A Multivariate Multilevel Longitudinal Functional Model for Repeatedly Observed Human Movement Data

Edward Gunning[1] & Steven Golovkine[1] & Andrew J. Simpkin[2] & Aoife Burke[3] & Sarah Dillon[4] & Shane Gore[3] & Kieran Moran[3] & Siobhan O'Connor[3] & Enda Whyte[3] & Norma Bargary[1]

[1] *MACSI, Department of Mathematics and Statistics, University of Limerick, Ireland, firstname.lastname@ul.ie*
[2] *School of Mathematical and Statistical Sciences, University of Galway, Ireland, andrew.simpkin@universityofgalway.ie*
[3] *Centre for Injury Prevention and Performance, Athletic Therapy and Training; School of Health and Human Performance, Dublin City University, Dublin, Ireland, firstname.lastname@dcu.ie*
[4] *School of Allied Health, Faculty of Education and Health Science, University of Limerick, Limerick, Ireland, sarah.dillon@ul.ie*

Résumé. La recherche en biomécanique et en mouvement humain implique souvent la mesure régulière de plusieurs variables cinématiques ou cinétiques tout au long d'un mouvement, ce qui produit des données qui se présentent sous la forme de courbes, lisses, multivariées et variables dans le temps. Ces données se prêtent naturellement à l'analyse des données fonctionnelles. De plus, il est courant d'enregistrer le même mouvement de manière répétée pour chaque individu, ce qui permet d'obtenir des courbes corrélées pouvant être considérées comme des données fonctionnelles longitudinales.

Nous présentons une nouvelle approche de modélisation de données fonctionnelles longitudinales multivariées et hiérarchiques, en l'appliquant à des données cinématiques de coureurs amateurs recueillies lors d'une course sur tapis roulant. Pour chaque foulée, les angles de la hanche, des genoux et des chevilles des coureurs sont modélisés conjointement comme des fonctions multivariées dépendantes de covariables spécifiques au sujet. Des effets aléatoires fonctionnels multivariés variant longitudinalement sont utilisés pour capturer la dépendance entre les foulées adjacentes et les changements dans les fonctions multivariées au cours de la course. Nous représentons chaque observation en la décomposant dans une base en composantes principales fonctionnelles multivariées et nous modélisons les coefficients de base grâce des modèles scalaires longitudinaux à effets mixtes. Les effets aléatoires prédits sont utilisés pour comprendre et visualiser les changements dans les données fonctionnelles multivariées au cours de la course.

Dans notre application, cette méthode nous permet de quantifier les effets des covariables scalaires sur les données fonctionnelles multivariées. Il en résulte un effet statistiquement significatif de la vitesse de course au niveau des articulations de la hanche, du genou et de la cheville. L'analyse des effets aléatoires prédits révèle que la cinématique des individus est généralement stable, mais certains individus présentent de fortes variations au cours de la course.

Mots-clés. Analyse de données fonctionnelles longitudinales, Donnéees fonctionnelles multivariées, Analyse cinématique, Modèle mixte.

Abstract. Biomechanics and human movement research often involves measuring multiple kinematic or kinetic variables regularly throughout a movement, yielding data that present as smooth, multivariate, time-varying curves and are naturally amenable to functional data analysis. It is now increasingly common to record the same movement repeatedly for each individual, resulting in curves that are serially correlated and can be viewed as longitudinal functional data.

We present a new approach for modelling multivariate multilevel longitudinal functional data, with application to kinematic data from recreational runners collected during a treadmill run. For each stride, the runners' hip, knee and ankle angles are modelled jointly as smooth multivariate functions that depend on subject-specific covariates. Longitudinally varying multivariate functional random effects are used to capture the dependence among adjacent strides and changes in the multivariate functions over the course of the treadmill run. We represent each observation using a multivariate functional principal components basis and model the basis coefficients using scalar longitudinal mixed effects models. The predicted random effects are used to understand and visualise changes in the multivariate functional data over the course of the treadmill run.

In our application, the method quantifies the effects of scalar covariates on the multivariate functional data, revealing a statistically significant effect of running speed at the hip, knee and ankle joints. Analysis of the predicted random effects reveals that individuals' kinematics are generally stable but certain individuals who exhibit strong changes during the run can also be identified.

Keywords. Longitudinal functional data analysis, Multivariate functional data, Kinematic analysis, Mixed-effects model

1 Introduction

Longitudinal functional data analysis (LFDA) concerns the modelling of the dependence among functions due to correlation over a longer (or different) timescale than the one on which they are measured. Examples include daily activity functions measured consecutively for a number of days for several subjects (Goldsmith *et al.*, 2015) or brain imaging profiles of patients measured at several hospital visits (Greven *et al.*, 2010), see Park and Staicu (2015).

Our motivating dataset comes from the Dublin City University running injury surveillance (RISC) study, where kinematic data from recreational runners were captured during a treadmill run with the goal of understanding running technique and its link to injury. We focus on modelling the sagittal plane hip, knee and ankle angles because the majority of running-related injuries occur in the lower limbs. During the treadmill run, the kinematic data were recorded for a large number of consecutive strides for each individual (see Figure 1). They were then segmented into individual strides, as a single stride is considered the

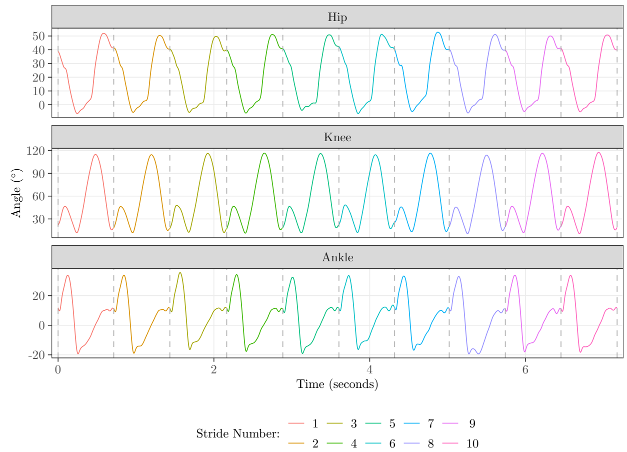


Figure 1: The right sagittal hip, knee and ankle angles of a single participant in the RISC dataset for the first ten strides of their treadmill run. The dashed vertical lines indicate touch down (i.e., when the foot first touches the ground), which represents the start and end of each stride.

most basic unit of analysis.

In this work, we want to employ a multivariate approach to capture the dependence among multiple joints (i.e., the hip, knee and ankle angles), rather than performing separate univariate analyses for each location. From an applied perspective, understanding the dependence (or co-ordination) among multiple joints is crucial for fully describing movement patterns (Glazier, 2021). Moreover, the participants were measured on both sides of the body, which adds a hierarchical structure to the data. We also need to include scalar covariate information in our model, e.g., sex, running speed and injury status. This motivates the development of a *multivariate multilevel longitudinal functional model*. The dataset contains more than 40000 multivariate functional observations from 284 unique individuals. To the best of our knowledge, this is the first piece of work to develop statistical methodology to appropriately analyse repeatedly observed multivariate kinematic data in human movement biomechanics.

2 Model

We denote the multivariate functional observation from the lth stride for the ith individual on side j as

$$\mathbf{y}_{ijl}(t) = \left(y_{ijl}^{(hip)}(t), y_{ijl}^{(knee)}(t), y_{ijl}^{(ankle)}(t)\right)^\top, l = 1, \dots n_{ij}, j \in \{\text{left, right}\} \text{ and } i = 1, \dots, N,$$

where N is the total number of individuals, n_{ij} is the total number of strides taken by individual i on side j, and $t \in [0, 100]$ is a normalised *functional time* interval with 0 representing the start of a stride and 100(%) representing the end. We also introduce a (normalised) *longitudinal time* variable $T \in [0, 1]$, such that T_{ijl} indexes the time in the treadmill run at which stride l occurs on side j for subject i. The start of the treadmill run is at $T = 0$ and the end at $T = 1$. Finally, we let $\mathbf{x}_{ij} = (x_{ij1}, \dots, x_{ijA})^\top$ denote the scalar covariates for subject i on side j. The covariates could be subject specific (e.g., sex, height) or subject-and-side specific (e.g., an indicator for a subject's dominant side). We assume that the covariates are fixed across strides and hence $\mathbf{x}_{ij}$ is not indexed by l.

Our proposed multivariate multilevel longitudinal functional model is

$$\mathbf{y}_{ijl}(t) = \boldsymbol{\beta}_0(t, T_{ijl}) + \sum_{a=1}^{A} x_{ija}\boldsymbol{\beta}_a(t) + \mathbf{u}_i(t, T_{ijl}) + \mathbf{v}_{ij}(t, T_{ijl}) + \boldsymbol{\varepsilon}_{ijl}(t).$$

where $\boldsymbol{\beta}_0(t, T_{ijl})$, $\mathbf{u}_i(t, T_{ijl})$ and $\mathbf{v}_{ij}(t, T_{ijl})$ are the multivariate intercept function, the subject-specific and the subject and side-specific multivariate functional random intercepts that varies smoothly in both functional and longitudinal time, $\boldsymbol{\beta}_a(t)$ is the multivariate functional fixed effect corresponding to the ath scalar covariate and $\boldsymbol{\varepsilon}_{ijl}(t)$ is the smooth multivariate functional random error that is specific to observation $\mathbf{y}_{ijl}(t)$. We assume that $\mathbf{y}_{ijl}(t)$ are centered.

The intercept function $\boldsymbol{\beta}_0(t, T)$ is assumed to be a smooth bivariate function of both functional time t and longitudinal time T. For $a = 1, \dots, A$, the fixed effect $\boldsymbol{\beta}_a(t)$ captures the influence of the ath scalar covariate on the expected level and shape of the response (Bauer *et al.*, 2018). We assume that the fixed effects are constant across T, which implies that the scalar covariates affect the average running kinematics, rather than the kinematics at any particular point in the treadmill run. For $i = 1, \dots, N$, the subject-specific random intercept $\mathbf{u}_i(t, T)$ captures correlation among observations from the same subject and the subject-and-side-specific random intercepts $\mathbf{v}_{ij}(t, T)$ capture correlation among observations from the same subject and side. These functions are assumed to be independent realisations of mean-zero multivariate Gaussian processes with matrix-valued covariance function $\mathbf{Q}(t, t', T, T')$ and $\mathbf{R}(t, t', T, T')$. Finally, the random errors are assumed to be independent realisations of a zero-mean multivariate Gaussian process with matrix-valued covariance function $\mathbf{S}(t, t')$. The multivariate functional random error represents the deviation that is specific to observation $\mathbf{y}_{ijl}(t)$. It is further assumed that the processes $\mathbf{u}_i(t, T)$, $\mathbf{v}_{ij}(t, T)$ and $\boldsymbol{\varepsilon}_{ijl}(t)$ are mutually uncorrelated.

2.1 Basis Representation of the Multivariate Functions

We first represent each multivariate functional observation by a basis expansion

$$\mathbf{y}_{ijl}(t) = \sum_{k=1}^{K} y^\star_{ijl,k}\boldsymbol{\psi}_k(t).$$

The basis functions $\{\boldsymbol{\psi}_k(t)\}_{k=1}^K$ are multivariate functions and $y^\star_{ijl,k}$ are scalar basis coefficients that weight the basis functions to produce the functional observations. The functions $\{\boldsymbol{\psi}_k(t)\}_{k=1}^K$ are estimated using a multivariate functional principal components analysis (MFPCA, Happ and Greven, 2018). We choose K, the number of functions to retain, such that a high percentage (e.g., 99.5%) of the variance in the data is explained. This allows the basis coefficients to be treated as transformed data rather than estimated parameters and modelled in place of the observed multivariate functions (Morris *et al.*, 2011).

2.2 Modelling the Basis Coefficients

We model the matrix $\mathbf{Y}^\star$ of basis coefficients in place of the observed multivariate functional data. We make the simplifying assumption that each of the K basis coefficients (i.e., each column of $\mathbf{Y}^\star$) can be modelled separately (Morris and Carroll, 2006). The model for the kth basis coefficient is

$$y^\star_{ijl,k} = \beta^\star_{0,k}(T_{ijl}) + \sum_{a=1}^{A} x_{ia}\beta^\star_{a,k} + u^\star_{i,k}(T_{ijl}) + v^\star_{ij,k}(T_{ijl}) + \varepsilon^\star_{ijl,k}, \tag{1}$$

which is a multilevel functional model in longitudinal time T (Di *et al.*, 2009). We choose to parameterise the longitudinally varying functions using a small number of unpenalised basis functions, because changes are expected to be smooth and simple. We also use the same set of basis functions $\{\xi_d(T)\}_{d=1}^{D}$ to represent each longitudinally varying term, giving

$$\beta^\star_{0,k}(T) = \sum_{d=1}^{D} \beta^\star_{0,k,d}\ \xi_d(T), \quad u^\star_{i,k}(T) = \sum_{d=1}^{D} u^\star_{i,k,d}\ \xi_d(T) \quad \text{and} \quad v^\star_{ij,k}(T) = \sum_{d=1}^{D} v^\star_{ij,k,d}\ \xi_d(T).$$

We use a small number of natural cubic B-spline basis functions to represent each term. Substituting the basis function evaluations into model (1) gives, for the kth basis coefficient, the model

$$y^\star_{ijl,k} = \sum_{d=1}^{D} \beta^\star_{0,k,d}\ \xi_d(T_{ijl}) + \sum_{a=1}^{A} x_{ia}\beta^\star_{a,k} + \sum_{d=1}^{D} u^\star_{i,k,d}\ \xi_d(T_{ijl}) + \sum_{d=1}^{D} v^\star_{ij,k,d}\ \xi_d(T_{ijl}) + \varepsilon^\star_{ijl,k},$$

where $(u^*_{i,k,1}, \ldots, u^*_{i,k,D})^\top \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}^*_k)$, $(v^*_{ij,k,1}, \ldots, v^*_{ij,k,D})^\top \sim \mathcal{N}(\mathbf{0}, \mathbf{R}^*_k)$ and $\varepsilon^*_{ijl,k} \sim \mathcal{N}(0, s_k)$. This is a scalar linear mixed effects model. The matrices $\mathbf{Q}^*_k$ and $\mathbf{R}^*_k$ are of dimension $D \times D$ and contain $D(D+1)/2$ free parameters to estimate.

2.3 Reconstructing the Model Terms

2.3.1 Fixed Effects

Rather than inspect individual parameter estimates, it is more natural to combine the estimated parameters across the basis coefficients to reconstruct and estimate the functional model terms. The estimated intercept function is given by

$$\widehat{\boldsymbol{\beta}}_0(t,T) = \sum_{k=1}^{K}\sum_{d=1}^{D} \widehat{\beta}^\star_{0,k,d}\ \xi_d(T)\boldsymbol{\psi}_k(t),$$

where $\widehat{\beta}^\star_{0,k,d}$ denotes the estimate of $\beta^\star_{0,k,d}$ from the mixed effects model. Likewise, the estimate of the functional fixed effect of the ath scalar covariate is given by

$$\widehat{\boldsymbol{\beta}}_a(t) = \sum_{k=1}^{K} \widehat{\beta}^\star_{a,k}\boldsymbol{\psi}_k(t), \quad a = 1\ldots, A.$$

The estimates of $\widehat{\mathrm{Var}}(\widehat{\beta}^{\star}_{a,k})$ from the mixed effects model can be combined across k to construct approximate pointwise and simultaneous confidence bands for $\boldsymbol{\beta}_a(t)$, as described in (Gunning *et al.*, 2023).

2.3.2 Covariance Structures

The matrix-valued covariance functions $\mathbf{Q}(t, t', T, T')$ and $\mathbf{R}(t, t', T, T')$ implied by the model are

$$\begin{aligned}\mathbf{Q}(t, t', T, T') &= \mathbb{E}[\mathbf{u}_i(t, T)\ \mathbf{u}_i(t', T')^\top] = \boldsymbol{\Psi}(t)^\top(\mathbb{I}_K \otimes \boldsymbol{\xi}(T))^\top \mathbf{Q}^*(\mathbb{I}_K \otimes \boldsymbol{\xi}(T'))\boldsymbol{\Psi}(t'),\\ \mathbf{R}(t, t', T, T') &= \mathbb{E}[\mathbf{v}_{ij}(t, T)\ \mathbf{v}_{ij}(t', T')^\top] = \boldsymbol{\Psi}(t)^\top(\mathbb{I}_K \otimes \boldsymbol{\xi}(T))^\top \mathbf{R}^*(\mathbb{I}_K \otimes \boldsymbol{\xi}(T'))\boldsymbol{\Psi}(t'),\end{aligned}$$

where $\boldsymbol{\Psi}(t)$ is the $K \times 3$ matrix containing the basis functions, $\boldsymbol{\xi}(T) = (\xi_1(T), \dots, \xi_D(T))^\top$, $\mathbf{Q}^*$ (resp. $\mathbf{R}^*$) is the block-diagonal matrix containing the matrices $\mathbf{Q}^*_1, \dots, \mathbf{Q}^*_K$ (resp. $\mathbf{R}^*_1, \dots, \mathbf{R}^*_K$) along its diagonal. Finally, the within-function covariance is

$$\mathbf{S}(t, t') = \boldsymbol{\Psi}(t)^\top \mathbf{S}^* \boldsymbol{\Psi}(t'), \quad \mathbf{S}^* = \mathrm{diag}\{s_1, \dots, s_K\}.$$

2.3.3 Individual Trajectories

Our methodology facilitates the prediction of subject-specific and subject-and-side-specific trajectories at any point in the treadmill run. The prediction of the subject-specific multivariate functional random intercept at any $T \in [0, 1]$ is given by

$$\widehat{\mathbf{u}}_i(t, T) = \sum_{k=1}^{K}\sum_{d=1}^{D} \widehat{u}^{\star}_{i,k,d}\ \xi_d(T)\boldsymbol{\psi}_k(t), \quad i = 1, \dots, N,$$

where $\widehat{u}^{\star}_{i,k,d}$ is the Best Linear Unbiased Predictor (BLUP) of $u^{\star}_{i,k,d}$ from the linear mixed effects model. The subject-and-side specific deviation is obtained analogously as

$$\widehat{\mathbf{u}}_i(t, T) + \widehat{\mathbf{v}}_{ij}(t, T) = \sum_{k=1}^{K}\sum_{d=1}^{D}\big(\widehat{u}^{\star}_{i,k,d} + \widehat{v}^{\star}_{ij,k,d}\big)\xi_d(T)\boldsymbol{\psi}_k(t),\ i = 1, \dots, N,\ \text{and } j \in \{\text{left}, \text{right}\}.$$

The predicted trajectories can be used, for example, to investigate change in technique over the course of the treadmill run as measured by the rate of change with respect to T.

3 Application

3.1 Data Collection, Extraction and Preparation

Recreational runners aged between 18 and 64 years of age with no history of injury in the last three months were recruited as participants for the RISC study. Prior to the baseline testing session, in which the kinematic data were collected, the participants completed an online survey regarding their injury history, training history and demographics. They ran for three minutes at a self-selected speed that represented their typical training pace, while kinematic data were collected using a 17-camera, three-dimensional motion analysis system for the first full minute of the run. The motion data (i.e., marker trajectories) were sampled at a rate of 200Hz and filtered using a fourth-order zero-lag Butterworth filter at 15 Hz to smooth out observational errors. From the filtered trajectories, the sagittal plane hip, knee and ankle angles were extracted bilaterally for the first minute of the treadmill run.

The long sequences of kinematic measurements (e.g., Figure 1) were segmented into individual strides based on the initial contact of the foot with the ground, which was identified using a custom algorithm. The univariate functional data for each stride were time normalised and registered to the point of the maximum knee flexion angle, which is a clear and easily identifiable landmark in each stride. Within each dimension, 80 cubic B-spline basis functions were used to provide a near-lossless representation of the univariate functions. For each stride, the longitudinal time variable T was created based on the time at which that stride started, with $T = 0$ representing the start of the subject's capture period. This variable was normalised by dividing by the subject's maximum capture time, so that $T \in [0, 1]$. The MFPCA, computed from the univariate basis expansions, yielded $K = 27$ basis functions to explain 99.5% of the variance in the multivariate functional data.

3.2 Modelling Results

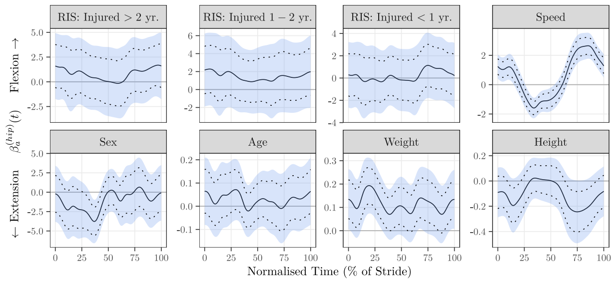
A constant function and four natural cubic B-splines were used as longitudinal basis functions, with unstructured $\mathbf{Q}_k^*$ and $\mathbf{R}_k^*$ matrices. We consider the following model:

$$\begin{aligned}\mathbf{y}_{ijl}(t) = \boldsymbol{\beta}_0(t, T_{ijl}) + \sum_{a=1}^{3} x_{ia}\boldsymbol{\beta}_a(t) + \text{ speed}_i \times \boldsymbol{\beta}_4(t) + \text{sex}_i \times \boldsymbol{\beta}_5(t) + \text{age}_i \times \boldsymbol{\beta}_6(t) \\ + \text{weight}_i \times \boldsymbol{\beta}_7(t) + \text{height}_i \times \boldsymbol{\beta}_8(t) + \mathbf{u}_i(t, T_{ijl}) + \mathbf{v}_{ij}(t) + \boldsymbol{\varepsilon}_{ijl}(t),\end{aligned}$$

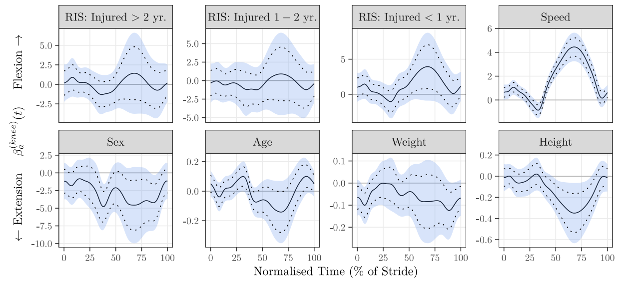
where x_{i1}, x_{i2} and x_{i3} are dummy-coded variables representing the "Injured more than 2 years ago", "Injured 1-2 years ago" and "Injured less than 1 year ago" categories of the retrospective injury status variable, where the reference category is "Never injured", speed_i is the self-selected running speed of subject i in km h^{-1}, sex_i is a dummy-coded variable for the sex of subject i (0 = male, 1 = female), age_i is the age of subject i in years, weight_i is the weight of subject i in kilograms and height_i is the height of subject i in centimetres. All numeric variables were centered to make the intercept function more interpretable.

3.2.1 Fixed Effects

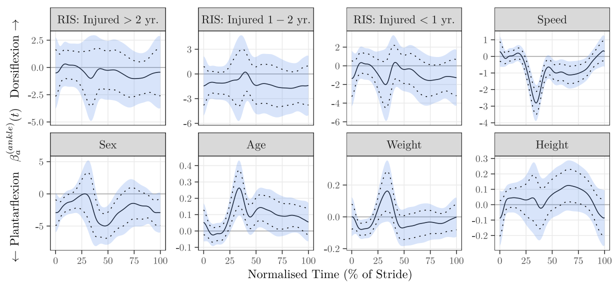
Analysis of the functional coefficients of the longitudinal basis functions used to model the intercept revealed that it was approximately constant in the longitudinal direction. Figure 2 displays the estimated coefficient functions that capture the effects of scalar covariates in our model. In all three dimensions, the simultaneous confidence bands for the retrospective injury status coefficient functions contain zero (solid grey horizontal line) for all t, indicating that



(a) Hip



(b) Knee



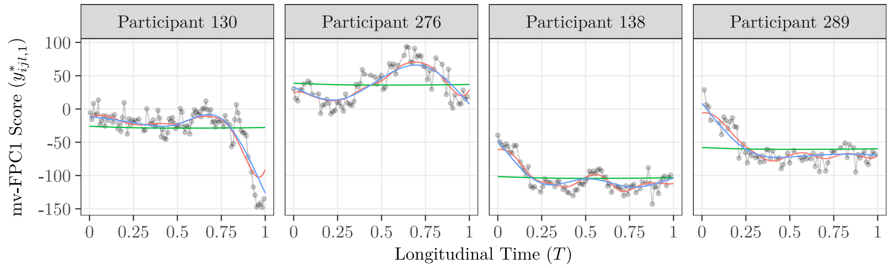
(c) Ankle

Figure 2: The estimated coefficient functions of the fixed effects from the fitted model. The black solid line represents the point estimate, the dotted black lines indicate pointwise 95% confidence intervals and the light blue ribbons represent 95% simultaneous confidence bands.

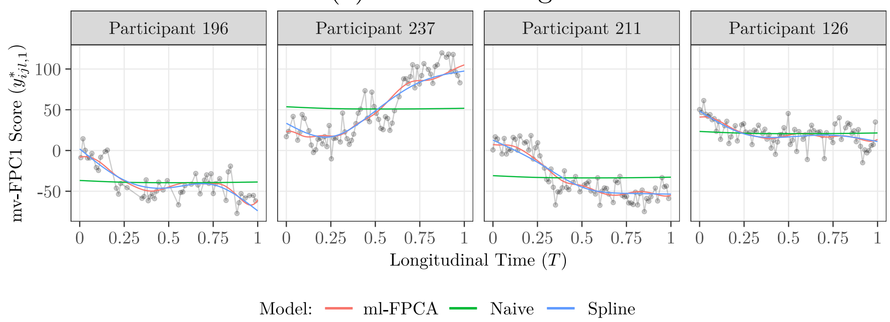
there is no evidence of a significant difference between any of the categories and the reference category of "Never injured". We observe a strong, noticeable effect of self-selected running speed in all three dimensions, as the simultaneous confidence bands only contain zero around the time that the point estimate crosses 0. Running at a higher speed is associated with greater hip flexion at initial contact and late in the swing phase ($t > 60\%$) and greater hip extension around the time of toe-off ($t \approx 38\%$), greater knee flexion which is most pronounced in the stance phase around the time of peak knee flexion angle ($t \approx 69\%$) and increased ankle plantarflexion which is most pronounced around the time of maximum plantarflexion ($t \approx 38\%$). The coefficient functions for the effect of sex are large in magnitude, reaching almost 5° in the knee and ankle. However, the corresponding confidence bands are wide and contain zero for almost all t, indicating a lot of uncertainty about this effect. There is limited evidence of an age, height or weight effect. Although the simultaneous confidence bands for these coefficient functions do not contain zero at certain points, the magnitude of each effect is small.

3.2.2 Random Effects

We present analysis of the fitted subject-and-side specific trajectories, which are obtained as BLUPs of the random effects. Figure 3 displays fits for subjects that were chosen according to summaries from the model. Firstly, we calculated the integrated squared first derivative with



(a) Rate of change



(b) Overall change

Figure 3: Observed and fitted values of the coefficients for subjects identified based on summaries from the model. **(a)** The top four subjects based on the integrated squared first derivative of their fitted trajectory with respect to longitudinal time. **(b)** The top four subjects based on the overall change during the treadmill run.

respect to longitudinal time of each subject's fitted profile, which provides a measure of the rate of change (or deviation from a constant fit) over the course of the treadmill run. Figure 3 (a) displays the first coefficients for the top four subjects ranked according to this metric. For ease of interpretation, we have only displayed the left side observations. All four subjects exhibit non-stationary patterns that are captured well by the longitudinal models. The naive model, which assumes that each individual's deviation is constant across longitudinal time, is inadequate. Figure 3 (b) displays another four subjects ranked according to the overall change in the subject's fitted profile over the course of the run, calculated as the absolute difference between the subjects' fitted profiles at $T = 0$ and $T = 1$. Non-stationary trends, which cannot be captured by the naive model, are evident again. It should be noted that these summaries were computed based on the full multivariate function but we have displayed the first coefficient. However, this coefficient captured the largest amount of variance in the longitudinal direction, so it is a reasonable choice.

As the coefficients in Figure 3 are a level of abstraction away from the multivariate functional data, we examine the fitted multivariate functions for a single individual. Based on Figure 3 (b), we choose to display Participant 237 because they exhibited a consistent, almost-linear evolution. Figure 4 (c) and (d) display the motion-capture animation at the time of peak knee flexion angle for this subject at the start (stride 1) and end (stride 80) of the treadmill run, respectively. The difference in the two pictures reflects the changes across longitudinal time, in particular the greater knee flexion at the end of the treadmill run.

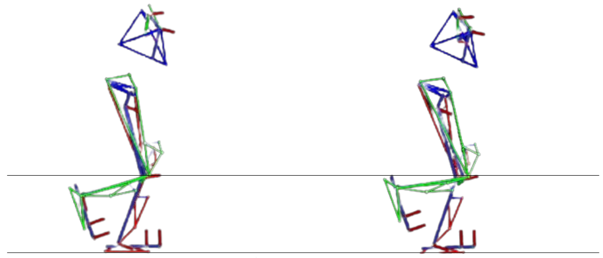


Figure 4: Individual analysis for Participant 237. **(Left)** The motion-capture animation for this subject at the time of peak knee flexion angle at the start of the treadmill run. **(Right)** The motion-capture animation for this subject at the time of peak knee flexion angle at the end of the treadmill run.

Bibliography

Bauer, A., Scheipl, F., Küchenhoff, H., & Gabriel, A.-A. (2018), An introduction to semi-parametric function-on-scalar regression. *Statistical Modelling*, 18(3-4), pp. 346-364.

Di, C.-Z., Crainiceanu, C. M., Caffo, B. S., & Punjabi, N. M. (2009), Multilevel functional principal component analysis. *The Annals of Applied Statistics*, 3(1), pp.458-488.

Happ, C., & Greven, S. (2018), Multivariate Functional Principal Component Analysis for Data Observed on Different (Dimensional) Domains. *Journal of the American Statistical Association*, 113(522), pp. 649-659.

Glazier, P. S. (2021), Beyond animated skeletons: How can biomechanical feedback be used to enhance sports performance? *Journal of Biomechanics*, 129, 110686.

Goldsmith, J., Zipunnikov, V., & Schrack, J. (2015), Generalized multilevel function-on-scalar regression and principal component analysis *Biometrics*, 71(2), pp. 344-353.

Greven, S., Crainiceanu, C. M., Caffo, B., & Reich, D. (2010), Longitudinal functional principal component analysis. *Electronic Journal of Statistics*, 4, pp. 1022-1054.

Gunning, E., Golovkine, S., Simpkin, A. J., Burke, A., Dillon, S., Gore, S., Moran, K. A., O'Connor, S., Whyte, & Bargary, N. (2023). Analyzing Kinematic Data from Recreational Runners using Functional Data Analysis. *Under Review.*

Morris, J. S., & Carroll, R. J. (2006), Wavelet-based functional mixed models. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 68(2), pp. 179-199.

Morris, J. S., Baladandayuthapani, V., Herrick, R. C., Sanna, P., & Gutstein, H. (2011), Automated analysis of quantitative image data using isomorphic functional mixed models, with application to proteomics data. *The Annals of Applied Statistics*, 5(2A), pp. 894-923.

Park, S. Y., & Staicu, A.-M. (2015), Longitudinal functional data analysis. *Stat*, 4(1), 212–226.