ELSEVIER

Gait and Posture 9 (1999) 10–17

GAIT & POSTURE

# Bootstrap prediction and confidence bands: a superior statistical method for analysis of gait data

Mark W. Lenhoff [a,*], Thomas J. Santner [b], James C. Otis [a], Margaret G.E. Peterson [a], Brian J. Williams [b], Sherry I. Backus [a]

[a] *The Hospital for Special Surgery, Motion Analysis Laboratory, 535 E 70 Street, New York City NY 10021, USA*
[b] *The Ohio State University, Columbus, Ohio, USA*

Received 2 June 1998; accepted 16 November 1998

**Abstract**

Gait analysis studies typically utilize continuous curves of data measured over the gait cycle, or a portion of the gait cycle. Statistical methods which are appropriate for use in studies involving a single point of data are not adequate for analysis of continuous curves of data. This paper determines the operating characteristics for two methods of constructing statistical prediction and confidence bands. The methods are compared, and their performance is evaluated using cross-validation methodology with a data set of the sort commonly evaluated in gait analysis. The methods evaluated are the often-used point-by-point Gaussian theory intervals, and the simultaneous bootstrap intervals of Sutherland et al. *The Development of Mature Walking*, MacKeith Press, London, 1988 and Olshen et al. Ann. Statist. 17 (1989) 1419–40. The bootstrap bands are shown to provide appropriate coverage for continuous curve gait data (86% coverage for a targeted coverage of 90%). The Gaussian bands are shown to provide inadequate coverage (54% for a targeted coverage of 90%). The deficiency in the Gaussian method can lead to inaccurate conclusions in gait studies. Bootstrap prediction and confidence bands are advocated for use as a standard method for evaluating gait data curves because the method is non-parametric and maintains nominal coverage levels for entire curves of gait data. 

*Keywords:* Bootstrap prediction; Confidence bands; Gait data

## 1. Introduction

The data analyzed in a typical gait study consists of continuous curves of data expressed as a function of the percentage of the gait cycle. Gait data that consists of a parameter vs. time (e.g. knee flexion angle vs percentage of gait cycle) must be handled using a different method than data that consists of a single observation (e.g. knee flexion angle at heel strike). Statistical methods appropriate for the analysis of less complex single point numerical data are inadequate when applied to continuous curves of gait data.

For numerical data, recall that prediction intervals contain, with pre-specified coverage probability, a new observation from the same population from which a statistical (or training) sample is drawn. For continuous, or gait curve data, the analogue of prediction intervals are *prediction bands*. Prediction bands contain, with pre-specified coverage probability, a new curve drawn from the same population as the training curves. One use of prediction bands is to classify new subjects as belonging to the same population, or not, as that from which the training curves are collected.

As an example, consider a prediction band calculated from a normal subject database that is applied to new subjects to determine if they should be classified as members of the normal population. If the curve of a

* Corresponding author.

0966-6362/99/$ - see front matter 
PII: S0966-6362(98)00043-5

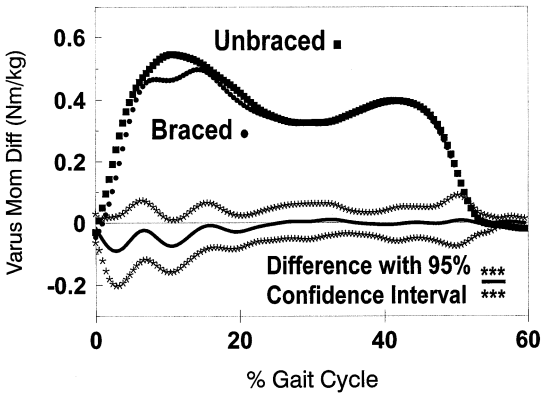


Fig. 1. Difference of braced and un-braced varus moment in ten subjects plotted together with 95% bootstrap confidence band for the mean of the difference.

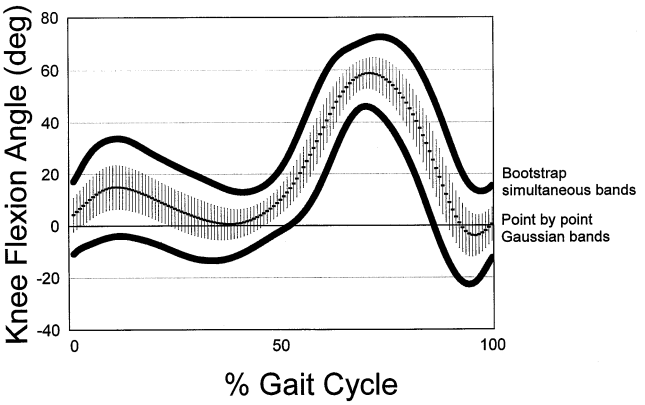


Fig. 2. For the knee flexion data in Fig. 4, individual 90% point-by-point prediction intervals (vertical lines), and the corresponding 90% bootstrap prediction bands (solid line).

new subject falls outside of the prediction band, it can be stated that the new subject is statistically different than the population in the normal subject database.

In some research studies it is important to determine *confidence bands* for the mean curve of a given population. For example, a researcher might want to characterize the effect upon the mean external varus moment of a group of subjects treated with a given type of knee brace. In this example, the researcher would wish to determine a confidence band for the mean difference in the external varus moment for subjects while wearing the brace and the external varus moment for the same subjects while not wearing the brace. If the confidence band about the difference of the braced and un-braced curves of each subject contains the horizontal line at value zero, then one could state that there was no significant difference between the two conditions. This application is illustrated in Fig. 1.

One popular method of constructing prediction bands is to apply Gaussian theory to the univariate data available at each percent of the gait cycle to determine a prediction interval for the data at each plotting position [3,4]. In this analysis, the collection of separate point-by-point prediction intervals is used as a prediction band. Confidence bands are formed in a similar manner except that confidence intervals based on the T distribution are computed for each fixed percent of the gait cycle. This method of analysis ignores the fact that many points are being considered simultaneously when an entire curve is considered. Such *point-by-point* bands are often used in gait analysis. For example, Fig. 2 is a plot of 100 separate 90% prediction intervals for knee flexion angle at 100 percentage points of the gait cycle, with the prediction band determined by the bootstrap method plotted as well.

If one wishes to examine only a few points in the gait cycle, a Bonferroni correction can be applied to the point-by-point (Gaussian) intervals to give the desired simultaneous coverage for the resulting band [5]. This correction widens the Gaussian limits. Unfortunately, the Bonferroni correction (and the width of the simultaneous limits at each point) increases with the *number* of points at which it is desired to form simultaneous prediction intervals. With data across the entire gait cycle, typically consisting of at least 100 points over the cycle, the limits determined by the Bonferroni method result in an extremely conservative evaluation. In the comparison below we omit comparison of the Bonferroni method with the point-by-point and bootstrap methods.

The bootstrap method is a computationally intensive technique for constructing bands which provide the desired coverage based on continuous curves such as is typical for gait analysis applications to angles, forces, and moments that are measured over the gait cycle or a portion of the gait cycle. Roughly, the bootstrap assesses the relationship between the true population and the sample by studying the relationship between the *given curves* treated as a pseudo-population and pseudo-samples drawn from these curves [6]. It uses the variability in the pseudo-samples to gauge the variability in samples that might be drawn from the true population and the location of center of the pseudo-population as an indicator of the location of the center of the true population. The method is represented pictorially in Fig. 3. The objective of this study was to quantify the true coverage probability for typical gait data curves using prediction bands constructed by the joining of point by point prediction intervals, and prediction bands constructed by the bootstrap method.

## 2. Methods

We estimated the true coverage probability of future curves using 90% prediction bands constructed by two methods as applied to knee flexion angle data for 28 normal subjects collected in the Motion Analysis Laboratory at the Hospital for Special Surgery. The normal subject curves are displayed in Fig. 4.

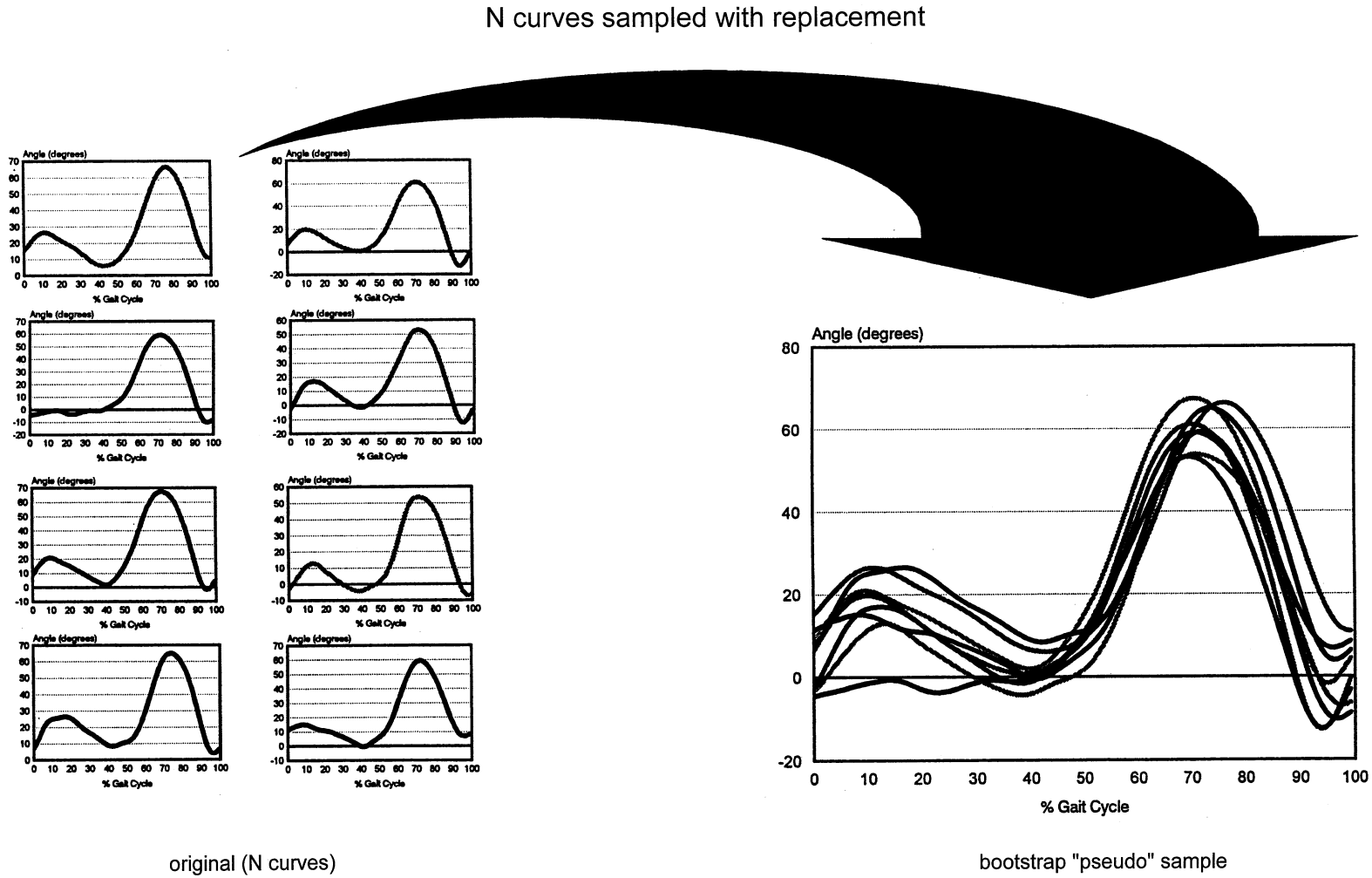


Fig. 3. Illustration of one bootstrap sample drawn from *N* curves. The variability and mean of many such samples bears the same relation to the population of the *N* original curves as do the *N* original curves to the underlying population.

The first method combines point-by-point Gaussian prediction intervals treating the result as a prediction band while the second method is the bootstrap band of Sutherland et al. and Olshen et al. [1,2]. The latter explicitly takes account of the simultaneous nature of the inference required over the percentage points of the gait cycle. A detailed mathematical explanation of the bootstrap band calculation is presented in Appendix A. All bootstrap bands calculated for this study were constructed using 400 iterations per band.

We quantified the achieved probability that new *curves* are covered for point-by-point prediction bands, and simultaneous bootstrap bands. Cross validation methodology was used to estimate the true achieved coverage probabilities for the two methods. To evaluate a given method, the idea of cross-validation is to remove one curve from the original data set, then calculate a prediction band using that method based on the remaining curves, and then determine if the band contains the removed curve, as illustrated in Fig. 5. This process is repeated for *each* of the curves in the original data set. For a given method of forming bands, the number of prediction bands constructed is equal to the number of curves in the data set. The proportion of deleted curves that are contained in the bands constructed from the corresponding data set with that curve deleted is an estimate of the true achieved coverage of the prediction bands.

The cross-validation technique was applied to the set of knee flexion curves. The goal was to estimate the true achieved coverage for both methods at a 90% limit. For the bootstrap bands, we wished to determine whether there was sufficient data to insure the true achieved coverage was close to the desired nominal level. We also wanted to verify that the achieved coverage for the bootstrap method was close to the desired nominal level. For point-by-point bands it was desired to quantify the deficiency by which the achieved coverage fell short of the nominal coverage.

## 3. Results

The results of the cross-validation calculations were as follows. For nominal 90% bootstrap prediction bands, the estimated true achieved coverage was 86% (24 of the 28 cross validation trials resulted in complete coverage of the deleted curve by the bootstrap band calculated from the remaining 27 curves). The standard error of the estimate is 6.6%; 90% is well with in two standard errors of 86%. This calculation suggests that with low harmonic curves, the achieved level of boot-

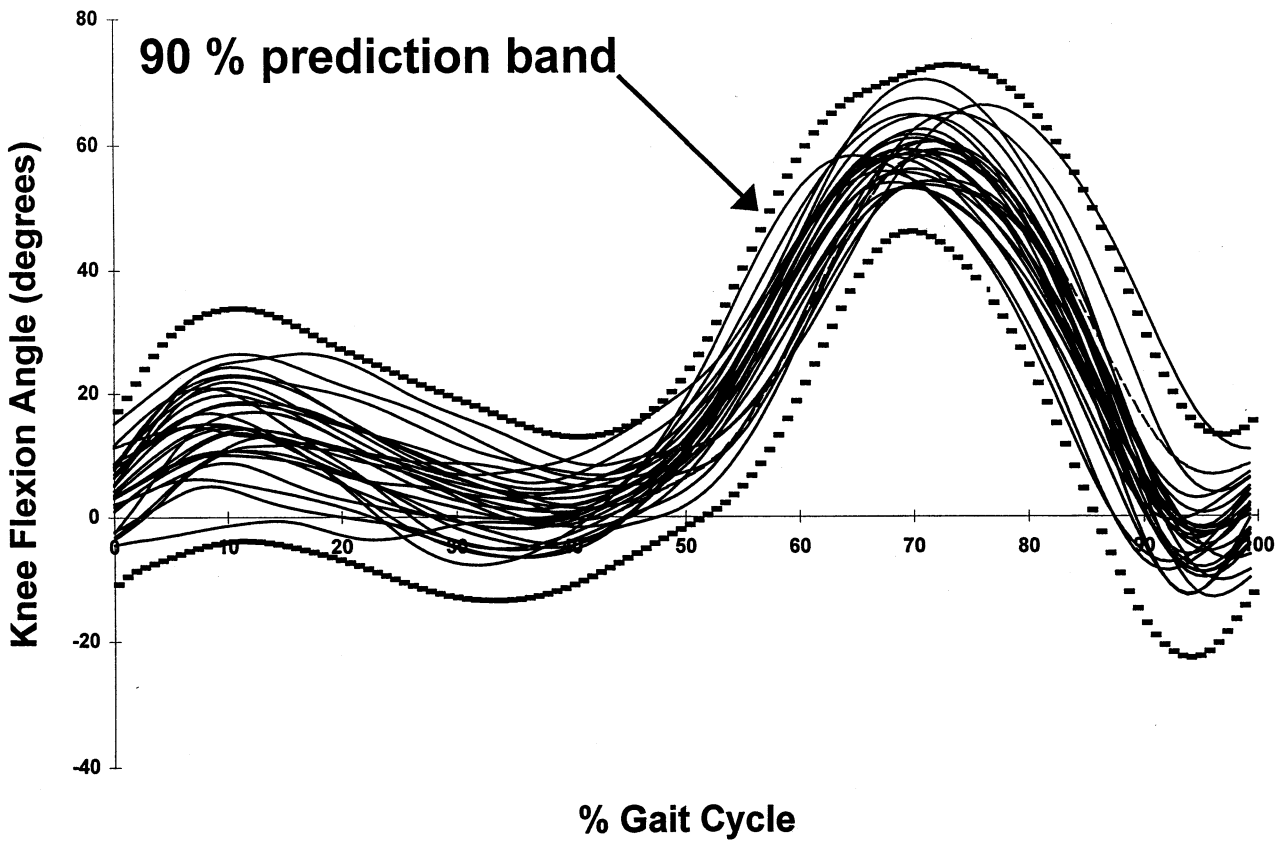


Fig. 4. Knee flexion angle versus percent gait cycle for 28 normal subjects, with a 90% bootstrap prediction band.

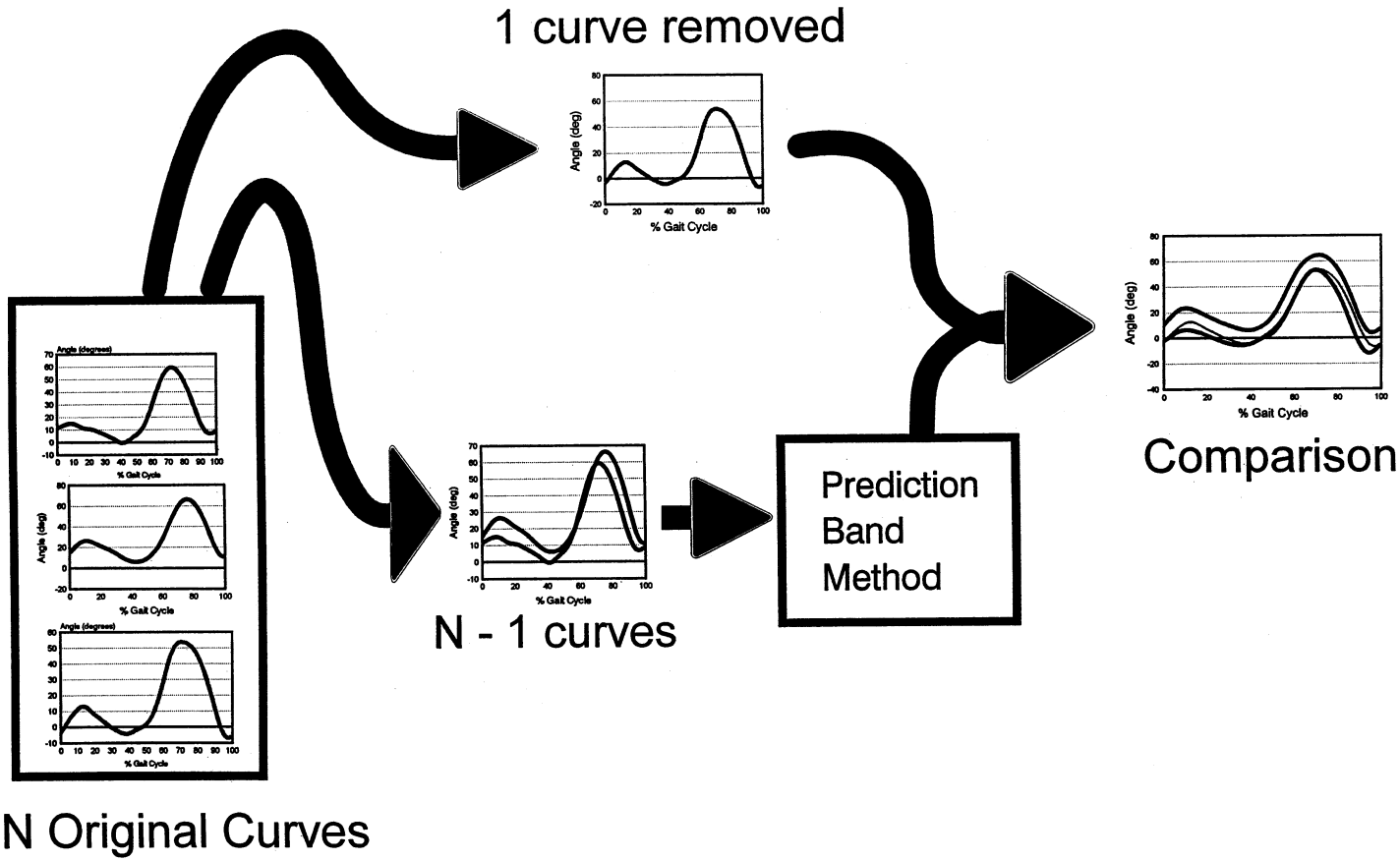


Fig. 5. Schematic illustration of the cross-validation method. For each cross-validation trial, one curve is removed from the original data set and compared to the band constructed from the remaining curves. This process was repeated once for each of the $N$ curves in the original data set, removing each curve once and only once. Coverage for a given cross-validation trial is adequate if the removed curve is completely within the statistical limits constructed from the remaining curves. In this study, coverage for new curves was estimated from all cross validation trials for each statistical method.

strap bands is roughly equal to the nominal level with as few as 25 or so curves.

The estimated true achieved coverage for prediction bands constructed from 90% point-by-point intervals was 54% (15 of the 28 curves removed during the cross-validation trials were covered). With the standard error of the estimate 9.4%, this estimate shows that the true coverage is much under the desired coverage, certainly not more than about 73% and possibly as low as 35% when 90% intervals are desired.

## 4. Discussion

This article quantifies the low achieved coverage for curves when bands are constructed by joining point-by-point intervals. In contrast, the article shows that the bootstrap bands introduced in Sutherland et al. [1] yield achieved coverage equal to nominal levels for data sets containing as few as 25 low frequency curves.

Confidence or prediction bands constructed using point-by-point methods have several shortcomings when applied to entire curves of continuous gait data. Gaussian theory for prediction or confidence intervals assumes that the data at each point of the gait cycle belong to a specific symmetric parametric family; this assumption must, in principle, be validated separately at each point at which the intervals are constructed. Even if all sets of gait data have Gaussian distributions, a band produced by *superimposing* point-by-point intervals is visually misleading. This is because the pair of curves formed in this way suggests that the coverage probability for *new* curves drawn from the same population as the training data is equal to the nominal level of the individual intervals. However, the probability of the simultaneous veracity of a set of statements is lower than the common individual probability of each one.

The bootstrap method does not rely on any parametric assumption concerning the distribution of the data at each point of the curve. Bootstrap prediction and confidence intervals can be theoretically justified asymptotically although their performance in small-sample settings has to be studied on a case-by-case basis.

The *bootstrap* is a *non-parametric* technique designed to form simultaneous confidence bands for a mean curve of population or prediction bands for a new curve drawn from the same population as the original data. The bootstrap method can also be used to form either type of band when the researcher is interested in a given, fixed part of the gait cycle such as between 40% and 100% of the gait cycle (e.g. stance or swing phase).

Finally, note that while Gaussian intervals are appropriate for studying a single point in the gait cycle such as knee flexion angle at toe-off, prediction intervals for *complex events* such as peak knee flexion angle requires that one knows simultaneous prediction intervals for all events at points near the peak excursion during the gait cycle. This is because one cannot a priori pin down the exact time in the gait cycle when the peak knee flexion angle occurs. The variability within and between subjects is sufficiently great that the peak excursion will not occur at the same percentage point of the gait cycle in each curve. A confidence interval for such a complex event can be validly computed using a simultaneous confidence band for knee flexion angle.

The primary disadvantage of the bootstrap method is its computationally intensive nature. The availability of inexpensive computing power has mitigated this disadvantage in recent years. Further work to quantify the minimum number of bootstrap iterations necessary for a given data set would prove beneficial in minimizing the amount of calculations performed. The data evaluated in this study utilized 400 bootstrap iterations. That number of iterations is shown to be adequate given the coverage which is demonstrated. Future work is necessary to determine the minimum number of bootstrap iterations for the method. A test for convergence vs number of iterations would be helpful to further refine the method. Because of the relatively few assumptions it requires, and the fidelity of its achieved coverage to its nominal coverage for low harmonic data, we recommend that bootstrap prediction and confidence bands be the gold standard method for evaluating gait curve data.

## Appendix A

We describe the Bootstrap Method for finding prediction bands based on data from $n$ curves. The data from the $i$th curve is denoted $Y_i(t_1), \ldots, Y_i(t_M)$ where, for convenience, we assume a common number $M$ of points is observed on each curve during the gait cycle and that $t$ ranges from 0 to $T$, i.e.

$$0 \le t_1 \le \cdots \le t_M \le T.$$

We regard the $n$ curves as perturbations, in a sense described below, of a true curve that can be represented by the finite Fourier sum

$$f(t) = \mu + \sum_{k=1}^{K} \left( \alpha_k \cos\left(\frac{2\pi k t}{T}\right) + \beta \sin\left(\frac{2\pi k t}{T}\right) \right) \qquad (0.1)$$

where $K$ is known. Thus $\mu$ is the overall mean, for example, and the form Eq. (0.1) asserts that the curve starts and ends at the same height. In practice, rather arbitrary curves can be accommodated by embedding them in a smooth way so the extended curve exhibits the symmetry properties that Eq. (0.1) entails.

An idealized version of the ith curve is viewed as being derived from the fundamental curve Eq. (0.1) by perturbing the coefficients of Eq. (0.1); this gives rise to

$$f_i(t) = \mu_i + \sum_{k=1}^{K}\left(\alpha_{i,k}\cos\left(\frac{2\pi kt}{T}\right) + \beta_{i,k}\sin\left(\frac{2\pi kt}{T}\right)\right) \tag{0.2}$$

where $\mu_i$, $\alpha_{i,1}$, $\beta_{i,1}$,..., $\alpha_{i,K}$, $\beta_{i,K}$ are unknown curve-specific coefficients. The coefficients for the $i$th curve are estimated by least squares assuming

$$Y_i(t_j) = f_i(t_j) + \epsilon_{ij}(1 \le j \le M) \tag{0.3}$$

is the model for the observed data where the $\epsilon_{ij}$ are uncorrelated, have mean zero and variance $\sigma_2$.

The vectors of fitted coefficients for the $n$ curves, $\boldsymbol{W}_i = (\hat{\mu}_i, \hat{\alpha}_{i,1}, \hat{\beta}_{i,1},..., \hat{\beta}_{i,K})^\top$ for $1 \le i \le n$, are regarded as independent draws from a $(2K+1)$-dimensional multivariate distribution with unknown mean vector ($\mu$, $\alpha_1$, $\beta_1$,..., $\beta_K$) and unknown covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{W}}$. Here $\top$ denotes the transpose of a vector or a matrix. The sample mean of the $\boldsymbol{W}_i$, $\bar{\boldsymbol{W}} = 1/n\ \Sigma_{i=1}^{n}\boldsymbol{W}_i$, is an estimate of the underlying true coefficient vector ($\mu$, $\alpha_1$, $\beta_1$,..., $\beta_K$). For example, the first component of $\bar{\boldsymbol{W}}$ is $\Sigma_{i=1}^{n}\hat{\mu}_i/n$ which is the average overall estimated mean. Similarly, the sample variance-covariance matrix of the $W_i$, defined as

$$\hat{\Sigma}_{\boldsymbol{W}} = \frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{W}_i - \bar{\boldsymbol{W}})^\top(\boldsymbol{W}_i - \bar{\boldsymbol{W}}),$$

is an estimator (biased) of the population variance-covariance $\boldsymbol{\Sigma}_{\boldsymbol{W}}$. Using these basic building blocks we estimate the following two quantities. The $i$th curve at $t$ is estimated by $\hat{f}_i(t) = \boldsymbol{W}_i^\top\ell(t)$ where

$$\ell(t) = \left(1, \cos\left(\frac{2\pi 1t}{T}\right), \sin\left(\frac{2\pi 1t}{T}\right),..., \cos\left(\frac{2\pi Kt}{T}\right) \sin\left(\frac{2\pi Kt}{T}\right)\right)$$

for $1 \le i \le n$. The true population curve at $t$, defined by Eq. (0.1), is estimated by

$$\hat{f}(t) = \bar{\boldsymbol{W}}^\top\ell(t) \tag{0.4}$$

We use

$$\hat{\sigma}_{\hat{f}(t)} = \sqrt{\ell(t)^\top\hat{\boldsymbol{\Sigma}}_{\boldsymbol{W}}\ell(t)}. \tag{0.5}$$

as a measure of the variability of $\hat{f}(t)$.

The prediction and confidence bands can be constructed as follows. Let $f_{n+1}(t)$ denote a future draw from the population mean curve that is defined by Eq. (0.1) and let $Y_{n+1}(t)$ be an observed curve generated by Eqs. (0.2) and (0.3). Given a desired confidence level $100(1-\alpha)\%$, we chose the constant $C_p$ so that

$$P\left\{\max_t\left(\frac{|\hat{f}_{n+1}(t) - \hat{f}(t)|}{\hat{\sigma}_{\hat{f}(t)}}\right) \le C_p\right\} = 1 - \alpha. \tag{0.6}$$

Then

$$\hat{f}(t) \pm C_p \times \hat{\sigma}_{\hat{f}(t)} \tag{0.7}$$

is $100(1-\alpha)\%$ prediction band over $[0, T]$ for a new curve.

Similarly if $C_c$ is chosen so that

$$P\left\{\max_t\left(\frac{|\hat{f}(t) - f(t)|}{\hat{\sigma}_{\hat{f}(t)}}\right) \le C_c\right\} = 1 - \alpha, \tag{0.8}$$

then

$$\hat{f}(t) \pm C_c \times \hat{\sigma}_{\hat{f}(t)} \tag{0.9}$$

is a $100(1-\alpha)\%$ confidence band for $f(t)$ over $[0, T]$.

The idea of the bootstrap is to chose $C_p$ and $C_c$ so that approximate versions of the probabilities Eqs. (0.6) and (0.8), respectively, are set equal to $1-\alpha$. These approximations are obtained by replacing the true stochastic mechanism by the empirical distribution formed from the population of curves.

In more detail, the following process is repeated $B$ times, say, where $B$ is large. In the $b$th cycle, $1 \le b \le B$, we select a sample of size $n$, with replacement, from the population of the $n$ original curves. Based on this pseudo-sample the estimator Eq. (0.4) is calculated; denote this quantity by $\hat{f}^b(t)$. We also calculate the spread estimate (0.5) based on the pseudo-sample; denote the corresponding value by $\hat{\sigma}_{\hat{f}^b(t)}$. Then the probability in Eq. (0.6) is approximately

$$\frac{1}{B}\times\sum_{b=1}^{B}\left[\frac{1}{n}\sum_{i=1}^{n}I\left(\max_t\left\{\frac{|\hat{f}_i(t) - \hat{f}^b(t)|}{\hat{\sigma}^b_{\hat{f}^b(t)}}\right\} \le C_p\right)\right] \tag{0.10}$$

where $I(E)$ is 1 or 0 according as event $E$ does or does not occur, respectively. In words, Eq. (0.10) is the average, over the bootstrap replications, of the proportion of the original data curves whose maximum standardized deviation from the bootstrap mean, $\hat{f}^b(t)$, is less than or equal $C_p$. If $C_p$ is chosen to make Eq. (0.10) equal to $1-\alpha$, then the prediction limits Eq. (0.7) will have approximate coverage $100(1-\alpha)\%$.

In a similar way

$$\frac{1}{B}\times\sum_{b=1}^{B}\left[\max_t\left\{\frac{|\hat{f}^b(t) - \hat{f}(t)|}{\hat{\sigma}^b_{\hat{f}^b(t)}}\right\} \le C_c\right] \tag{0.11}$$

is an approximation of the left-hand side probability in Eq. (0.8). Using Eq. (0.9) with $C_c$ computed to make Eq. (0.11) equal to $1-\alpha$ gives an approximate $100(1-\alpha)\%$ confidence band over $[0, T]$ for $f(t)$.

## References

[1] Sutherland DH, Olshen RA, Biden EN, Wyatt MP. The Development of Mature Walking. London: MacKeith Press, 1988.
[2] Olshen RA, Biden EN, Wyatt MP, Sutherland DH. Gait analysis and the bootstrap. Ann Statist 1989;17:1419–40.
[3] Bowker A, Lieberman G. Engineering Statistics, 2nd edition. Prentice Hall: New Jersey, 1972.

[4] Natrella M. Experimental Statistics, National Bureau of Standards Handbook 91. Washington, DC: Superintendent of Documents, US Government Printing Office, 1963: 20402.
[5] Hochberg Y, Tamhane A. Multiple Comparison Procedures. New York: Wiley, 1987.
[6] Efron B, Tibshirani RJ. An Introduction to the Bootstrap. London: Chapman and Hall, 1993.