other measures summarized, result in the loss of information that its use entails. Complementary, interesting problems can be identified when using more informative gait points. Recent statistical methodologies can be used to address this problem [3, 46].

Functional multilevel models are an essential weapon in the challenge to exploit information from monitoring athletes or patients, to optimize decision-making using different sources of information and measurements, made at different resolution levels. These tools can help integrate and analyze the information together, obtain a representation of the individuals along with different levels of hierarchy, and establish the different forms of variability in the different levels considered. These tools are remarkable if we want to analyze all training records or physiological variables of a group of athletes over a season or different micro-macro-cycles [31, 18]. For example, there is not yet a sufficiently good methodology to represent the

information inherently as proposed by these models [36, 43, 23]. Despite being an exciting research topic with high relevance, we believe that there are not many methodologies to address relevant problems in biomechanics to date. For example, a specific need of this field could be to build a multilevel model that considers the different time lengths of a step, and does not lose the information on the step geometry with the standardization of all the strides to the $[0,1]$ interval.

The multilevel models have allowed us to calculate the intraclass correlation coefficient between the two interval training sessions taking into account the 20 steps recorded in each session. To the best of our knowledge, this is a novel approach in this area since the traditional approaches previously used to measure reliability rely on the compression of information in the average curve and only between two conditions [44]. This constitutes an important analytic advance, since with the inclusion of the 20 steps in the model in each test, we have more information, and with the new procedure, we can see if there are statistical differences between the different levels of hierarchy or groups of patients/athletes taking into account the potential differences in the study design.

An important aspect to consider in analyzing the results is that the individuals' movement patterns seem unique. This is not new, and several papers have exempted the individuality of human walking and running [20]. In this sense, since the biomechanical patterns are probably grouped in clusters [41, 22], standard hypothesis tests applied to the whole sample are not the best way to establish biomechanical differences. There are some discrepancies between studies when examining these issues. Also, in the biomechanics literature, as in other biomedical literature areas, there is some controversy about the use of p-value [2], and the use of other approaches such as effect size [5] or e-values [61] may be recommended.

A limitation of this study is a relatively small sample size (19 participants), together with the fact that we are analyzing the biomechanical variations of the knee, without taking into account the possible multivariate structure of knee movement. Recently, some papers have emerged about this topic in another application [40, 8]. However, the literature with the multilevel model is sparse [60]. The present importance of computational and methodological limitations (see the discussion about the additive

model for [50]) In this paper, due to the reduced number of data, we think that we can gain a greater interpretation in this type of study of a more exploratory character with this procedure. Moreover, this work's main purpose is to illustrate the use of classical multilevel models with biomechanical data.

The rise of biosensors [17, 42, 54] in the area of biomechanics and medicine is causing an unprecedented revolution in the evaluation of athletes and patients care. It is likely that in the coming years, many of the clinical decisions will also be supported by the values predicted from the algorithms in many contexts, such as the prediction of injuries [11, 59] or optimal surgery recovery [24, 30] so in sport and general populations. Undoubtedly, the introduction of the data analysis techniques discussed here will help practitioners analyze objects that vary in a continuum repeatedly and that appear more and more frequently in biomedical data [15].

Acknowledgments

This work has received financial support from the Xunta de Galicia - Consellería de Cultura, Educación e Universidade (Centro de investigación de Galicia accreditation 2019-2022 ED431G-2019/04 and the European Union (European Regional Development Fund - ERDF).

Ethics Statement

The studies involving human participants were reviewed and approved by Northumbria University. The patients/participants provided their written informed consent to participate in this study.

References

[1] Belić, M., Bobić, V., Badža, M., Šolaja, N., Đurić-Jovičić, M., and Kostić, V. S. (2019). Artificial intelligence for assisting diagnostics and assessment of parkinson's disease—a review. *Clinical neurology and neurosurgery*, 184:105442.

[2] Benjamin, D. J., Berger, J. O., Johannesson, M., Nosek, B. A., Wagenmakers, E.-J., Berk, R., Bollen, K. A., Brembs, B., Brown, L., Camerer, C., et al. (2018). Redefine statistical significance. *Nature human behaviour*, 2(1):6–10.

[3] Berrendero, J. R., Cuevas, A., and Torrecilla, J. L. (2016). Variable selection in functional data classification: a maxima-hunting proposal. *Statistica Sinica*, pages 619–638.

[4] Bittencourt, N. F., Meeuwisse, W., Mendonça, L., Nettel-Aguirre, A., Ocarino, J., and Fonseca, S. (2016). Complex systems approach for sports injuries: moving from risk factor identification to injury pattern recognition—narrative review and new concept. *British journal of sports medicine*, 50(21):1309–1314.

[5] Browne, R. H. (2010). The t-test p value and its relationship to the effect size and p (x¿ y). *The American Statistician*, 64(1):30–33.

[6] Buford, T. W., Roberts, M. D., and Church, T. S. (2013). Toward exercise as personalized medicine. *Sports medicine*, 43(3):157–165.

[7] Cardinale, M. and Varley, M. C. (2017). Wearable training-monitoring technology: applications, challenges, and opportunities. *International journal of sports physiology and performance*, 12(s2):S2–55.

[8] Carroll, C., Müller, H.-G., and Kneip, A. (2021). Cross-component registration for multivariate functional data, with application to growth curves. *Biometrics*, 77(3):839–851.

[9] Cederbaum, J., Scheipl, F., and Greven, S. (2018). Fast symmetric additive covariance smoothing. *Computational Statistics & Data Analysis*, 120:25–41.

[10] Ceyssens, L., Vanelderen, R., Barton, C., Malliaras, P., and Dingenen, B. (2019). Biomechanical risk factors associated with running-related injuries: a systematic review. *Sports medicine*, 49(7):1095–1115.

[11] Clermont, C. A., Duffett-Leger, L., Hettinga, B. A., and Ferber, R. (2020). Runners' perspectives on 'smart'wearable technology and its use for preventing injury. *International Journal of Human–Computer Interaction*, 36(1):31–40.

[12] Cui, E., Leroux, A., Smirnova, E., and Crainiceanu, C. M. (2021). Fast univariate inference for longitudinal functional models. *Journal of Computational and Graphical Statistics*, pages 1–12.

[13] Donoghue, O. A., Harrison, A. J., Coffey, N., and Hayes, K. (2008). Functional data analysis of running kinematics in chronic achilles tendon injury. *Medicine and science in sports and exercise*, 40(7):1323–1335.

[14] Dorn, T. W., Schache, A. G., and Pandy, M. G. (2012). Muscular strategy shift in human running: dependence of running speed on hip and ankle muscle performance. *Journal of Experimental Biology*, 215(11):1944–1956.

[15] Dunn, J., Runge, R., and Snyder, M. (2018). Wearables and the medical revolution. *Personalized medicine*, 15(5):429–448.

[16] Febrero Bande, M. and Oviedo de la Fuente, M. (2012). Statistical computing in functional data analysis: The r package fda. usc.

[17] Ferber, R., Osis, S. T., Hicks, J. L., and Delp, S. L. (2016). Gait biomechanics in the era of data science. *Journal of biomechanics*, 49(16):3759–3761.

[18] Halson, S. L. (2014). Monitoring training load to understand fatigue in athletes. *Sports medicine*, 44(2):139–147.

[19] Hemingway, B. S., Greig, L., Jovanovic, M., and Swinton, P. (2020). A narrative review of mathematical fitness-fatigue modelling for applications in exercise science: model dynamics, methods, limitations, and future recommendations.

[20] Horst, F., Lapuschkin, S., Samek, W., Müller, K.-R., and Schöllhorn, W. I. (2019). Explaining the unique nature of individual gait patterns with deep learning. *Scientific reports*, 9(1):1–13.

[21] Ibrahim, S. A. (2021). Artificial intelligence for disparities in knee pain assessment. *Nature Medicine*, 27(1):22–23.

[22] Jauhiainen, S., Pohl, A. J., Äyrämö, S., Kauppi, J.-P., and Ferber, R. (2020). A hierarchical cluster analysis to determine whether injured runners exhibit similar kinematic gait patterns. *Scandinavian journal of medicine & science in sports*, 30(4):732–740.

[23] Kalkhoven, J. T., Watsford, M. L., Coutts, A. J., Edwards, W. B., and Impellizzeri, F. M. (2021). Training load and injury: causal pathways and future directions. *Sports Medicine*, pages 1–14.

[24] Karas, M., Marinsek, N., Goldhahn, J., Foschini, L., Ramirez, E., and Clay, I. (2020). Predicting subjective recovery from lower limb surgery using consumer wearables. *Digital Biomarkers*, 4(1):73–86.

[25] Karas, M., Straczkiewicz, M., Fadel, W., Harezlak, J., Crainiceanu, C. M., and Urbanek, J. K. (2019). Adaptive empirical pattern transformation (ADEPT) with application to walking stride segmentation. *Biostatistics*. kxz033.

[26] Karas, M., Urbanek, J. K., Illiano, V. P., Bogaarts, G., Crainiceanu, C. M., and Dorn, J. F. (2021). Estimation of free-living walking cadence from wrist-worn sensor accelerometry data and its association with SF-36 quality of life scores. *Physiological Measurement*, 42(6):065006.

[27] Kokoszka, P. and Reimherr, M. (2017). *Introduction to functional data analysis*. CRC press.

[28] Koldenhoven, R. M. and Hertel, J. (2018). Validation of a wearable sensor for measuring running biomechanics. *Digital biomarkers*, 2(2):74–78.

[29] Kosorok, M. R. and Laber, E. B. (2019). Precision medicine. *Annual review of statistics and its application*, 6:263–286.

[30] Kowalski, R. G., Hammond, F. M., Weintraub, A. H., Nakase-Richardson, R., Zafonte, R. D., Whyte, J., and Giacino, J. T. (2021). Recovery of consciousness and functional outcome in moderate and severe traumatic brain injury. *JAMA neurology*.

[31] Lambert, M. I. and Borresen, J. (2010). Measuring training load in sports. *International journal of sports physiology and performance*, 5(3):406–411.

[32] Lencioni, T., Carpinella, I., Rabuffetti, M., Marzegan, A., and Ferrarin, M. (2019). Human kinematic, kinetic and emg data during different walking and stair ascending and descending tasks. *Scientific data*, 6(1):1–10.

[33] Li, H., Kozey Keadle, S., Staudenmayer, J., Assaad, H., Huang, J. Z., and Carroll, R. J. (2015). Methods to assess an exercise intervention trial based on 3-level functional data. *Biostatistics*, 16(4):754–771.

[34] Li, Y., Wang, N., and Carroll, R. J. (2013). Selecting the number of principal components in functional data. *Journal of the American Statistical Association*, 108(504):1284–1294.

[35] Malone, J. J., Lovell, R., Varley, M. C., and Coutts, A. J. (2017). Unpacking the black box: applications and considerations for using gps devices in sport. *International journal of sports physiology and performance*, 12(s2):S2–18.

[36] Matabuena, M. and Rodríguez-López, R. (2019). An improved version of the classical banister model to predict changes in physical condition. *Bulletin of Mathematical Biology*, 81(6):1867–1884.

[37] Messier, S. P., Legault, C., Schoenlank, C. R., Newman, J. J., Martin, D. F., and DeVita, P. (2008). Risk factors and mechanisms of knee injury in runners. *Medicine & Science in Sports & Exercise*, 40(11):1873–1879.

[38] Morris, M. E., Huxham, F., McGinley, J., Dodd, K., and Iansek, R. (2001). The biomechanics and motor control of gait in parkinson disease. *Clinical biomechanics*, 16(6):459–470.

[39] Müller, R. and Büttner, P. (1994). A critical discussion of intraclass correlation coefficients. *Statistics in medicine*, 13(23-24):2465–2476.

[40] Park, J., Brunel, N., and Chassat, P. (2022). Curvature and torsion estimation of 3d functional data: A geometric approach to build the mean shape under the frenet serret framework. *arXiv preprint arXiv:2203.02398*.

[41] Phinyomark, A., Osis, S., Hettinga, B. A., and Ferber, R. (2015). Kinematic gait patterns in healthy runners: A hierarchical cluster analysis. *Journal of biomechanics*, 48(14):3897–3904.

[42] Phinyomark, A., Petri, G., Ibáñez-Marcelo, E., Osis, S. T., and Ferber, R. (2018). Analysis of big data in gait biomechanics: Current trends and future directions. *Journal of medical and biological engineering*, 38(2):244–260.

[43] Piatrikova, E., Willsmer, N. J., Altini, M., Jovanović, M., Mitchell, L. J., Gonzalez, J. T., Sousa, A. C., and Williams, S. (2021). Monitoring the heart rate variability responses to training loads in competitive swimmers using a smartphone application and the banister impulse-response model. *International Journal of Sports Physiology and Performance*, 1(aop):1–9.

[44] Pini, A., Markström, J. L., and Schelin, L. (2019). Test–retest reliability measures for curve data: An overview with recommendations and supplementary code. *Sports biomechanics*, pages 1–22.

[45] Pini, A. and Vantini, S. (2017). Interval-wise testing for functional data. *Journal of Nonparametric Statistics*, 29(2):407–424.

[46] Poß, D., Liebl, D., Kneip, A., Eisenbarth, H., Wager, T. D., and Barrett, L. F. (2020). Superconsistent estimation of points of impact in non-parametric regression with functional predictors. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 82(4):1115–1140.

[47] Riazati, S., Caplan, N., Matabuena, M., and Hayes, P. R. (2020). Fatigue induced changes in muscle strength and gait following two different intensity, energy expenditure matched runs. *Frontiers in Bioengineering and Biotechnology*, 8:360.

[48] Rossi, A., Pappalardo, L., Cintia, P., Iaia, F. M., Fernández, J., and Medina, D. (2018). Effective injury forecasting in soccer with gps training data and machine learning. *PloS one*, 13(7):e0201264.

[49] Selvin, E., Crainiceanu, C. M., Brancati, F. L., and Coresh, J. (2007). Short-term variability in measures of glycemia and implications for the classification of diabetes. *Archives of internal medicine*, 167(14):1545–1551.

[50] Sergazinov, R., Leroux, A., Cui, E., Crainiceanu, C., Aurora, R. N., Punjabi, N. M., and Gaynanova, I. (2022). A case study of glucose levels during sleep using fast function on scalar regression inference. *arXiv preprint arXiv:2205.08439*.

[51] Shang, H. L. (2014). A survey of functional principal component analysis. *AStA Advances in Statistical Analysis*, 98(2):121–142.

[52] Shou, H., Eloyan, A., Lee, S., Zipunnikov, V., Crainiceanu, A., Nebel, M., Caffo, B., Lindquist, M., and Crainiceanu, C. M. (2013). Quantifying the reliability of image replication studies: the image intraclass correlation coefficient (i2c2). *Cognitive, Affective, & Behavioral Neuroscience*, 13(4):714–724.

[53] Shou, H., Zipunnikov, V., Crainiceanu, C. M., and Greven, S. (2015). Structured functional principal component analysis. *Biometrics*, 71(1):247–257.

[54] Straczkiewicz, M., James, P., and Onnela, J.-P. (2021). A systematic review of smartphone-based human activity recognition methods for health research. *NPJ Digital Medicine*, 4(1):1–15.

[55] Uhlrich, S. D., Kolesar, J. A., Kidziński, Ł., Boswell, M. A., Silder, A., Gold, G. E., Delp, S. L., and Beaupre, G. S. (2020). Personalization improves the biomechanical efficacy of foot progression angle modifications in individuals with medial knee osteoarthritis. *medRxiv*.

[56] Ullah, S. and Finch, C. F. (2013). Applications of functional data analysis: A systematic review. *BMC medical research methodology*, 13(1):1–12.

[57] Van Gent, R., Siem, D., van Middelkoop, M., Van Os, A., Bierma-Zeinstra, S., and Koes, B. (2007). Incidence and determinants of lower extremity running injuries in long distance runners: a systematic review. *British journal of sports medicine*, 41(8):469–480.

[58] Van Gheluwe, B., Kirby, K. A., Roosen, P., and Phillips, R. D. (2002). Reliability and accuracy of biomechanical measurements of the lower extremities. *Journal of the American Podiatric Medical Association*, 92(6):317–326.

[59] Van Hooren, B., Goudsmit, J., Restrepo, J., and Vos, S. (2020). Real-time feedback by wearables in running: Current approaches, challenges and suggestions for improvements. *Journal of sports sciences*, 38(2):214–230.

[60] Volkmann, A., Stöcker, A., Scheipl, F., and Greven, S. (2021). Multivariate functional additive mixed models. *Statistical Modelling*, page 1471082X211056158.

[61] Vovk, V. and Wang, R. (2019). E-values: Calibration, combination, and applications. *Forthcoming in the Annals of Statistics*.

[62] Warmenhoven, J., Bargary, N., Liebl, D., Harrison, A., Robinson, M., Gunning, E., and Hooker, G. (2020). Pca of waveforms and functional pca: A primer for biomechanics. *Journal of Biomechanics*, page 110106.

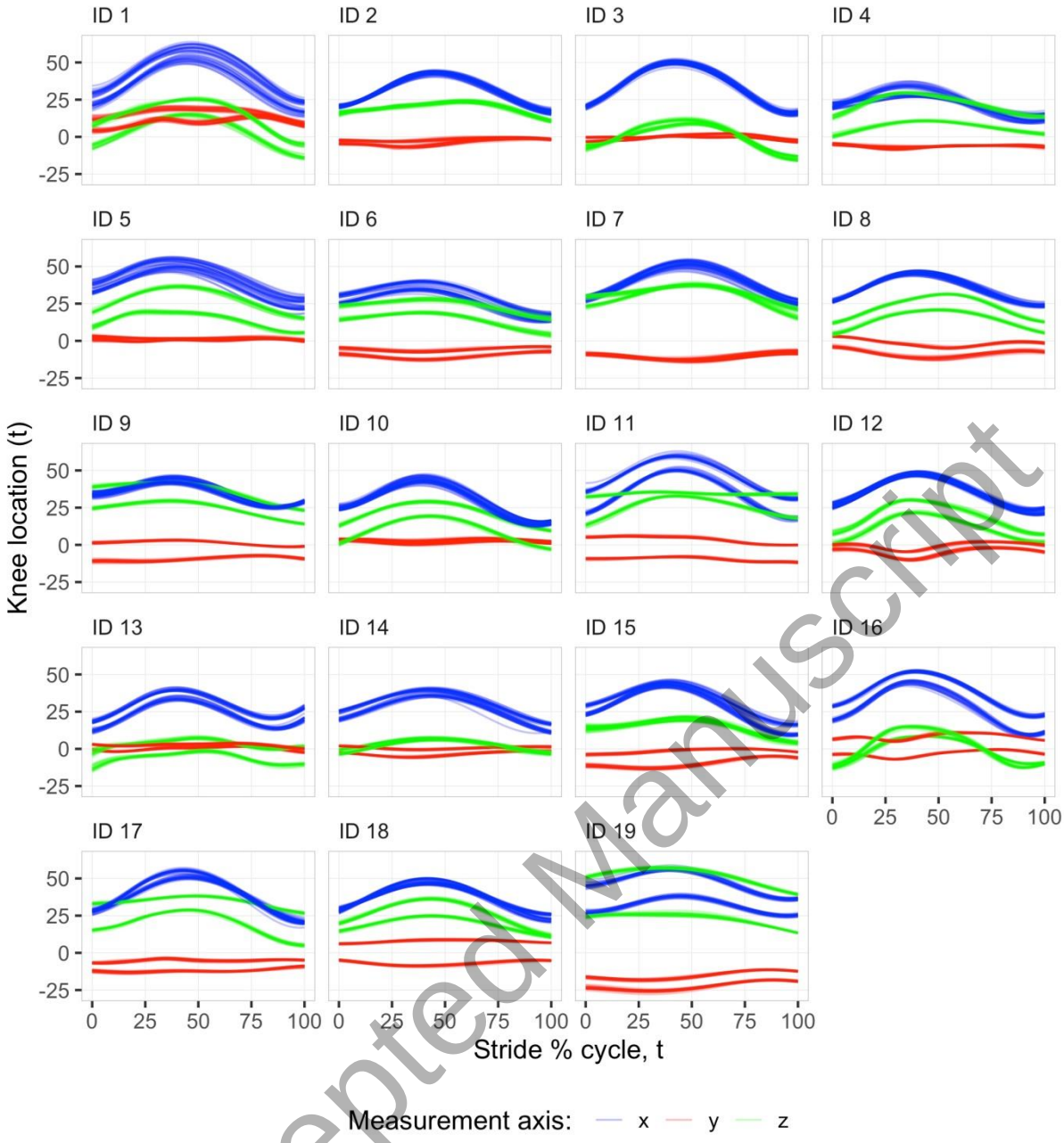

Fig. 1 Knee location trajectories in recorded along 20 strides in two HIIT sessions for 19 participants. Each plot shows data for one participant from both HIIT sessions. Measurements from the same axis are marked with the same color.

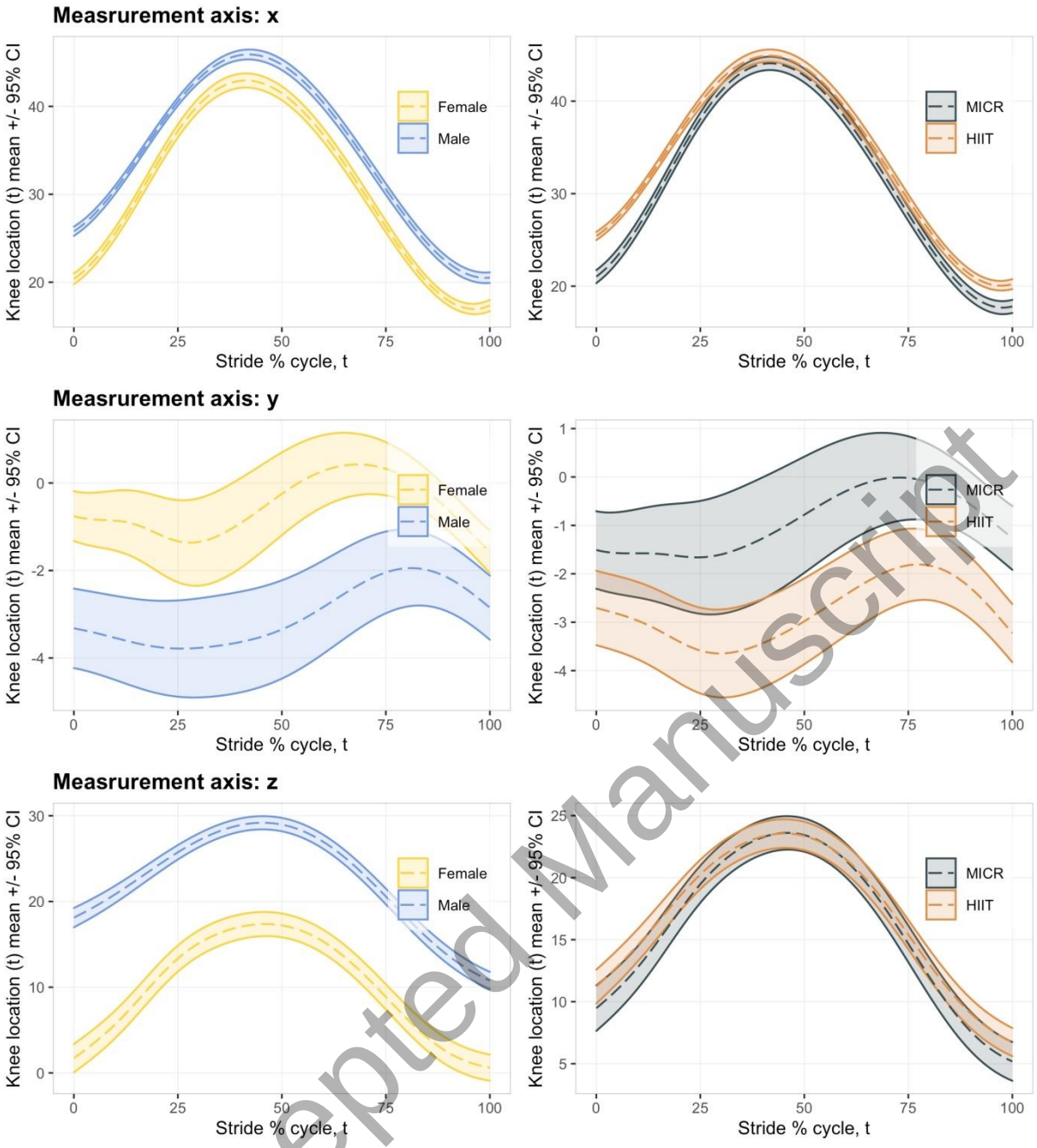

Fig. 2 Sample mean and ± 95% confidence intervals (CIs) of the mean computed point-wise from knee location data collected from 20 running strides per individual during one HIIT and one MICR session. Sample statistics were computed separately for data collected from each measurement axis (*x*, *y*, *z*), in strata according to gender (female, male) and race type (MICR, HIIT).

Fig. 3 Scatterplots of functional scores 1 and 2 of level 1, 2 and 3, calculated with N3 multilevel models. Each horizontal plots panel corresponds to one measurement axis (*x*, *y*, *z*). Point colors denote participant ID. Point shapes denote participant gender ("F" for female, "M" for male).

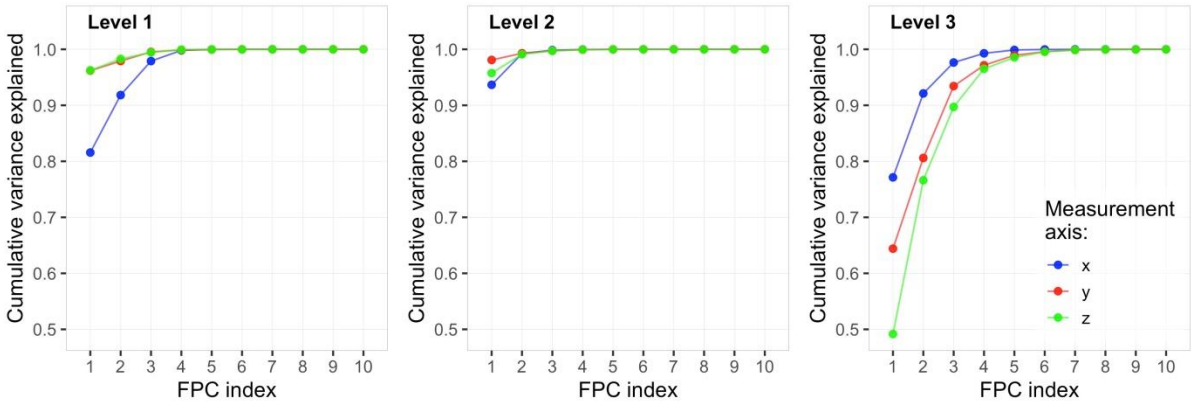

Fig. 4 Cumulative variance explained by subsequent functional principal components of level 1, 2 and 3, calculated with N3 multilevel models. Results are color-coded, with three colors representing results from separate measurement axis-specific models (*x*, *y*, *z*).

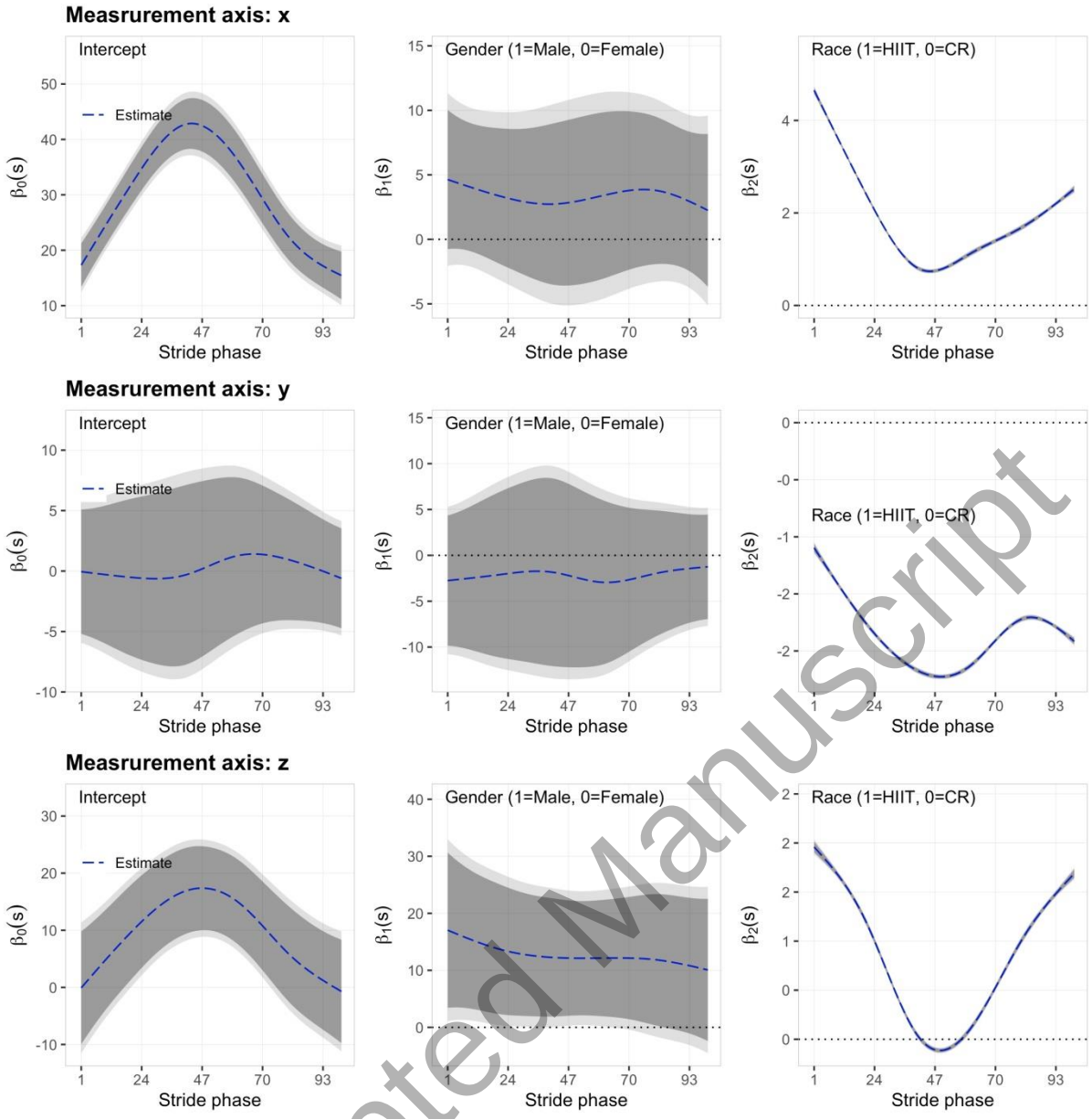

Fig. 5 Fixed effects estimates (dashed blue line), 95% point-wise confidence intervals (dark gray shaded area), and 95% joint confidence intervals (light gray shaded area) in multilevel functional regression with biomechanical profiles set as functional response, covariates for gender (coded 1 for male, 0 for female) and race type (coded 1 for HIIT, 0 for CR) and participant-specific random functions for race type (coded 1 for HIIT, 0 for CR).

Table 1 Structured functional models. For nested models, $i = 1, 2, \dots, n$; $j = 1, 2, \dots, n_i$; $k = 1, 2, \dots, K_{ij}$; $i_1 = 1, 2, \dots, I_1$, $i_2 = 1, 2, \dots, I_{2i1}, \dots, i_r = 1, 2, \dots, i_{r_{i_1}, i_2, \dots, i_{r_1}}$. For crossed designs, $i = 1, 2, \dots, n$; $j = 1, 2, \dots, J$; $k = 1, 2, \dots, n_{ij}$; (C2s) "Two-way sub" stands for "Two-way crossed design with subsampling"; (CM) contains combinations of anys $(s = 1, 2, \dots, r)$ subset of the latent processes, as well as repeated measurements within each cell. $S_1, S_2, \dots, S_d$ $\in \{i_{k_1} i_{k_2} \dots, i_{k_s}, u : k_1, k_2, \dots, k_s \in (1, 2, \dots, r), u \in (\varnothing, 1, 2, \dots, I_{i_1 i_2, \dots, i_r}), s \leq r\}$, u is the index for repeated observation in cell $(i_{k1}, i_{k2}, \dots, i_{kr})$. $\epsilon(t)$ is a random error $N(0, \sigma^2)$.

	Model	Model formula
Nested	(N1) One-way	$X^{i}(t) = \mu(t) + Z^{i}(t) + \epsilon^{i}(t)$
	(N2) Two-way	$X^{i,j}(t) = \mu(t) + Z^{i}(t) + W^{i,j} + \epsilon^{i,j}(t)$
	(N3) Three-way	$X^{i,j,k}(t) = \mu(t) + Z^{i}(t) + W^{i,j} + U^{i,j,k} + \epsilon^{i,j,k}(t)$
	(NM) Multi-way	$X^{i_1,i_2,\dots,i_r}(t) = \mu(t) + R^{i_1}_{(1)}(t) + R^{i_2}_{(2)} + \dots + R^{i_r}_{(r)} + \epsilon^{i_1,i_2,\dots,i_r}(t)$
Crossed	(C2) Two-way	$X^{i,j}(t) = \mu(t) + \eta^{j}(t) + Z^{i}(t) + W^{i,j} + \epsilon^{i,j}(t)$
	(C2s) Two-way sub	$X^{i,j,k}(t) = \mu(t) + \eta^{j} + Z^{i}(t) + W^{i,j} + U^{i,j,k} + \epsilon^{i,j,k}(t)$
	(CM) Multi-way	$X^{i_1,i_2,\dots,i_r u}(t) = \mu(t) + R^{S_i}(t) + R^{S_2} + \dots + R^{S_r} + \epsilon^{i_1,i_2,\dots,i_r u}(t)$

Table 2 Sample mean and 95% confidence intervals (CIs) of the mean for two scores (score 1, score 2) along the three hierarchical levels (level 1, level 2, level 3) considered in the three-level multilevel functional model, summarized separately for females and males.

	Axis	Score level	Score index	Score mean [95% CI] Female	Score mean [95% CI] Male
1	x	Level 1	Score 1	207.7 [183.7, 231.7]	239.0 [220.4, 257.6]
2	x	Level 1	Score 2	-8.4 [-13.1, -3.7]	-23.2 [-31.4, -14.9]
3	x	Level 2	Score 1	104.0 [88.6, 119.4]	121.3 [106.0, 136.6]
4	x	Level 2	Score 2	25.3 [20.9, 29.8]	23.0 [19.1, 26.8]
5	x	Level 3	Score 1	5.0 [4.1, 6.0]	5.8 [4.7, 6.9]
6	x	Level 3	Score 2	1.8 [1.4, 2.3]	2.0 [1.6, 2.5]
7	y	Level 1	Score 1	-7.3 [-19.9, 5.3]	-30.2 [-64.6, 4.2]
8	y	Level 1	Score 2	-1.0 [-4.6, 2.6]	-0.1 [-3.1, 3.0]
9	y	Level 2	Score 1	-3.8 [-18.8, 11.2]	-15.3 [-36.0, 5.4]
10	y	Level 2	Score 2	0.7 [-1.1, 2.6]	0.6 [-1.8, 3.1]
11	y	Level 3	Score 1	-0.2 [-0.5, 0.2]	-0.8 [-1.3, -0.2]
12	y	Level 3	Score 2	0.1 [-0.1, 0.2]	0.0 [-0.2, 0.3]
13	z	Level 1	Score 1	68.1 [22.0, 114.2]	149.2 [114.8, 183.5]
14	z	Level 1	Score 2	36.6 [27.0, 46.1]	44.8 [40.1, 49.6]
15	z	Level 2	Score 1	36.0 [11.7, 60.2]	76.5 [46.0, 107.0]
16	z	Level 2	Score 2	16.1 [10.1, 22.2]	15.6 [10.7, 20.5]
17	z	Level 3	Score 1	1.6 [1.1, 2.2]	3.4 [2.9, 4.0]
18	z	Level 3	Score 2	0.8 [0.4, 1.2]	1.8 [1.4, 2.2]