

# Advanced Statistical Analysis of Biomechanical Time Series

Principal Components Analysis

Functional Data Analysis

Statistical Parametric Mapping

Drew Harrison

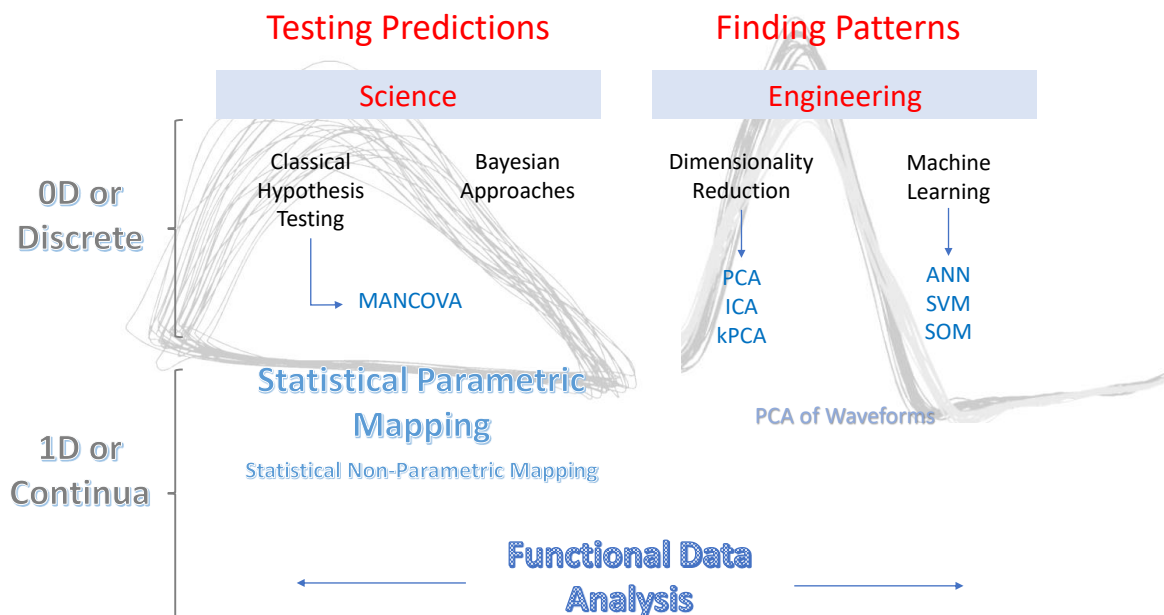
Physical Education and Sport Sciences, University of Limerick

John Warmenhoven

Exercise & Sports Science, University of Sydney

 @johnwarmenhoven

Australian Institute of Sport



## Functional Data Analysis (FDA)

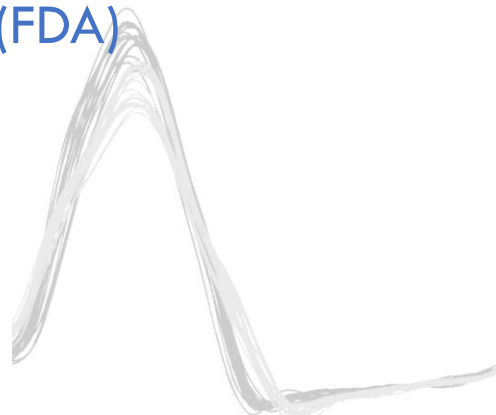
Data points on a time-series are considered as a single entity.

Assumed to be generated by some relatively smooth underlying function.

Involves fitting basis expansions to time-series data (B-Splines, Fourier, etc.).

High control over smoothing and registration of data if it is necessary.

Extrapolation to conventional multivariate statistical practices.



## Functional Data Analysis (FDA)

Data points on a time-series are considered as a single entity.

Assumed to be generated by some relatively smooth underlying function.

Involves fitting basis expansions to time-series data (B-Splines, Fourier, etc.).

High control over smoothing and registration of data if it is necessary.

Extrapolation to conventional multivariate statistical practices.

### *General FDA Steps:*

1. Fitting functions with the addition of smoothing using a roughness penalty.
2. Registration techniques using time-warping functions.
3. The implementation of statistical techniques such as hypothesis testing (t-tests, ANOVAs, regressions, etc.), or dimension reduction approaches (PCA, etc.).

Functional PCA (*f*PCA) is the most commonly used in biomechanics and human movement.

# Functional Data Analysis (FDA)

Number of approaches for fitting functions are available, and some common approaches are:

1. Fourier:
  - a) Stable functions (i.e. strong local features and where the curvature tends to be of the same order).
  - b) Periodicity of the signal be reflected to some degree in the data.
2. B-Splines:
  - a) Spline functions are the most common choice of approximation system for non-periodic functional data or parameters.
  - b) Derived from polynomials, with increased flexibility around unstable local behaviour.

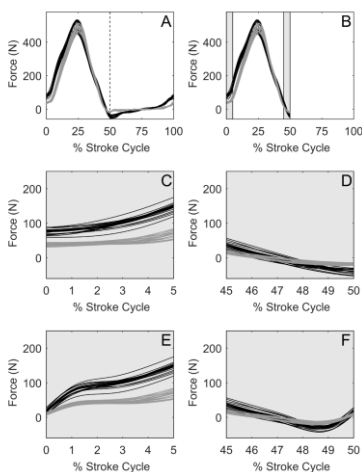
Fourier and B-Spline fitting processes are demonstrated in Warmenhoven et al. (2017).

## General FDA Steps:

1. Fitting functions with the addition of smoothing using a roughness penalty.
2. Registration techniques using time-warping functions.
3. The implementation of statistical techniques such as hypothesis testing (t-tests, ANOVAs, regressions, etc.), or dimension reduction approaches (PCA, etc.).

Functional PCA (*f*PCA) is the most commonly used in biomechanics and human movement.

# Functional Data Analysis (FDA)



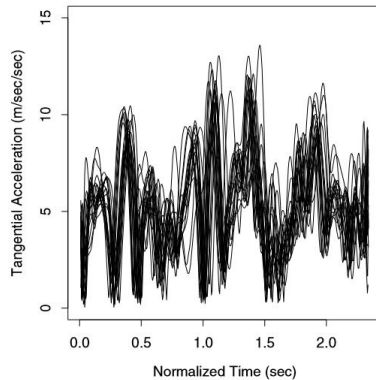
Fourier and B-Spline fitting processes are demonstrated in Warmenhoven et al. (2017).

## General FDA Steps:

1. Fitting functions with the addition of smoothing using a roughness penalty.
2. Registration techniques using time-warping functions.
3. The implementation of statistical techniques such as hypothesis testing (t-tests, ANOVAs, regressions, etc.), or dimension reduction approaches (PCA, etc.).

Functional PCA (*f*PCA) is the most commonly used in biomechanics and human movement.

# Functional Data Analysis (FDA)



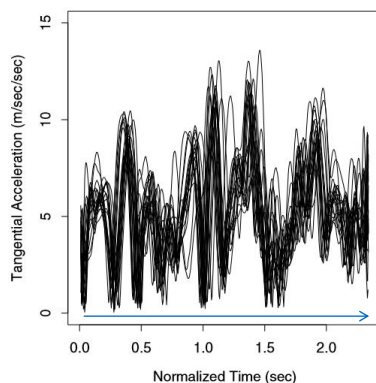
Handwriting data from Ramsay & Silverman (2002).

## General FDA Steps:

1. Fitting functions with the addition of smoothing using a roughness penalty.
2. Registration techniques using time-warping functions.
3. The implementation of statistical techniques such as hypothesis testing (t-tests, ANOVAs, regressions, etc.), or dimension reduction approaches (PCA, etc.).

Functional PCA ( $f$ PCA) is the most commonly used in biomechanics and human movement.

# Functional Data Analysis (FDA)



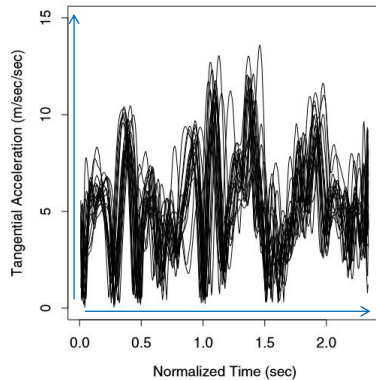
Handwriting data from Ramsay & Silverman (2002).

## General FDA Steps:

1. Fitting functions with the addition of smoothing using a roughness penalty.
2. Registration techniques using time-warping functions.
3. The implementation of statistical techniques such as hypothesis testing (t-tests, ANOVAs, regressions, etc.), or dimension reduction approaches (PCA, etc.).

Functional PCA ( $f$ PCA) is the most commonly used in biomechanics and human movement.

# Functional Data Analysis (FDA)



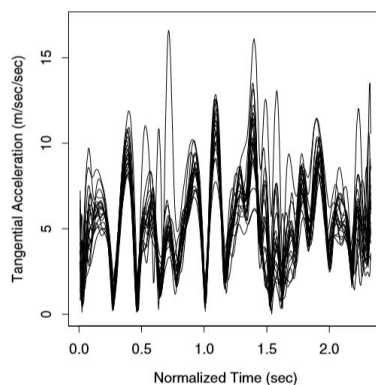
Handwriting data from Ramsay & Silverman (2002).

## General FDA Steps:

1. Fitting functions with the addition of smoothing using a roughness penalty.
2. Registration techniques using time-warping functions.
3. The implementation of statistical techniques such as hypothesis testing (t-tests, ANOVAs, regressions, etc.), or dimension reduction approaches (PCA, etc.).

Functional PCA ( $f$ PCA) is the most commonly used in biomechanics and human movement.

# Functional Data Analysis (FDA)



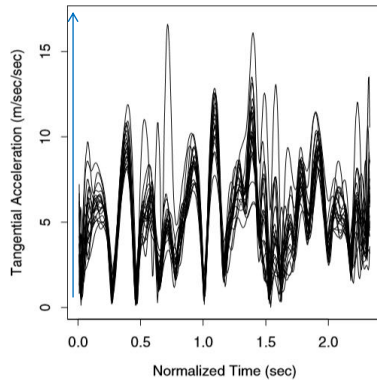
Handwriting data from Ramsay & Silverman (2002).

## General FDA Steps:

1. Fitting functions with the addition of smoothing using a roughness penalty.
2. Registration techniques using time-warping functions.
3. The implementation of statistical techniques such as hypothesis testing (t-tests, ANOVAs, regressions, etc.), or dimension reduction approaches (PCA, etc.).

Functional PCA ( $f$ PCA) is the most commonly used in biomechanics and human movement.

# Functional Data Analysis (FDA)

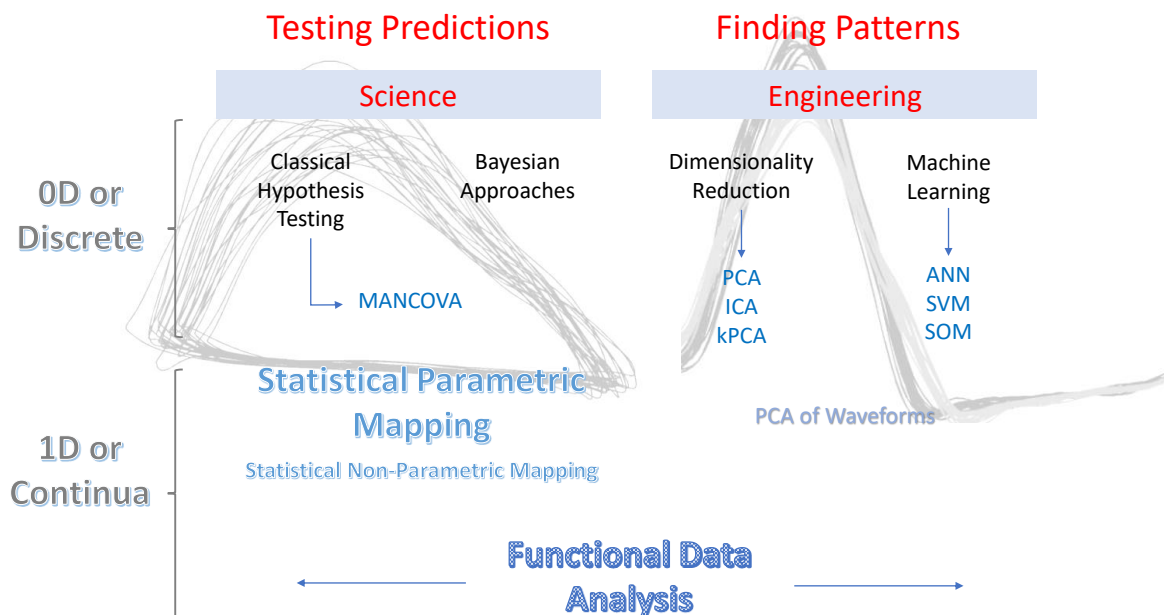


Handwriting data from Ramsay & Silverman (2002).

## General FDA Steps:

1. Fitting functions with the addition of smoothing using a roughness penalty.
2. Registration techniques using time-warping functions.
3. The implementation of statistical techniques such as hypothesis testing (t-tests, ANOVAs, regressions, etc.), or dimension reduction approaches (PCA, etc.).

Functional PCA (fPCA) is the most commonly used in biomechanics and human movement.



# FDA Techniques

## Finding patterns and exploring variability

- Functional Principal Components Analysis (fPCA)
- Functional Canonical Correlation Analysis (fCCA)

## Testing predictions

- Functional Linear Models (FLM)
  - Functional Predictor > Scalar Outcome
  - Scalar Predictor > Functional Outcome
  - Functional Predictor > Functional Outcome
- Principal Differential Analysis (PDA)

Ramsay & Silverman (2005)

Ramsay & Silverman (2002)

Ramsay, Hooker & Graves (2009)



# FDA Techniques

## Finding patterns and exploring variability

- Functional Principal Components Analysis (fPCA)
- Functional Canonical Correlation Analysis (fCCA)

## Testing predictions

- Functional Linear Models (FLM)
  - Functional Predictor > Scalar Outcome
  - Scalar Predictor > Functional Outcome
  - Functional Predictor > Functional Outcome
- Principal Differential Analysis (PDA)





## FDA Techniques

### Finding patterns and exploring variability

- Functional Principal Components Analysis (fPCA)
- Functional Canonical Correlation Analysis (fCCA)

### Testing predictions

- Functional Linear Models (FLM)
  - Functional Predictor > Scalar Outcome
  - Scalar Predictor > Functional Outcome
  - Functional Predictor > Functional Outcome
- Principal Differential Analysis (PDA)

Displays dominant modes of variation in data.

Provides outputs relative to each mode of variation that allow for patterns to be tested relative to research questions.

This is the most common FDA technique used in applied in biomechanics, with use in:

- **Jumping** (Ryan, Harrison & Hayes, 2007).
- **Race-walking** (Dona, Preatoni, Cobelli, Rodano & Harrison, 2009).
- **Rowing** (Warmenhoven et al., 2017)
- **Gait** (Donoghue, Harrison, Coffey, & Hayes, 2008).
- **Weightlifting** (Kipp & Harris, 2014).

## FDA Techniques

### Finding patterns and exploring variability

- Functional Principal Components Analysis (fPCA)
- Functional Canonical Correlation Analysis (fCCA)

### Testing predictions

- Functional Linear Models (FLM)
  - Functional Predictor > Scalar Outcome
  - Scalar Predictor > Functional Outcome
  - Functional Predictor > Functional Outcome
- Principal Differential Analysis (PDA)

Identifies components of variability present in each of two sets of observations, which are highly correlated with one another.

Great potential for coordination research.

Relatively underutilised in biomechanics research.

A good example of biomechanical application is Leurgans, Moyeed & Silverman (1993), with use in paediatric gait analysis.



## FDA Techniques

### Finding patterns and exploring variability

- Functional Principal Components Analysis (*f*PCA)
- Functional Canonical Correlation Analysis (*f*CCA)

### Testing predictions

- **Functional Linear Models (FLM)**
  - Functional Predictor > Scalar Outcome
  - Scalar Predictor > Functional Outcome
  - Functional Predictor > Functional Outcome
- Principal Differential Analysis (PDA)

Applying the general linear model to functional data.

These describe the relationship between outcome variables and an explanatory variables.

In biomechanics research the following techniques have been applied:

*Functional t-tests* in ACL jumping research (Baumgart, Hoppe & Freiwald, 2017).

*Functional ANOVAs* in gait under pain effusion (Park, Seeley, Francom, Reese & Hopkins, 2017).

*Functional Regressions* in gait exploring the effects of walking velocity (Dura et al., 2010).

## FDA Techniques

### Finding patterns and exploring variability

- Functional Principal Components Analysis (*f*PCA)
- Functional Canonical Correlation Analysis (*f*CCA)

### Testing predictions

- **Functional Linear Models (FLM)**
  - Functional Predictor > Scalar Outcome
  - Scalar Predictor > Functional Outcome
  - Functional Predictor > Functional Outcome
- **Principal Differential Analysis (PDA)**

Has applications to both “*finding patterns*” and “*testing predictions*”

In the context of dimension reduction:

- *f*PCA looks for linear differential operators to explain variation between curves.
- PDA looks for linear differential operators to explain variation between derivatives of curves.

For FLMs, a PDA model looks for a linear differential operator to represent covariation between a variable ( $x$ ) and its derivative ( $Dx$ ).

This is largely underexplored in the context of biomechanical research themes and contexts.

# FDA Techniques

## Extensions beyond Ramsay & Silverman...



### Review

A force profile analysis comparison between functional data analysis, statistical parametric mapping and statistical non-parametric mapping in on-water single sculling

John Warmenhoven<sup>a,\*</sup>, Andrew Harrison<sup>b</sup>, Mark A. Robinson<sup>c</sup>, Jos Vanrenterghem<sup>d</sup>, Norma Bargary<sup>e</sup>, Richard Smith<sup>f</sup>, Stephen Cobley<sup>g</sup>, Conny Draper<sup>h</sup>, Cyril Donnelly<sup>i</sup>, Todd Pataky<sup>j</sup>

<sup>a</sup> Exercise and Sports Science, University of Sydney, Australia

<sup>b</sup> Physical Education and Sport Sciences, University of Limerick, Ireland

<sup>c</sup> Research Institute for Sport and Exercise Sciences, Liverpool John Moores University, United Kingdom

<sup>d</sup> Department of Rehabilitation Sciences, KU Leuven, Belgium

<sup>e</sup> Department of Mathematics and Statistics, University of Limerick, Ireland

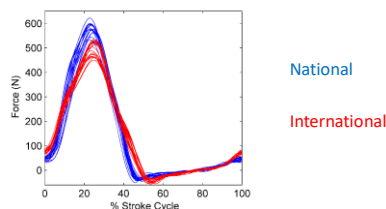
<sup>f</sup> School of Human Sciences (Exercise and Sport Science), University of Western Australia, Australia

<sup>g</sup> Department of Health and Human Sciences, Kyoto University, Japan

Finally, it should also be acknowledged that there are advancements in FDA beyond the scope of the FDA technique applied in the present study.<sup>34</sup> From the perspective of FDA hypothesis testing techniques it appears that there are two main approaches (parametric and non-parametric), which fall within basis function approximation methods and overall tests. With reference to procedures concerned with testing the equality of coefficients from a basis function approximation, parametric methods include the works of Fan and Li,<sup>35</sup> Cuevas et al.,<sup>36</sup> and Spitzner<sup>37</sup>; and nonparametric methods include the work of Zhang and Chen,<sup>38</sup> Delgado,<sup>39</sup> and Cao et al.<sup>40</sup> The FDA t-test used in the present study<sup>28</sup> was explored due to its implementation with biomechanical data in experimental human movement research, and also the ease with which software can be accessed by applied clinicians and researchers from the FDA website.

## Finding Patterns with PCA and fPCA

Data used to describe these techniques come from sample data set one in Warmenhoven et al. (2017).



PCA of Waveforms and fPCA can be described together...

### Step

- Representation of data
- Smoothing
- Control for phase variation

### PCA of Waveforms

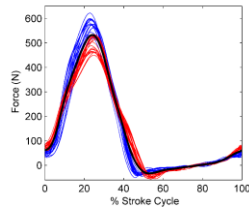
- Data points
- Filtering as a part of pre-processing
- Piecewise spline interpolation

### fPCA

- Coefficients of a function
- Smoothing parameter applied during function fitting
- Curve registration using time warping functions

# Finding Patterns with PCA and fPCA

Data used to describe these techniques come from sample data set one in Warmenhoven et al. (2017).



PCA of Waveforms and fPCA can be described together...

## Step

Mean of curves

## PCA of Waveforms

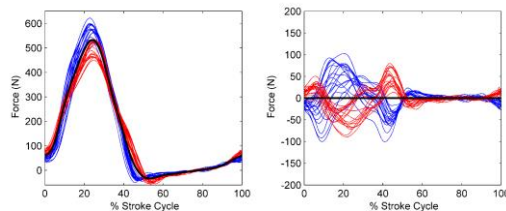
Data points

## fPCA

Coefficients of a function

# Finding Patterns with PCA and fPCA

Data used to describe these techniques come from sample data set one in Warmenhoven et al. (2017).



Variance matrix calculated as a step in deriving the covariance matrix

PCA of Waveforms and fPCA can be described together...

## Step

Mean of curves

Variability structures in the original data

## PCA of Waveforms

Data points

Covariance matrix

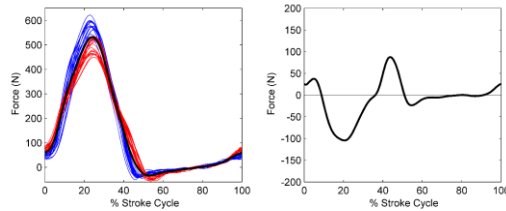
## fPCA

Coefficients of a function

Covariance function

# Finding Patterns with PCA and fPCA

Data used to describe these techniques come from sample data set one in Warmenhoven et al. (2017).

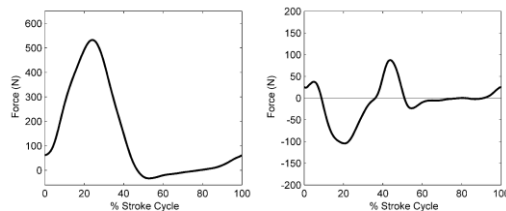


PCA of Waveforms and fPCA can be described together...

| Step                      | PCA of Waveforms                                     | fPCA  |
|---------------------------|--|---|
| Orthogonal decomposition  | Eigenvectors (PCs)                                   | Eigenfunctions (fPCs)                                 |
| Proportion of variability | Eigenvalues used to calculate % variation of each PC | Eigenvalues used to calculate % variation of each fPC |

# Finding Patterns with PCA and fPCA

Data used to describe these techniques come from sample data set one in Warmenhoven et al. (2017).

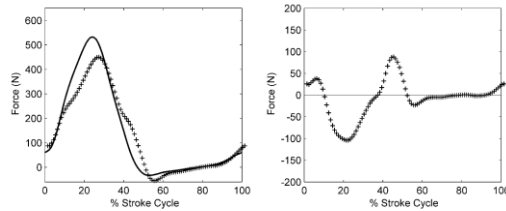


Interpreting variability that is shown by PCA of Waveforms and fPCA...

1. Retain the mean vector (PCA) or function (fPCA) from the original data.

## Finding Patterns with PCA and fPCA

Data used to describe these techniques come from sample data set one in Warmenhoven et al. (2017).

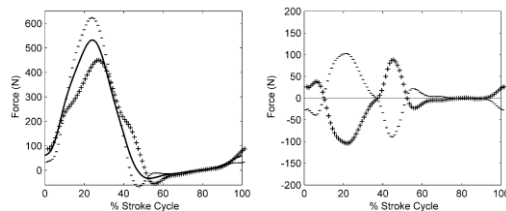


Interpreting variability that is shown by PCA of Waveforms and fPCA...

1. Retain the mean vector (PCA) or function (fPCA) from the original data.
2. Add...

## Finding Patterns with PCA and fPCA

Data used to describe these techniques come from sample data set one in Warmenhoven et al. (2017).



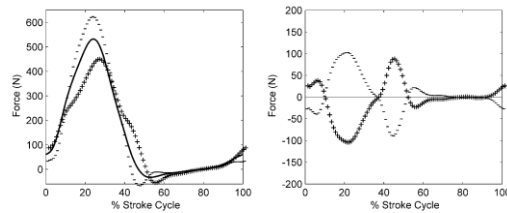
This process of visualization can be implemented in an identical way for both PCA and fPCA.

Interpreting variability that is shown by PCA of Waveforms and fPCA...

1. Retain the mean vector (PCA) or function (fPCA) from the original data.
2. Add... and subtract the PC vector or fPC function from the mean.
3. Plot the addition of the PC/fPC intuitively using "+" and the subtraction using "-" symbols.
4. A constant (c) can be used to scale the size of each PC or fPC when plotted relative to the mean.

# Finding Patterns with PCA and fPCA

Data used to describe these techniques come from sample data set one in Warmenhoven et al. (2017).



PCA of Waveforms and fPCA can be described together...

## Step

Weights relative to PCs/fPCs

## PCA of Waveforms

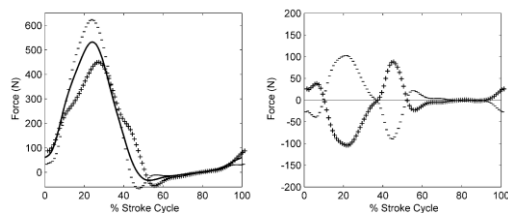
PC scores

## fPCA

fPC scores

# Finding Patterns with PCA and fPCA

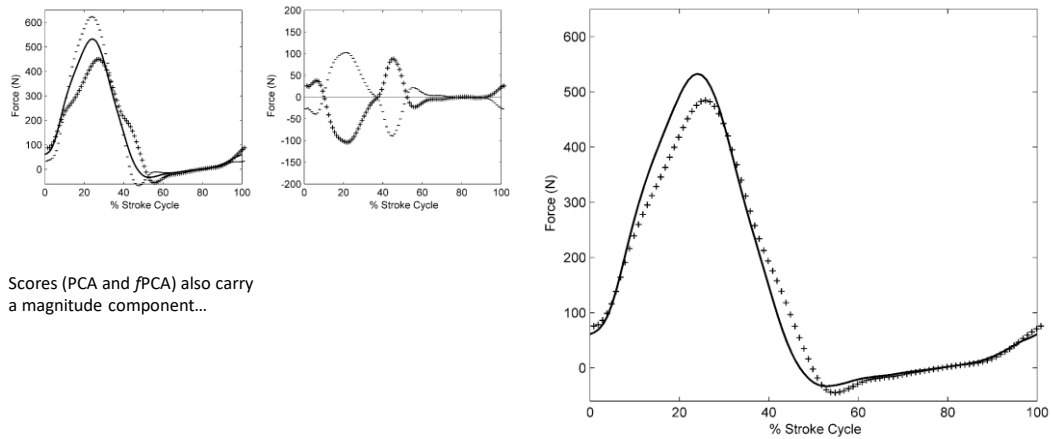
Data used to describe these techniques come from sample data set one in Warmenhoven et al. (2017).



Scores (PCA and fPCA) also carry a magnitude component...

## Finding Patterns with PCA and fPCA

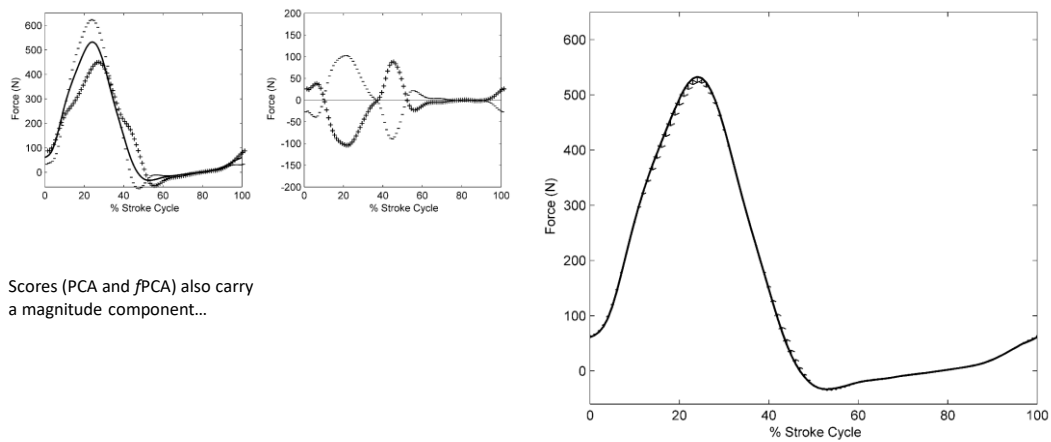
Data used to describe these techniques come from sample data set one in Warmenhoven et al. (2017).



Scores (PCA and fPCA) also carry a magnitude component...

## Finding Patterns with PCA and fPCA

Data used to describe these techniques come from sample data set one in Warmenhoven et al. (2017).

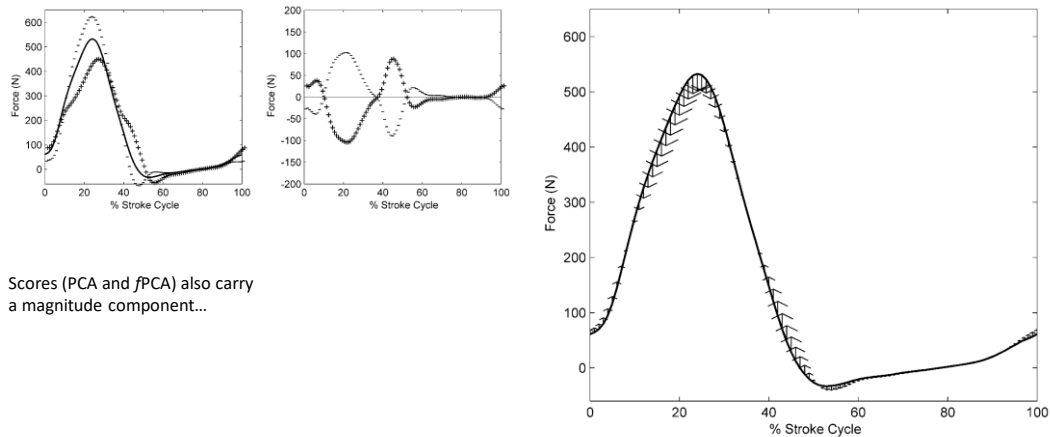


Scores (PCA and fPCA) also carry a magnitude component...



## Finding Patterns with PCA and *f*PCA

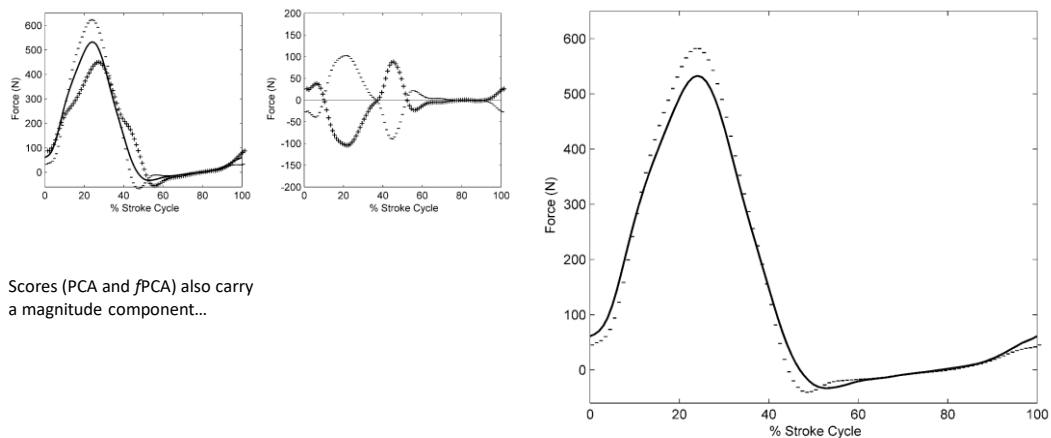
Data used to describe these techniques come from sample data set one in Warmenhoven et al. (2017).



Scores (PCA and *f*PCA) also carry a magnitude component...

## Finding Patterns with PCA and *f*PCA

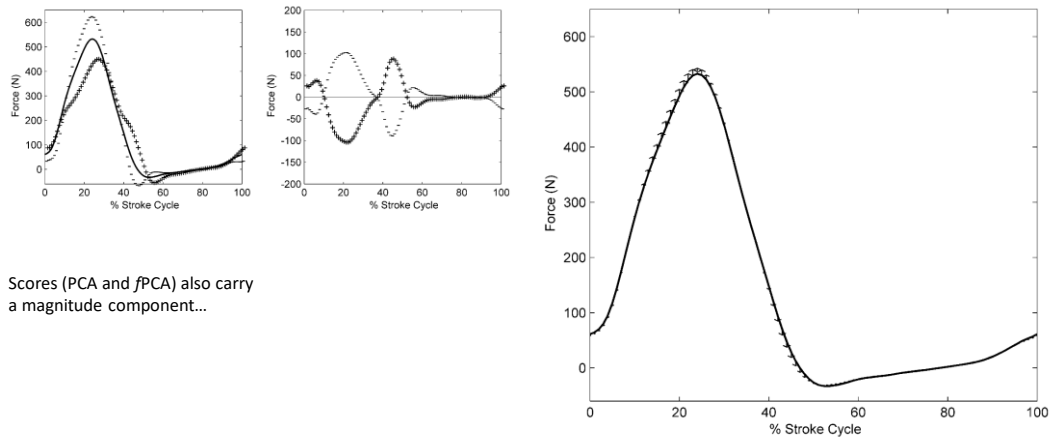
Data used to describe these techniques come from sample data set one in Warmenhoven et al. (2017).



Scores (PCA and *f*PCA) also carry a magnitude component...

## Finding Patterns with PCA and fPCA

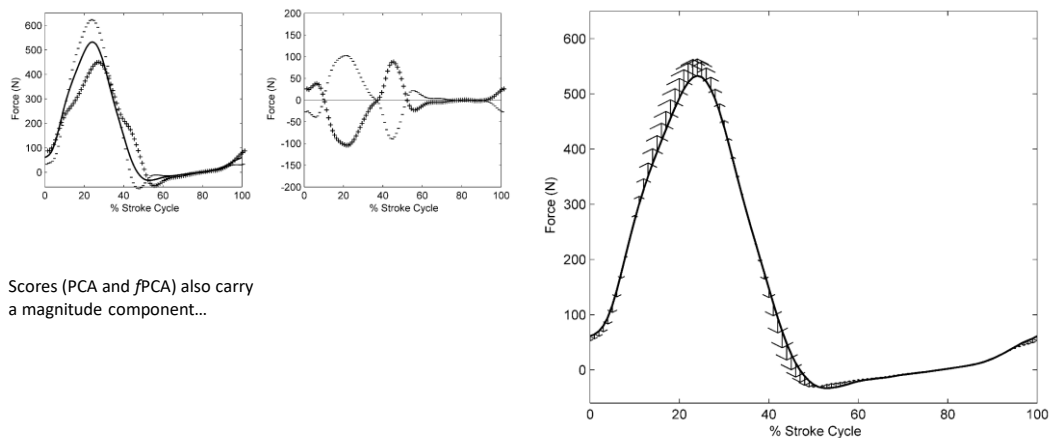
Data used to describe these techniques come from sample data set one in Warmenhoven et al. (2017).



Scores (PCA and fPCA) also carry a magnitude component...

## Finding Patterns with PCA and fPCA

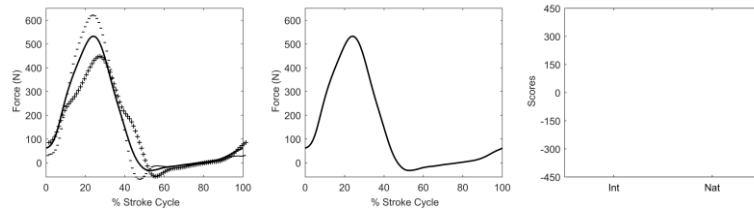
Data used to describe these techniques come from sample data set one in Warmenhoven et al. (2017).



Scores (PCA and fPCA) also carry a magnitude component...

# Finding Patterns with PCA and fPCA

Data used to describe these techniques come from sample data set one in Warmenhoven et al. (2017).



PCA of Waveforms and fPCA can be described together...

Step

Weights relative to PCs/fPCs

PCA of Waveforms

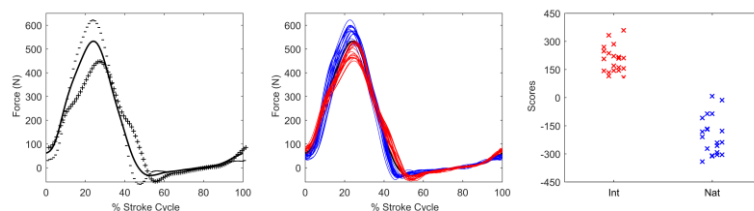
PC scores

fPCA

fPC scores

# Finding Patterns with PCA and fPCA

Data used to describe these techniques come from sample data set one in Warmenhoven et al. (2017).



PCA of Waveforms and fPCA can be described together...

Step

Weights relative to PCs/fPCs

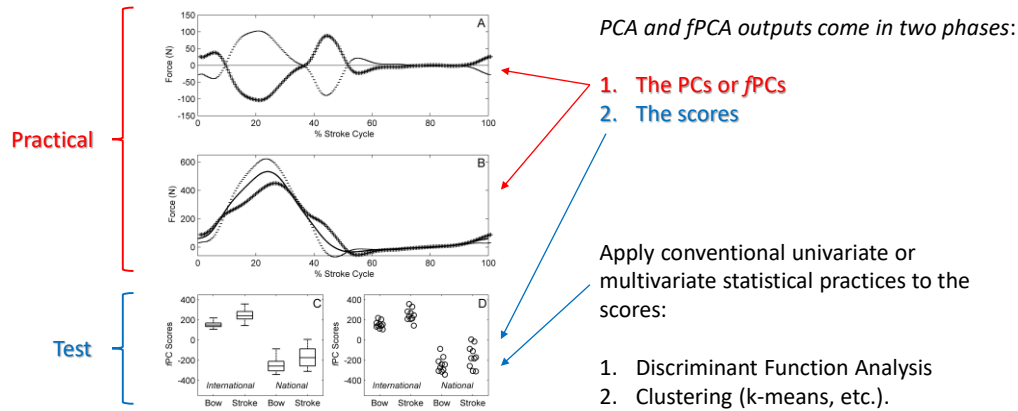
PCA of Waveforms

PC scores

fPCA

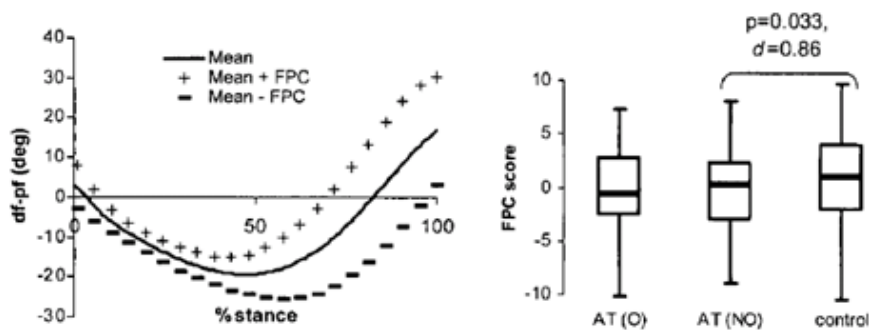
fPC scores

## Finding Patterns with PCA and fPCA



This is Figure 1 from Warmenhoven et al. (2017).

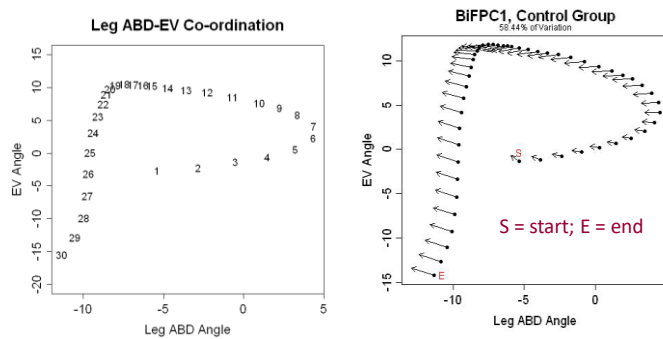
## Group comparisons of fPCs



Ankle dorsi-flexion on stance phase in running, analysing fPC scores for injured subjects wearing orthotics, [AT(O)], without orthotics, [AT(NO)] and uninjured limb [control].

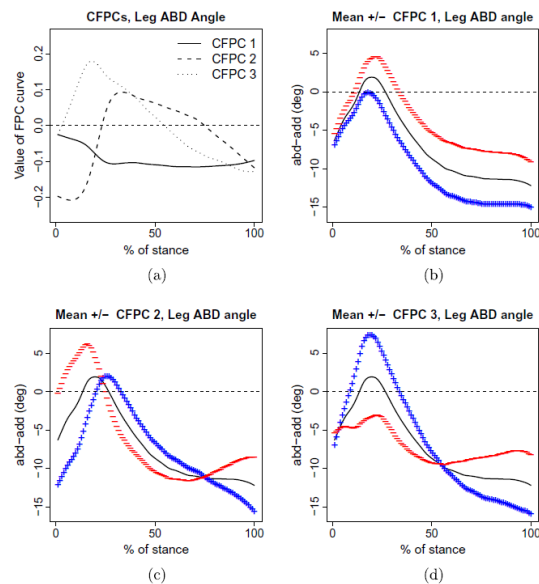
Coffey et al. (2011).

## Depicting and analysing coordination: Bivariate *f*PCA



At A, the ensemble average angle-angle plot of the leg ABD vs EV coordination is plotted. B. Shows the bivariate *f*PC as a vector relative to the ensemble mean function.

## Common *f*PCA for repeated measures analysis



## Common *f*PCA for repeated measures analysis

| Component | AT(NO)            |       | Control           |       |
|-----------|-------------------|-------|-------------------|-------|
|           | CFPC              | FPC   | CFPC              | FPC   |
| 1         | 10.46<br>(69.26%) | 11.03 | 36.34<br>(90.42%) | 36.40 |
| 2         | 1.62<br>(10.75%)  | 1.56  | 2.81<br>(7.00%)   | 2.86  |
| 3         | 2.64<br>(17.50%)  | 2.21  | 0.52<br>(1.30%)   | 0.56  |

- *Cf*PCA provides information on the factors that influence variation across groups and the amount of total variation accounted by each factor
- In the injured group [AT(NO)] the same factors influence movement but with *Cf*PCA shows minimal variation accounted by *Cf*PC3 in control group.

## Functional Linear Models

Difficult to derive a theoretical null distribution for any given test statistic for functional data (Ramsay, Hooker & Graves, 2009)...

Need to account for:

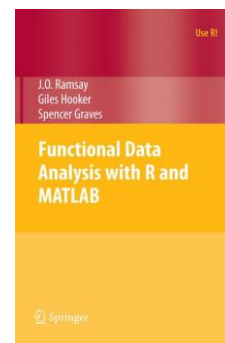
- Selection of a smoothing parameter
- the smoothing itself

Package employs a **permutation test** methodology.

This involves constructing a null distribution from the observed data directly.

Executed by rearranging the vector of responses while keeping the covariates in the same order and trying to fit the model again.

Ramsay, Hooker & Graves (2009)



The advantage of this is that we no longer need to rely on distributional assumptions.

## Functional t-test

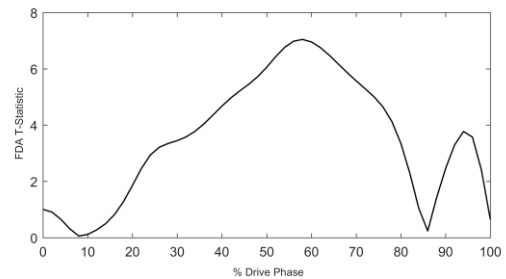
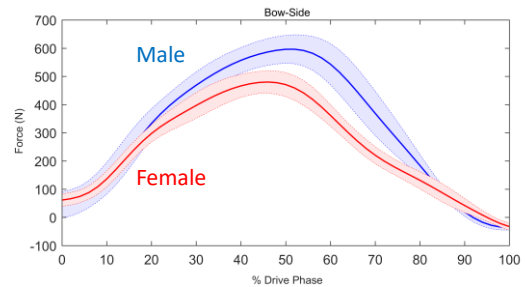
Used to assess the simple experimental scenario of whether a difference exists between **two independent groups**.

Data taken from Warmenhoven et al. (2018) and looks at differences in gender for characteristics of force application in rowing.

After function fitting, smoothing and registration... functional **descriptive statistics** can be plotted and inspected relative to the results of the t-test.

Creation of a pointwise absolute t-statistic...

- Provides a sense of the relative separation of the two groups of functions.

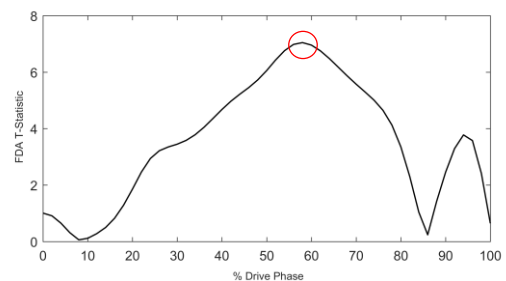
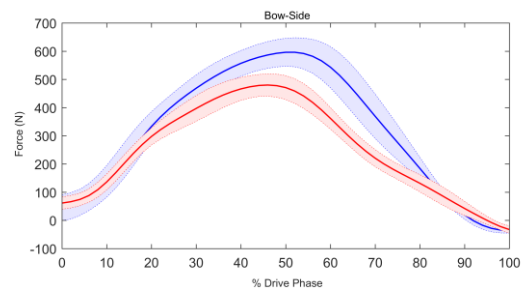


## Functional t-test

A formal hypothesis test requires:

- value or statistic to test, and;
- probability value indicating the result of the test

The test statistic used is the maximum value of the multivariate T-test,  $T(t)$ .





## Functional t-test

A formal hypothesis test requires:

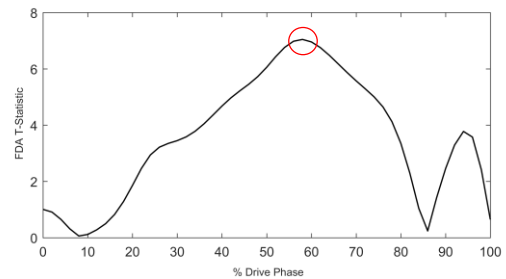
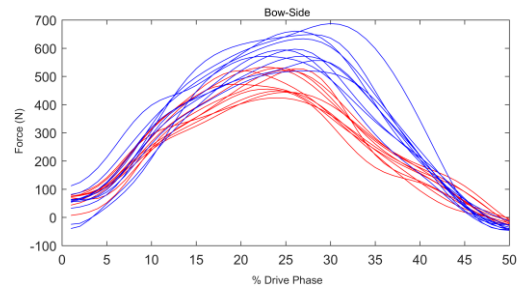
- value or statistic to test, and;
- probability value indicating the result of the test

The test statistic used is the maximum value of the multivariate T-test,  $T(t)$ .

To find a critical value of this statistic:

- randomly shuffle the labels of the curves
- recalculate the maximum of  $T(t)$  with the new labels

Repeating this many times allows a null distribution to be constructed.



## Functional t-test

A formal hypothesis test requires:

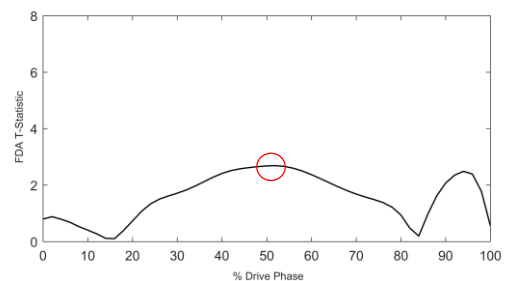
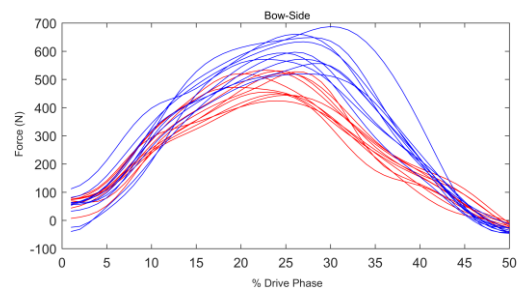
- value or statistic to test, and;
- probability value indicating the result of the test

The test statistic used is the maximum value of the multivariate T-test,  $T(t)$ .

To find a critical value of this statistic:

- randomly shuffle the labels of the curves
- recalculate the maximum of  $T(t)$  with the new labels

Repeating this many times allows a null distribution to be constructed.



## Functional t-test

A formal hypothesis test requires:

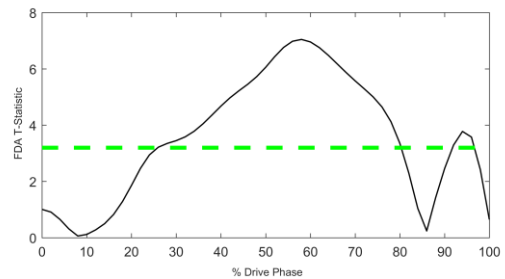
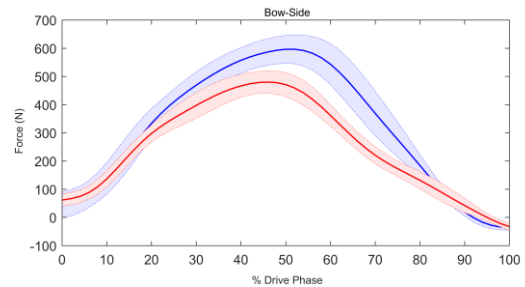
- value or statistic to test, and;
- probability value indicating the result of the test

The test statistic used is the maximum value of the multivariate T-test,  $T(t)$ .

To find a critical value of this statistic:

- randomly shuffle the labels of the curves
- recalculate the maximum of  $T(t)$  with the new labels

Repeating this many times allows a null distribution to be constructed.



## Functional t-test

A formal hypothesis test requires:

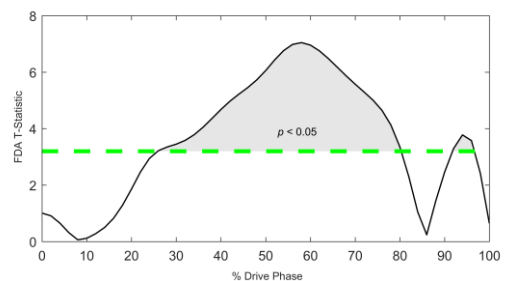
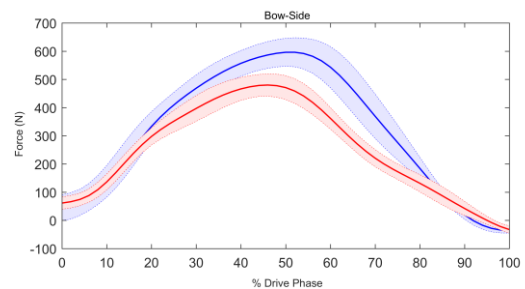
- value or statistic to test, and;
- probability value indicating the result of the test

The test statistic used is the maximum value of the multivariate T-test,  $T(t)$ .

To find a critical value of this statistic:

- randomly shuffle the labels of the curves
- recalculate the maximum of  $T(t)$  with the new labels

Repeating this many times allows a null distribution to be constructed.



## References

- Baumgart, C., Hoppe, M. W., & Freiwald, J. (2017). Phase-Specific Ground Reaction Force Analyses of Bilateral and Unilateral Jumps in Patients With ACL Reconstruction. *Orthopaedic Journal of Sports Medicine*, 5(6), 2325967117710912.
- Dona, G., Preatoni, E., Cobelli, C., Rodano, R., & Harrison, A. J. (2009). Application of functional principal component analysis in race walking: an emerging methodology. *Sports Biomechanics*, 8(4), 284-301.
- Donoghue, O. A., Harrison, A. J., Coffey, N., & 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.
- Durá, J. V., Belda, J. M., Poveda, R., Page, Á., Laparra, J., Das, J., . . . García, A. C. (2010). Comparison of functional regression and nonfunctional regression approaches to the study of the walking velocity effect in force platform measures. *Journal of applied biomechanics*, 26(2), 234-239.
- Kipp, K., & Harris, C. (2014). Patterns of barbell acceleration during the snatch in weightlifting competition. *Journal of sports sciences*, 33(14), 1467-1471.
- Leurgans, S. E., Moyeed, R. A., & Silverman, B. W. (1993). Canonical correlation analysis when the data are curves. *Journal of the Royal Statistical Society. Series B (Methodological)*, 725-740.
- Park, J., Seeley, M. K., Francom, D., Reese, C. S., & Hopkins, J. T. (2017). Functional vs. Traditional Analysis in Biomechanical Gait Data: An Alternative Statistical Approach. *Journal of human kinetics*, 60(1), 39-49.
- Ramsay, J. O., Hooker, G., & Graves, S. (2009). *Functional data analysis with R and MATLAB*: Springer Science & Business Media.
- Ramsay, J. O., & Silverman, B. W. (2002). *Applied functional data analysis: methods and case studies* (Vol. 77). New York: Springer.
- Ramsay, J. O., & Silverman, B. W. (2005). *Functional data analysis*: Wiley Online Library.
- Ryan, W., Harrison, A., & Hayes, K. (2006). Functional data analysis of knee joint kinematics in the vertical jump. *Sports Biomechanics*, 5(1), 121-138.
- Warmenhoven, J., Cobley, S., Draper, C., Harrison, A. J., Bargary, N., & Smith, R. (2017). Considerations for the use of functional principal components analysis (fPCA) in sports biomechanics: examples from on-water rowing. *Sports Biomechanics*, In Press.
- Warmenhoven, J., Harrison, A. J., Robinson, M. A., Vanrenterghem, J., Bargary, N., Smith, R., . . . Pataky, T. (2017). A comparison of functional data analysis and statistical parametric mapping in sports biomechanics: The independent samples t-test. *Journal of Science & Medicine in Sport*, In Press.