



# Smoothing spline analysis of variance models: A new tool for the analysis of cyclic biomechanical data

Nathaniel E. Helwig<sup>a,b,\*</sup>, K. Alex Shorter<sup>c</sup>, Ping Ma<sup>d</sup>, Elizabeth T. Hsiao-Wecksler<sup>e</sup>

<sup>a</sup> Department of Psychology, University of Minnesota, Minneapolis, MN 55455-0366, United States

<sup>b</sup> School of Statistics, University of Minnesota, Minneapolis, MN 55455-0493, United States

<sup>c</sup> Department of Mechanical Engineering, University of Michigan, Ann Arbor, MI, 48109-2125, United States

<sup>d</sup> Department of Statistics, University of Georgia, Athens, GA 30602-5029, United States

<sup>e</sup> Department of Mechanical Science and Engineering, University of Illinois, Urbana, IL 61801-2906, United States

## ARTICLE INFO

### Article history:

Accepted 31 July 2016

### Keywords:

Smoothing spline

Cyclic data

Gait analysis

Functional data analysis

## ABSTRACT

Cyclic biomechanical data are commonplace in orthopedic, rehabilitation, and sports research, where the goal is to understand and compare biomechanical differences between experimental conditions and/or subject populations. A common approach to analyzing cyclic biomechanical data involves averaging the biomechanical signals across cycle replications, and then comparing mean differences at specific points of the cycle. This pointwise analysis approach ignores the functional nature of the data, which can hinder one's ability to find subtle differences between experimental conditions and/or subject populations. To overcome this limitation, we propose using mixed-effects smoothing spline analysis of variance (SSANOVA) to analyze differences in cyclic biomechanical data. The SSANOVA framework makes it possible to decompose the estimated function into the portion that is common across groups (i.e., the average cycle, AC) and the portion that differs across groups (i.e., the contrast cycle, CC). By partitioning the signal in such a manner, we can obtain estimates of the CC differences (CCDs), which are the functions directly describing group differences in the cyclic biomechanical data. Using both simulated and experimental data, we illustrate the benefits of using SSANOVA models to analyze differences in noisy biomechanical (gait) signals collected from multiple locations (joints) of subjects participating in different experimental conditions. Using Bayesian confidence intervals, the SSANOVA results can be used in clinical and research settings to reliably quantify biomechanical differences between experimental conditions and/or subject populations.

© 2016 Elsevier Ltd. All rights reserved.

## 1. Introduction

Experimental data collected during cyclic motion are commonly used in the study of injury, orthopedic, rehabilitation, and sports biomechanics (Becker et al., 1995; DeVita et al., 1997; Griffin et al., 1995; Knoll et al., 2004; Risberg et al., 2009). In such studies, kinetic and kinematic time series data from limb and body movements are collected during a cyclic task, and, in many cases, several replications of the cyclic task are observed to obtain better estimates of the underlying biomechanical signals. The goal of an experimental biomechanics study is to investigate or compare differences in data collected from different experimental conditions and/or subject populations. These differences can be structural, physiological, or artificially created using physical constraints to modify behavior in a

controlled manner to investigate an aspect of human motion (Collins et al., 2005; Huang and Kuo, 2014). Key to estimating and comparing differences in the underlying behavior is the creation of an average cycle from the multiple task replications.

As a preliminary step, cyclic biomechanical data are often digitally filtered and then time-normalized to represent the behavior in terms of % cycle (Helwig et al., 2011). After time-normalization, cyclic biomechanical signals are frequently estimated by averaging data across task replications. Typically, average curves are used to qualitatively characterize motion, and single points at features of interest along the curve are used to quantitatively compare experimental conditions, such as minimum/maximum and mean differences between experimental conditions and/or subject populations (Becker et al., 1995; Chmielewski et al., 2005; DeVita et al., 1997; Diop et al., 2005; Griffin et al., 1995; Knoll et al., 2004; Risberg et al., 2009). Both the average data and cycle-to-cycle variability can be used to characterize behavior (Hausdorff et al., 2001; James and Stergiou, 2004; O'Connor et al., 2012; Romei et al., 2004; Huang and Kuo, 2014). However, current methods tend to reduce continuous

\* Corresponding author at: Department of Psychology, University of Minnesota, Minneapolis, MN 55455-0366, United States.

E-mail addresses: [helwig@umn.edu](mailto:helwig@umn.edu) (N.E. Helwig), [kshorter@umich.edu](mailto:kshorter@umich.edu) (K.A. Shorter), [pingma@uga.edu](mailto:pingma@uga.edu) (P. Ma), [ethw@illinois.edu](mailto:ethw@illinois.edu) (E.T. Hsiao-Wecksler).

time series data to discrete points for statistical comparison (DiBerardino et al., 2012; Shorter et al., 2008), which ignores the functional nature of the biomechanical data.

In this work, we propose to use functional differences instead of the traditional pointwise comparisons for the statistical analysis of biomechanical data. Specifically, we demonstrate the benefits of analyzing cyclic biomechanical data using a mixed-effects smoothing spline analysis of variance (SSANOVA) framework (Gu, 2013; Gu and Ma, 2005; Helwig, 2015; Wang, 1998a, 1998b; Zhang et al., 1998), which is a nonparametric extension of a linear mixed-effects regression (LMER) model (Bates et al., 2015). Note that LMER models are a popular statistical approach for analyzing correlated data, and have been proven useful for biomechanics research (e.g., Nimeskern et al., 2015). Furthermore, functional data analysis (FDA) methods have proven useful for normalizing, describing, and analyzing a variety of types of biomechanical, physiological, and behavioral time series data (e.g., Forner-Cordero et al., 2006; Lucero and Koenig, 2000; Page et al., 2006).

Our paper reveals how the LMER and FDA methodologies can be combined to examine functional differences in correlated data. The proposed framework can analyze differences in cyclic data comprising multiple task replications collected from subjects in different groups (experimental conditions or subject populations). The SSANOVA approach estimates the underlying mean functions (e.g., joint angle trajectories) for each subject and group, as well as the functions describing the mean differences between the groups. Consequently, the SSANOVA approach makes it possible to examine functional (instead of pointwise) mean differences in cyclic biomechanical data. We hypothesize that the SSANOVA will provide a powerful statistical framework for analyzing group differences in cyclic biomechanical data, and will offer clear benefits over the classic pointwise analysis approach.

## 2. Methods

Supplementary online material (SOM) provides detailed information about our proposed model and model fitting process. The SOM also includes analysis code and data to help researchers better understand, replicate, and extend/modify the work presented in this paper.

### 2.1. Mean difference methods

#### 2.1.1. Overview of problem

Suppose we have a periodic biomechanical signal (e.g., angle trajectory) collected over the course of a cycle (e.g., gait cycle) from  $S > 1$  subjects participating in  $G > 1$  experimental conditions. Also, suppose we have observed  $R_{sg} \geq 1$  replications of the cycle from the  $s$ -th subject in the  $g$ -th condition, and that we have linearly length normalized each cycle to consist of 101 time points representing 0–100% of the cycle (Helwig et al., 2011). The goal is to determine whether there exist any significant mean differences between the biomechanical signals for the different experimental conditions.

#### 2.1.2. Regions of deviation

The Regions of Deviation (ROD) approach (DiBerardino et al., 2012; Shorter et al., 2008) is a pointwise analysis technique that attempts to determine which time points exhibit significant mean differences between biomechanical signals from different experimental conditions. To test for significant differences, the ROD approach performs a  $t$  test at each time point of interest. In the case of  $G=2$  groups, only one  $t$  test is needed per time point; however, for  $G > 2$  groups, multiple  $t$  tests would be needed at each time point. The primary limitation of the ROD approach—and related pointwise analysis techniques—is that such methods ignore the functional nature of biomechanical data, which can hinder one's ability to discover and understand subtle differences in the data. Furthermore, using pointwise analysis methods requires either (i) a priori selection of the time points to investigate or (ii) examining differences at all time points, which requires repeated significance testing on the same correlated data.

#### 2.1.3. Smoothing spline ANOVA

We propose a two-way mixed-effects SSANOVA model of the form

$$y_i = \eta(\mathbf{x}_i) + b_{s_i} + e_i \quad (1)$$

for  $i = 1, \dots, n$  where  $n = 101 \sum_s \sum_g R_{sg}$  is the total number of data points after vectorizing the data,  $y_i$  is the recorded biomechanical signal,  $\mathbf{x}_i = (t_i, g_i)$  is the fixed effect predictor vector corresponding to the  $i$ -th data point with  $t_i \in \{0, \dots, 100\}$  denoting the time point (percent cycle) and  $g_i \in \{1, \dots, G\}$  denoting the group membership (condition),  $s_i \in \{1, \dots, S\}$  is the subject indicator,  $b_{s_i}$  is the random effect corresponding to the  $i$ -th data point such that  $b_{s_i} \stackrel{\text{iid}}{\sim} N(0, \tau\sigma^2)$  with  $\tau > 0$ , and  $e_i \stackrel{\text{iid}}{\sim} N(0, \sigma^2)$  is random error that is independent of the  $b_{s_i}$  effects. Note that the error variance  $\sigma^2$  captures the variation within an individual's data, whereas the variance component  $\tau\sigma^2$  captures the variation between individuals' data. The intraclass correlation coefficient (ICC) is given by  $\rho = \tau/(1 + \tau)$ , which is the correlation between two data points collected from the same subject.

The function  $\eta$  can be decomposed into main and interaction effects, such as

$$\eta(\mathbf{x}_i) = \eta_0 + \eta_1(t_i) + \eta_2(g_i) + \eta_{12}(t_i, g_i) \quad (2)$$

where  $\eta_0$  is a constant function,  $\eta_1$  is the main effect of time (percent cycle),  $\eta_2$  is the main effect of the experimental condition (group), and  $\eta_{12}$  is the time-group interaction effect. Using the SSANOVA decomposition in Eq. (2), it is possible to separate  $\eta$  into the average cycle ( $AC = \eta_0 + \eta_1$ ) and the contrast cycle ( $CC = \eta_2 + \eta_{12}$ ), where the AC captures the periodic shape over the biomechanical cycle that is shared by all of the subjects regardless group membership, and the CC captures how a specific group  $g$  differs from (i.e., contrasts) the overall average shape. In addition, the SSANOVA model makes it possible to examine the contrast cycle difference (CCD) between two groups

$$\delta(t|g, g_*) = \eta(t, g) - \eta(t, g_*) \quad (3)$$

where  $g, g_* \in \{1, \dots, G\}$  denote the group memberships. The CCD quantifies differences between the biomechanical signals of groups  $g$  and  $g_*$  across the entire cycle, and can be used to identify regions of the cycle where functional differences exist between the groups.

Using the Bayesian interpretation of a smoothing spline (Gu and Wahba, 1993; Nychka, 1988; Wahba, 1983), it is possible to form Bayesian confidence intervals (CIs) around the estimated function  $\hat{\eta}$  and CCD  $\hat{\delta}$  across the entire cycle. The SSANOVA fitted values can be written as  $\hat{\eta}(\mathbf{x}) = \mathbf{f}_x \hat{\beta} = \sum_{j=1}^k \hat{\beta}_j f_j(\mathbf{x})$ , where  $\mathbf{f}_x$  is the transpose of  $\mathbf{f}_x = (f_j(\mathbf{x}))_{k \times 1}$  which is a vector of reproducing kernel functions evaluated at  $\mathbf{x} = (t, g)$ , and  $\hat{\beta} = \{\hat{\beta}_j\}_{k \times 1}$  is the estimated function coefficient vector. The Bayesian CIs for  $\hat{\eta}$  and  $\hat{\delta}$  have the form

$$\begin{aligned} \hat{\eta} &\pm C_{\alpha/2} \sqrt{V_{\eta|\mathbf{y}}} \\ \hat{\delta} &\pm C_{\alpha/2} \sqrt{V_{\delta|\mathbf{y}}} \end{aligned} \quad (4)$$

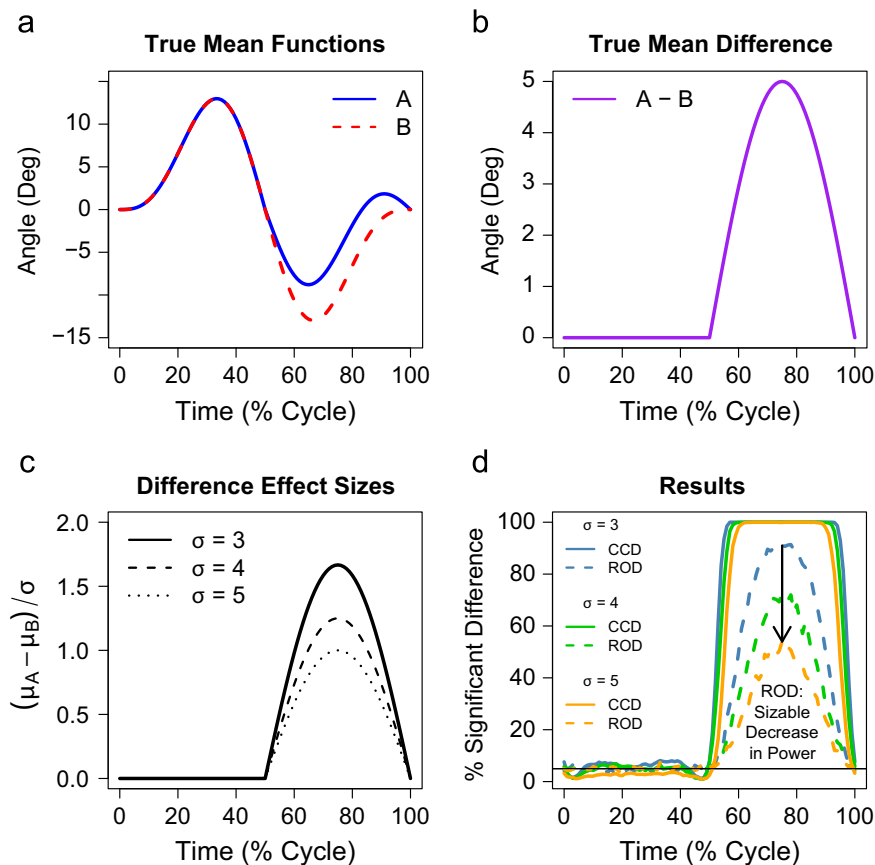
where  $C_{\alpha/2}$  is some critical value for a level  $\alpha$  significance test,  $V_{\eta|\mathbf{y}} = \mathbf{f}_x' V_{\beta} \mathbf{f}_x$  is the posterior variance of  $\eta$  evaluated at  $\mathbf{x} = (t, g)$ ,  $V_{\beta}$  denotes the  $k \times k$  posterior covariance matrix of  $\hat{\beta}$ , and  $V_{\delta|\mathbf{y}} = \mathbf{f}_{xx'}' V_{\beta} \mathbf{f}_{xx'}$  is the posterior variance of  $\delta$  with  $\mathbf{f}_{xx'} = \mathbf{f}_x - \mathbf{f}_{x^*}$  denoting the difference between the kernel functions evaluated at  $\mathbf{x} = (t, g)$  and  $\mathbf{x}^* = (t, g^*)$ .

For classic SSANOVA models (i.e., those without random effects), the critical value  $C_{\alpha/2}$  is the quantile of a standard normal distribution. For mixed-effects SSANOVA models, we recommend using the quantile of a  $t$  distribution with degrees of freedom equal to  $S - 1$  given that there are only  $S$  independent sampling units (i.e., subjects). Furthermore, using this definition of  $C_{\alpha/2}$  makes the critical values of the SSANOVA and ROD approaches equivalent, which facilitates comparisons between the two methods. Note that the CI for  $\delta$  can be used to test for differences between the groups: time points where the CI does not contain zero are determined to exhibit significant mean differences between the groups. Thus, the CCDs proposed here provide a functional and powerful approach that can be used to find regions of statistically significant differences in gait signals.

### 2.2. Simulation study

We designed a simple simulation study to compare the significance testing results of the classic ROD approach versus the proposed CCDs with Bayesian CIs. In the simulation we assume that we have recorded some cyclic biomechanical signal from  $S=10$  subjects who have participated in  $G=2$  experimental conditions. We assume that each subject's data has been normalized to have  $T=101$  time points, which represent 0, ..., 100% of the biomechanical cycle. For simplicity, we assume that we only have  $R_{sg} = 1$  replication from each subject in each condition. The true mean functions used for the simulation are plotted in Fig. 1a. Note that the true mean functions are identical for time points 0–50% (and time point 100%), and have differences at time points 51–99%. The magnitudes of the group mean differences across the cycle are plotted in Fig. 1b–c, which reveals that the maximum group mean difference occurs at 75% of the cycle.

Using the functions in Fig. 1a, we generated data by randomly sampling  $b_{s_i} \stackrel{\text{iid}}{\sim} N(0, \tau\sigma^2)$  and  $e_i \stackrel{\text{iid}}{\sim} N(0, \sigma^2)$ , and defining the response according to Eq. (1). We investigated three different levels of the error standard deviation  $\sigma \in \{3, 4, 5\}$ , which span the range of values in our observed data. We fixed  $\tau = 0.5$  throughout the simulation, which produces an ICC that is representative of our observed data, i.e.,



**Fig. 1.** Top: true mean functions used in simulation (left) and true mean difference (right). Bottom: difference effect size at different noise levels (left) and significance testing results using contract cycle difference (CCD) and regions of deviation (ROD) approaches (right). In subplot d), note that the ROD approach has a sizable decrease in power as the error standard deviation increases, whereas the CCD approach has a minimal decrease in power.

$\rho = \tau / (1 + \tau) = 1/3$ . For each level of  $\sigma$ , we generated 1000 replications of the data, and applied the ROD and CCD approaches to each generated sample. To assess the performance of the approaches, we compare the percentage of times (across the 1000 replications) that each method declared a significant difference at each time point using an  $\alpha = 0.05$  level.

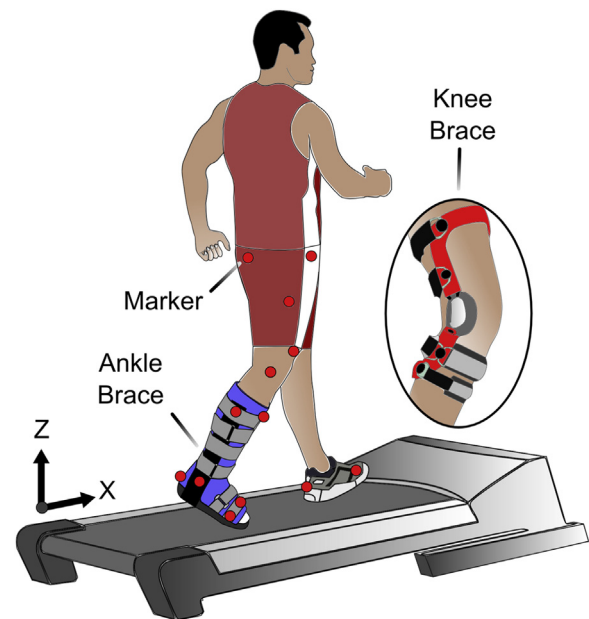
### 2.3. Data example

Locomotion data from Shorter et al. (2008) were used to demonstrate the benefits of our proposed model for the analysis of cyclic biomechanical data. Data were collected from ten healthy male subjects ( $21 \pm 2$  years) during steady state treadmill walking under three conditions: (a) normal, non-braced (NB) walking, (b) knee-braced (KB) walking such that right knee motion was restricted to full extension, and (c) ankle-braced (AB) walking such that right ankle motion was restricted to neutral position (i.e., perpendicular to the shank).

Subjects walked at a self-selected pace on a treadmill. The same self-selected speed was used for all walking conditions and was identified during unconstrained gait. Kinematic marker data were collected using a six-camera infrared motion analysis system at 120 Hz (Vicon, Oxford, UK; Model 460). Markers were placed on the pelvis and lower limbs of the subjects. Tightly fitting spandex shorts and tops were worn to minimize any motion artifact. Fig. 2 provides an illustration of the experiment, and Shorter et al. (2008) give additional details regarding the data and data collection procedure.

The kinematic data were first low-pass filtered at 8 Hz using a fourth-order, zero-lag, Butterworth filter. Next, bilateral hip, knee, and ankle flexion/extension angle waveforms were calculated from the filtered data (Vaughan et al., 1999). For each subject and condition, joint angle data were then segmented into gait cycles (GCs; defined from heel-strike to heel-strike) separately for each limb. Lastly, joint angle data were linearly length normalized to 101 time points representing 0–100% GC (Helwig et al., 2011). The data analyzed in this study comprise ten consecutive (bilateral) GCs for each subject in each condition.

Using the mixed-effects smoothing spline ANOVA model proposed in Eq. (1), we fit a model to the joint angle data from each of the six joints separately. All models were fit using the `bigssa` function in `bigsplines` package (Helwig, 2016) in the R environment (R Core Team, 2016). For each model, a periodic smoothing



**Fig. 2.** Depiction of the data collection set-up, including the ankle and knee braces.

spline was used for  $t$  (% GC), whereas  $g$  (group: NB, KB, AB) was treated as a nominal variable. The variances of the random subject effects and the random errors (i.e.,  $\tau\sigma^2$  and  $\sigma^2$ ) were estimated from the data using a restricted maximum likelihood approach (Helwig, 2015). The generalized cross-validation method

(Craven and Wahba, 1979) was used to estimate the model's smoothing parameters. Finally, we used recent SSANOVA approximations presented in Helwig and Ma (2015) for fast and reliable computation. See the SOM for further details of our model fitting procedure, including our analysis code and data.

### 3. Results

#### 3.1. Simulation study

Fig. 1d illustrates the significance testing results for the CCD and ROD approaches. Note that both approaches fail to find a difference

about 95% of the time for time points 0–50% (and time point 100%), which are the points where no true difference exists. However, during the 51–99% window, the CCD approach correctly declares a difference with higher probability than the ROD approach, and the benefit of the CCD approach increases as the error variance increases. We note that if the number of replications  $R_{\text{sg}}$  were increased, the ROD approach would have more power and would begin to resemble the performance of the CCD approach. However, these simulation results reveal that—even with a single replication of the cycle per subject—the SSANOVA approach has the power to reliably find subtle differences in biomechanical signals.

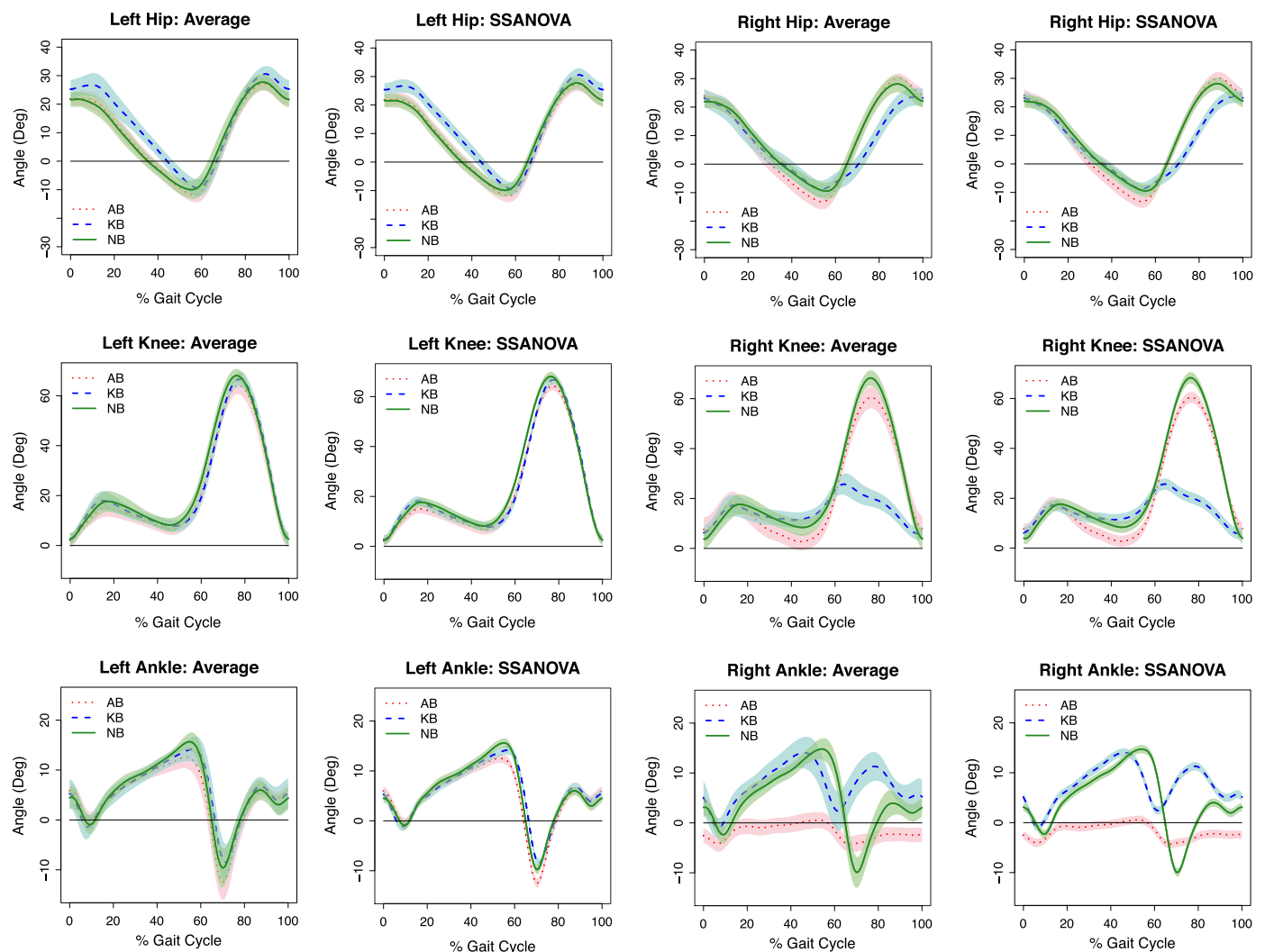
#### 3.2. Data example

The SSANOVA models fit the locomotion data well, with total  $R^2$  values ranging from 77% to 95% (Table 1). The average and SSANOVA predicted joint angle trajectories are plotted in Fig. 3, along with 95% CIs. Note that the SSANOVA predictions closely resemble the pointwise averages, which is expected. The primary difference is that the standard errors (and CIs) for the pointwise approach ignore the cycle-to-cycle variability in the data by averaging each subject's data across the cycle replications. In contrast, the SSANOVA models both the within and between cycle variability in the data, which can result in substantial decreases in the standard errors—and, thus, narrower CIs. This is particularly true for the

**Table 1**  
Results for the SSANOVA models.

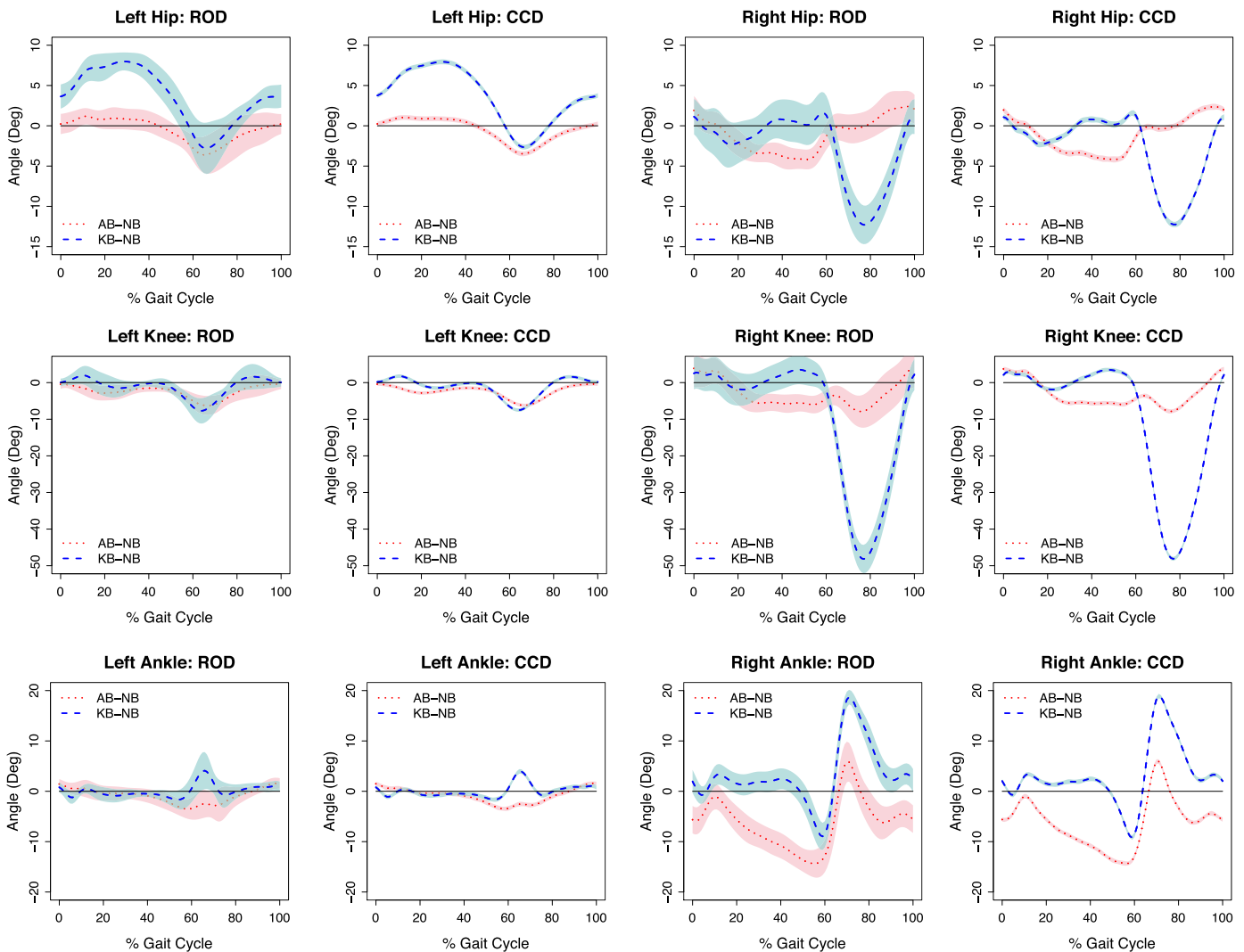
Model	$R^2$	$\hat{\tau}$	$\hat{\sigma}$	$\hat{\rho}$
Hip L	0.91	1.42	2.71	0.59
Hip R	0.91	1.45	2.57	0.59
Knee L	0.95	0.46	3.77	0.32
Knee R	0.91	0.57	4.15	0.36
Ankle L	0.77	0.15	3.08	0.13
Ankle R	0.77	0.13	3.04	0.12

Note:  $\hat{\rho} = \hat{\tau} / (1 + \hat{\tau}) = \text{ICC}$ .



**Fig. 3.** Average joint angle trajectory in each condition with 95 interval, along with the SSANOVA predicted joint angle trajectory in each condition with 95% Bayesian confidence interval.





**Fig. 4.** Mean difference joint angle trajectory in each condition with 95 confidence interval (i.e., ROD), along with SSANOVA predicted joint angle trajectory difference in each condition with 95% Bayesian confidence interval (i.e., CCD).

knee and ankle data, where the cycle-to-cycle variability is small. The primary benefit of the SSANOVA approach is evident in Fig. 4, which plots the estimated average difference between the braced and non-braced conditions using both the ROD and CCD approaches. Fig. 4 reveals that the SSANOVA provides a more precise estimate of the differences in the biomechanical data.

Knee bracing caused large changes to the joint motion of the affected (right) leg: (a) reduction of right hip flexion ( $10\text{--}15^\circ$ ) during swing (60–90% GC), (b) reduction of right knee flexion ( $40\text{--}50^\circ$ ) throughout swing (60–100% GC), and (c) increased right ankle dorsiflexion ( $10\text{--}15^\circ$ ) during early swing (60–80% GC). In contrast, the knee bracing created small increases in left hip flexion ( $5\text{--}10^\circ$ ) throughout stance (0–60% GC), and small changes to the left knee and ankle joints ( $5\text{--}10^\circ$ ) around swing initiation (55–60% GC). Ankle bracing created smaller compensatory changes in right joint movement: (a) increases in right hip extension ( $<5^\circ$ ) during mid-to-late stance (20–50% GC), (b) reduced right knee flexion during stance (20–60% GC) and mid-swing (75–85% GC), and (c) significant reductions in right ankle motion ( $10\text{--}20^\circ$ ) throughout the GC. Ankle bracing also created increases in left hip flexion ( $<5^\circ$ ) during late stance and early swing (50–80% GC), as well as reduced left knee flexion ( $5\text{--}10^\circ$ ) and left ankle dorsiflexion and plantarflexion ( $<5^\circ$ ) during late stance and early swing (40–70% GC).

## 4. Discussion

### 4.1. Summary of results

Our results reveal the potential of the proposed SSANOVA framework for the statistical analysis of differences in cyclic biomechanical data. Using a simulation study, we demonstrated that the proposed SSANOVA approach has more statistical power than the classic ROD approach, which implies that the SSANOVA has greater potential to find subtle differences in biomechanical data. Furthermore, our example using experimentally-collected kinematic data demonstrates the potential of the SSANOVA approach for analyzing real biomechanical data. Together, the simulation study and real data example clearly highlight the primary benefit of taking a functional analysis approach—instead of the classic pointwise analysis approach. By partitioning the data into the AC and CC, the SSANOVA approach is able to examine differences using only the relevant information in the CC, while the irrelevant information in the AC is ignored. The result is a more precise estimate of the functional differences in the data.

The real data example revealed that the effects of the ankle bracing were most evident in the ipsilateral (i.e., right) limbs' behaviors. During stance the restricted motion at the right ankle limited the free rotation of the shank, and reduced plantar-flexion

at push-off. To compensate, right knee and hip flexion were significantly decreased during stance, with reduced right knee flexion continuing into swing. Differences seen at the joints of the contralateral leg occur mainly around the stance to swing transition. The effects of the knee bracing were more evident, and were pronounced in both the ipsilateral and contralateral limb behavior. The brace on the right knee severely limited right knee flexion, hindering limb advancement during swing. To compensate for this reduced knee flexion, subjects dorsi-flexed the right ankle during early and mid-swing to improve the clearance of the foot during limb advancement. Furthermore, left hip flexion was significantly increased during stance at the same time that right hip flexion was significantly decreased during swing, which created a hip-hiking motion to provide additional clearance for the right foot.

#### 4.2. Implications for gait analysis

The SSANOVA approach presented in this paper addresses fundamental shortcomings of commonly used approaches for averaging and statistical analysis of biomechanical data. First, by simultaneously analyzing multiple replications of biomechanical data, a SSANOVA model provides improved estimates of the underlying biomechanical signals corresponding to each experimental condition (or subject population), as well as information about the cycle-to-cycle variability in the data (via the  $\tau$  and  $\rho$  parameters). Secondly, this approach makes it possible to identify entire regions of the movement data where experimental conditions (or subject populations) significantly differ from one another (via the CCD). This ability to identify entire regions of movement data that are affected by an experimental condition, pathology, or impairment in a statistically rigorous manner presents a significant contribution to gait and locomotion analysis. In contrast to the ROD approach (Shorter et al., 2008; DiBerardino et al., 2012), the work presented here provides a functional (instead of a pointwise) approach to the statistical analysis, which leverages the functional nature of the biomechanical data. As such, the CCDs (with Bayesian CIs) presented in this paper make it possible to identify statistically significant regions of deviation between conditions in the functional data.

#### 4.3. Modifications and extensions

It should be noted that the methodology presented in this paper could be adjusted and/or extended in a variety of ways depending on the situation. For example, if one wants to include multiple joints (e.g., left and right limb) in the same model, a three-way SSANOVA model could be used such as  $y_i = \eta(t_i, g_i, k_i) + b_{s_i} + e_i$ , where  $k_i \in \{L, R\}$  indicates the limb (left/right) corresponding to the  $i$ -th data point and the other terms can be interpreted as previously described. As another example, if the subjects are nested in different groups, then the  $b_{s_i} \stackrel{iid}{\sim} N(0, \tau\sigma^2)$  effects could be replaced with  $b_{s_i, g_i} \stackrel{iid}{\sim} N(0, \tau_{g_i}\sigma^2)$  effects, which would allow each group to have a unique variance component. Furthermore, it is possible to include all six joints (bilateral hip, knee, ankle) in one model; in this case, we could replace the  $b_{s_i} \stackrel{iid}{\sim} N(0, \tau\sigma^2)$  effects with  $b_{s_i, k_i} \stackrel{iid}{\sim} N(0, \tau_{k_i}\sigma^2)$  effects, which would allow each joint to have a unique variance component. Finally, using an unconstrained smoothing spline for the time effect, the proposed mixed-effects SSANOVA models could be used to analyze individual and group differences in non-cyclic biomechanical data.

#### 4.4. Limitations

The primary limitation of the proposed methodology is that—like classic pointwise analysis approaches—the SSANOVA approach assumes that (i) the cyclic biomechanical data have been partitioned

into cycles before analysis, and (ii) the cycle start and end points are correctly identified, so that corresponding time points are reasonably comparable after the time normalization. However, with real cyclic biomechanical data such as gait data, the identification of the cycle start and end points is not necessarily simple, and there may be errors in the identifications. For steady state gait data, we have found that data-driven approaches can reliably identify GC start and end points, making the above two assumptions reasonable for these particular data. However, identifying the cycle start and end points may not be so reliable for other types of cyclic biomechanical data. An interesting extension of the proposed methodology—that is beyond the scope of this paper—would be to fit an SSANOVA model to multi-cycle biomechanical data, and use the SSANOVA results to identify the cycle start and end points, e.g., from the estimated derivatives of the function.

#### 4.5. Conclusions

In this paper, we demonstrated the benefits of analyzing cyclic biomechanical data using a mixed-effects SSANOVA model. Unlike traditional univariate analysis approaches, our proposed model enables the analysis of functional differences by treating biomechanical signals in a functional manner. In addition, this approach can simultaneously analyze multiple cycle replications from many subjects who belong to different subject groups/conditions. Thus, the framework presented here will be useful in both laboratory and clinical settings for the analysis of biomechanical signals, and the identification of significant functional differences between experimental conditions and/or subject groups.

#### Conflict of interest statement

The authors have no conflicting interests.

#### Acknowledgments

This work was funded by start-up funds from the University of Minnesota, start-up funds from the University of Michigan, the Mary Jane Neer Disability Research Fund at the University of Illinois, and NSF grants DMS 1440037, DMS 1438957, and EEC 0540834.

#### Appendix A. Supplementary data

Supplementary data associated with this article can be found in the online version at <http://dx.doi.org/10.1016/j.jbiomech.2016.07.035>.

#### References

- Bates, D., Mächler, M., Bolker, B.M., Walker, S.C., 2015. Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* 67, 1–48.
- Becker, P.H., Rosenbaum, D., Kriese, T., Gerngro, H., Clase, L., 1995. Gait asymmetry following successful surgical treatment of ankle fractures in young adults. *Clin. Orthop. Relat. Res.* 311, 262–269.
- Chmielewski, T.L., Hurd, W.J., Rudolph, K.S., Axe, M.J., Snyder-Mackler, L., 2005. Perturbation training improves knee kinematics and reduces muscle co-contraction after complete unilateral anterior cruciate ligament rupture. *Phys. Ther.* 85, 740–749.
- Collins, S.H., Adamczyk, P.G., Kuo, A.D., 2009. Dynamic arm swinging in human walking. *Proceedings of the Royal Society B: Biological Sciences* 276, pp. 3679–3688.
- Craven, P., Wahba, G., 1979. Smoothing noisy data with spline functions: estimating the correct degree of smoothing by the method of generalized cross-validation. *Numer. Math.* 31, 377–403.

- DeVita, P., Hortobagyi, T., Barrier, J., Torry, M., Glover, K.L., Speroni, D.L., Money, J., Mahar, M.T., 1997. Gait adaptations before and after anterior cruciate ligament reconstruction surgery. *Med. Sci. Sport. Exerc.* 29, 853–859.
- DiBerardino 3rd, L.A., Ragetly, C.A., Hong, S., Griffon, D.J., Hsiao-Weckslers, E.T., 2012. Improving regions of deviation gait symmetry analysis with pointwise *t* tests. *J. Appl. Biomech.* 28, 210–214.
- Diop, M., Rahmani, A., Belli, A., Gautheron, V., Geyssant, A., Cottalorda, J., 2005. Influence of speed variation and age on ground reaction forces and stride parameters of children's normal gait. *Int. J. Sport. Med.* 26, 682–687.
- Forner-Cordero, A., Koopman, H.J.F.M., van der Helm, F.C.T., 2006. Describing gait as a sequence of states. *J. Biomech.* 39, 948–957.
- Griffin, M.P., Olney, S.J., McBride, I.D., 1995. Role of symmetry in gait performance of stroke subjects with hemiplegia. *Gait Posture* 3, 132–142.
- Gu, C., 2013. *Smoothing Spline ANOVA Models*, 2nd edition. Springer-Verlag, New York.
- Gu, C., Ma, P., 2005. Optimal smoothing in nonparametric mixed-effect models. *Ann. Stat.* 33, 1357–1379.
- Gu, C., Wahba, G., 1993. Smoothing spline ANOVA with component-wise Bayesian “confidence intervals”. *J. Comput. Graph. Stat.* 2, 97–117.
- Hausdorff, J.M., Rios, D.A., Edelberg, H.K., 2001. Gait variability and fall risk in community-living older adults: a 1-year prospective study. *Arch. Phys. Med. Rehabil.* 82, 1050–1056.
- Helwig, N.E., 2015. Efficient estimation of variance components in nonparametric mixed-effects models with large samples. *Statistics and Computing* (Advance online publication), pp. 1–18. <http://dx.doi.org/10.1007/s11222-015-9610-5>.
- Helwig, N.E., 2016. *bigsplines: Smoothing Splines for Large Samples*. R package version 1.0-8. URL <<http://cran.r-project.org/package=bigsplines>>.
- Helwig, N.E., Hong, S., Hsiao-Weckslers, E.T., Polk, J.D., 2011. Methods to temporally align gait cycle data. *J. Biomech.* 44, 561–566.
- Helwig, N.E., Ma, P., 2015. Fast and stable multiple smoothing parameter selection in smoothing spline analysis of variance models with large samples. *J. Comput. Graph. Stat.* 24, 715–732.
- Helwig, N.E., Ma, P., 2016. Smoothing spline ANOVA for super-large samples: Scalable computation via rounding parameters. *Statistics and Its Interface* (in press).
- Huang, T.-w.P., Kuo, A.D., 2014. Mechanics and energetics of load carriage during human walking. *J. Exp. Biol.* 217, 605–613.
- James C.R., 2004. Considerations of movement variability in biomechanics research, Stergiou N., (Ed), In: *Innovative Analyses of Human Movement*, Human Kinetics, Champaign, IL, 29–62.
- Knoll, Z., Kocsis, L., Kiss, R.M., 2004. Gait patterns before and after anterior cruciate ligament reconstruction. *Knee Surg. Sport. Traumatol. Arthrosc.* 12, 7–14.
- Lucero, J.C., Koenig, L.L., 2000. Time normalization of voice signals using functional data analysis. *J. Acoust. Soc. Am.* 108, 1408–1420.
- Nimeskern, L., Pleumeekers, M.M., Pawson, D.J., Koevoet, W.L., Lehtoviita, I., Soyka, M.B., Rösli, C., Holzmann, D., van Osch, G.J., Müller, R., Stok, K.S., 2015. Mechanical and biochemical mapping of human auricular cartilage for reliable assessment of tissue-engineered constructs. *J. Biomech.* 48, 1721–1729.
- Nychka, D., 1988. Bayesian confidence intervals for smoothing splines. *J. Am. Stat. Assoc.* 83, 1134–1143.
- O'Connor, S.M., Xu, H.Z., Kuo, A.D., 2012. Energetic cost of walking with increased step variability. *Gait Posture* 36, 102–107.
- Page, A., Ayala, G., León, M.T., Peydro, M.F., Prat, J.M., 2006. Normalizing temporal patterns to analyze sit-to-stand movements by using registration of functional data. *J. Biomech.* 39, 2526–2534.
- R Core Team, 2016. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL: <<http://www.R-project.org/>>.
- Risberg, M.A., Moksnes, H., Storevold, A., Holm, I., Snyder-Mackler, L., 2009. Rehabilitation after anterior cruciate ligament injury influences joint loading during walking but not hopping. *Br. J. Sport. Med.* 43, 423–428.
- Romei, M., Galli, M., Motta, F., Schwartz, M., Crivellini, M., 2004. Use of the normalcy index for the evaluation of gait pathology. *Gait Posture* 19, 85–90.
- Shorter, K.A., Polk, J.D., Rosengren, K.S., Hsiao-Weckslers, E.T., 2008. A new approach to detecting asymmetries in gait. *Clin. Biomech.* 23, 459–467.
- Vaughan, C.L., Davis, B.L., O'Connor, J.C., 1999. *Dynamics of Human Gait*, 2nd edition. Kiboho, Cape Town, South Africa.
- Wahba, G., 1983. Bayesian “confidence intervals” for the cross-validated smoothing spline. *J. R. Stat. Soc. Ser. B* 45, 133–150.
- Wang, Y., 1998a. Mixed effects smoothing spline analysis of variance. *J. R. Stat. Soc. Ser. B* 60, 159–174.
- Wang, Y., 1998b. Smoothing spline models with correlated random errors. *J. Am. Stat. Assoc.* 93, 341–348.
- Zhang, D., Lin, X., Raz, J., Sowers, M., 1998. Semiparametric stochastic mixed models for longitudinal data. *J. Am. Stat. Assoc.* 93, 710–719.