DATA REDUCTION AS A NECESSARY TOOL IN BIOMECHANICAL KINEMATIC DATA ANALYSIS

Joseph O'Halloran, Orna Donoghue and Ross Anderson

Biomechanics Research Unit, PESS Department, University of Limerick, Ireland

The purpose of this study was to evaluate four possible methods of data reduction for the subsequent use of the data with an Artificial Neural Network (ANN). Functional data analysis and dynamical systems theory approaches were used to investigate the kinematic dynamics of gait and rowing movements. Five rowing participants completed a 2000m row on a RowPerfect ergometer and 24 gait participants completed treadmill running. The results indicate that each of the four methods provides possibilities for use in an ANN by utilising data reduction. In particular the continuous relative phase, with its use of two joint position and velocity, can compress four variables into one, and can maintain the trends of the data.

KEY WORDS: data reduction, dynamical systems theory, functional data analysis

INTRODUCTION:

The purpose of this study was to examine the current methods of data reduction being utilised within biomechanics with the subsequent goal of employing the method deemed most suitable in an Artificial Neural Network (ANN). The use of ANNs has previously been shown to predict power output of a rowing stroke to within 21 Watts (O'Halloran & Anderson, 2006). However a difficulty encountered during this research was the high dimensionality of the input data. It is hypothesised that if the dimensionality of the data can be successfully reduced without reducing any of the essential components of the signal then this data can be used in conjunction with an unsupervised ANN to optimise the rowing stroke.
Modern technologies allow biomechanists to collect vast quantities of data, at high rates for increasingly longer periods of time. This collection of additional data allows us to gain a more holistic view of the activity. Ideally, the entire data set of each component measured would be used; however this becomes impractical and has led to lack of clarity in the output. Thus, the objectives of the paper are to investigate the current methods used in the analysis of kinematic data, assess the usefulness of each of these methods and propose the avenue to take for successful input to an unsupervised ANN. The methods considered are functional data analysis (FDA), continuous relative phase (CRP), vector coding (VC), and cross correlation (CC).

METHOD:

Data Collection:

The data was collected in two distinct phases; one set of rowing data, and one set of gait data. Prior to the recording of both sets of data, the participants completed an informed consent form and pre-test questionnaire and received an individualised information sheet. Participants were familiarised with the testing procedures and any possible risks were outlined. Ethical approval was obtained from the University Research Ethics Committee.
The rowing data was captured from five rowers (age 26±4.6 yrs; height 173±5 cm; weight 74±4.3 kg). The participants were asked to perform a 2000m row on a RowPerfect ergometer (RowPerfect, CARE RowPerfect, The Netherlands). This distance was completed in 326±78 strokes. The participants' kinematic data was captured at 200Hz (Motion Analysis Inc, California, USA). Reflective body markers were placed on the fifth metatarsal, lateral malleolus of the tibia, lateral condyle of the femur, greater trochanter of the femur, acromion process, lateral epicondyle of the humerus and on the styloid process of the ulna. From these joint markers five joint angles for the entire movement could be identified.
The gait data was captured from twelve subjects with a history of chronic, low-grade achilles tendon (AT) injury (1 female, 11 male, age: 39±8.1 yrs; height: 175±5cm; weight: 73±8.5 kg)

and twelve control subjects (1 female, 11 male, age: 44±8.4 yrs; height: 178±5cm; weight: 79±12.2 kg). All AT subjects displayed levels of pronation during running, which the collaborating podiatrist judged likely to be related to the clinical presentation of AT injury. These subjects were provided with custom-made orthoses in the year prior to testing. Three-dimensional rearfoot and lower limb kinematics were obtained in orthoses (O) and no orthoses (NO) conditions during treadmill running at self-selected speeds. Control subjects participated in the no orthoses condition only. This provided three groups of data: AT(O), AT(NO) and controls.

Data Analysis:
Four methods of data analysis were used. The gait data was analysed using FDA and VC techniques. The rowing data was analysed using CRP and CC techniques.
Some of the techniques used in this study come from methods based on dynamical systems theory (Hamill *et al.*, 1999). These methods typically incorporate multiple component variables into the final measure (Hamill, 2006). Angle-angle diagrams provide information about the movement of one segment relative to another segment; however this data is limited to qualitative interpretation. Hamill *et al*. (2000) and Heiderscheit *et al.* (2002) proposed a modified VC technique based on the work of Sparrow *et al*. (1987), which converts an angle-angle diagram to a position-time plot. The orientation of the vector between two successive data points is calculated according to the formula below and continued for all data points of the angle-angle plot.

$$\theta = \tan^{-1}\left(\frac{y_2 - y_1}{x_2 - x_1}\right)$$

where: y_1, y_2 = *successive coordinates of angle on y-axis*

x_1, x_2 = *successive coordinates of angle on x-axis*

This provides a continuous coupling angle curve across stance which can be used to determine quantitatively the degree of coordination between two segments. Variability during the various stages of stance can then be calculated using circular statistics (Wheat and Glazier, 2006). CC is a method by which the degree of similarity between two sets of numbers can be quantified. The procedure is very simple, yet the general concept is used in a variety of advanced analysis techniques. These techniques are all based on the fact that if one carries out a point by point multiplication of two data sets, the sum of the products will be a quantification of their relationship. The cross correlation can be calculated using the following formula:

$$r_{xy} = \sum_{i=0}^{N-1} (x_i / \bar{x})(y_i / \bar{y})$$

where: x_i = *angle 1 at i^th^ point;* $\bar{x}$ *=mean angle 1*

y_i = *angle 2 at i^th^ point;* $\bar{y}$ *=mean angle 2*

FDA is an advanced statistical method of detecting patterns in multidimensional signals. An essential part of FDA is the application of principal component analysis (PCA) techniques to functional data. PCA is used to reduce the dimensionality of a problem by extracting relevant information from high-dimensional data sets. It does this by examining the relevant principal components that describe most of the variance in the signal. FDA has advantages as it identifies, from the entire dataset, the uncorrelated functions that describe the patterns of movement, and it is suitable for time series and coordination data. Daffertshoffer *et al.* (2004) suggested that examining the patterns of movement is a more appropriate approach in biomechanical analysis of many tasks; however it has had limited use in biomechanics research to date. FDA involved several steps that have been outlined previously in Donoghue *et al*. (2006). CRP is a method of examining coordination. It does so by means of in-phase and out-of-phase movements. Over the past 20 years, a growing body of literature has purported that these types of analytical techniques are more sensitive than basic kinematic or kinetic parameters at identifying the coordination of a musculo skeletal system (Hamill, 2006). The use of CRP to examine coupling relationships is based on kinematic data obtained across the entire movement.

RESULTS AND DISCUSSION:

A representation output of each method is presented ; first the gait data then the rowing data. Figure 1 shows the ankle eversion-ankle dorsi flexion (EV-ADF) VC angles across stance for representative subjects. The VC angle can be used to provide a means for coupling two joints together across the entire movement. Potential disadvantages are that VC provides spatial information but not temporal information and may pose problems when there is a change in direction of joint motion (Wheat and Glazier, 2006). The use of angle-angle diagrams rather than phase plane plots in VC is advantageous as it allows a more intuitive interpretation of the data and does not require vigorous normalisation procedures compared to other approaches such as CRP.

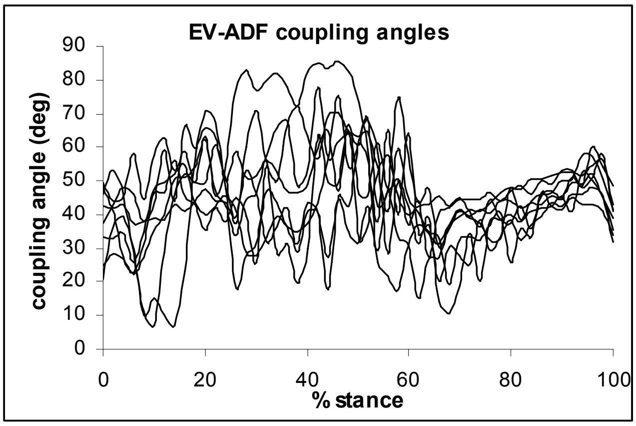


Figure 1 EV-ADF coupling angles across stance for representative subjects

During FDA analysis four functional principal components (FPCs) were extracted for each of angles; each reflecting a distinct feature of the angle-time curve. FDA relies mainly on graphical presentation and due to space constraints, representative data for just one of these FPCs is presented here. A useful technique to interpret the FPCs is to examine plots of the overall mean function and the functions obtained by adding and subtracting a suitable multiple of the FPC in question.

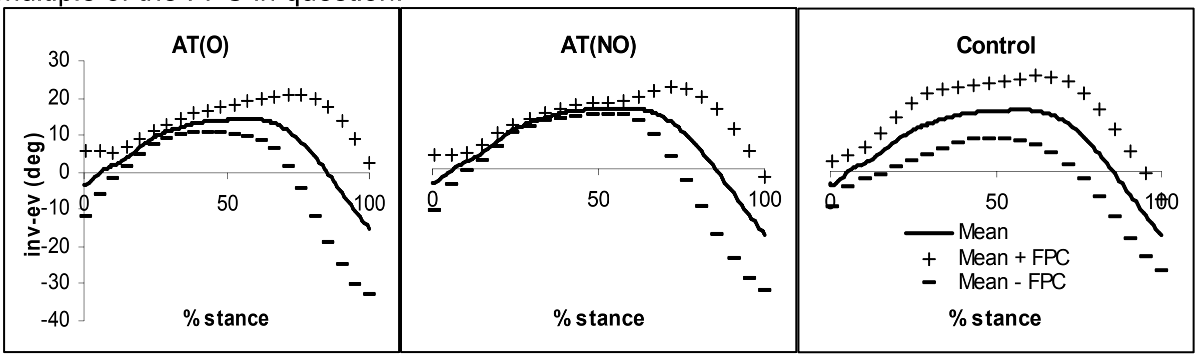
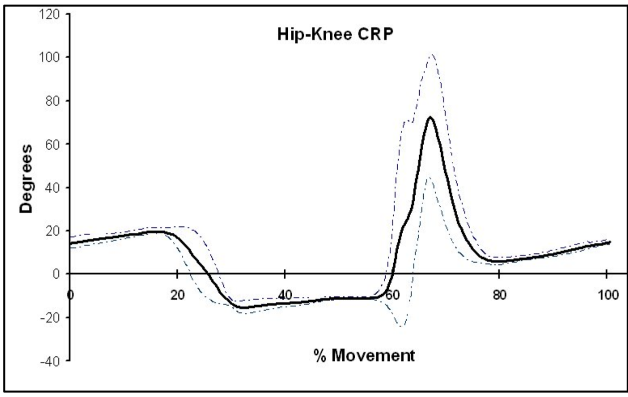


Figure 2 The effects of adding and subtracting FPC1 to the mean eversion angle time series curves for AT(O), AT(NO) and control groups individually.

Figure 2 shows the first FPC extracted for eversion angle for each group, which describes variation around the overall mean curve. Clear differences can be seen with controls showing greater variation in the loading patterns between 10-50% of stance compared to AT subjects. While supporting the findings from traditional approaches that involve more severe forms of data reduction, FDA revealed additional information about the movement patterns and the role of variability during stance. Figure 3 indicates a mean CRP phase with the effects of adding or subtracting a standard deviation from this curve. The CRP of two joint angles can be used to compress these two variables into one, which is representative of both. CRP across the entire movement can be entered into an ANN as a representation of both joint angles. However, caution must be expressed in establishing whether the movements are desired to be in-phase (0^0) or out-of-phase (180^0). Also key elements such as the requirement data normalisation and for sinusoidal data must be adhered to. Figure 4 illustrates a CC of the knee and ankle joints. A normalised CC, representative of the entire data curve, without calculating the correlation coefficient, can be used as input data for the ANN. A key point when examining the CC as future ANN input data is the number of joints to be considered. When more than two are involved the CC acts as a rudimentary measure, and trends, maintained with two joint CC, are within the data set are lost.



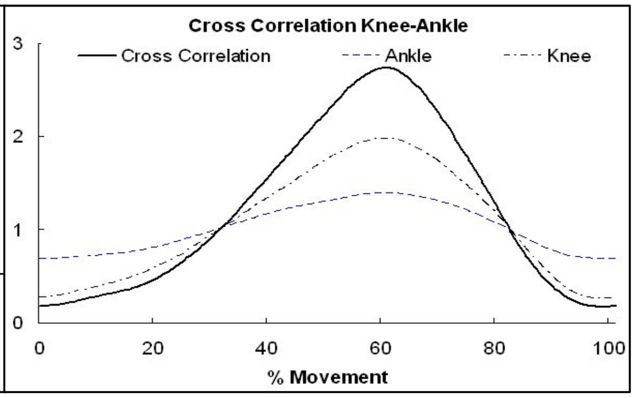


Figure 3 CRP across movement of knee and hip phase angles

Figure 4 Cross Correlation of knee and ankle angles

CONCLUSION:

Data reduction is necessary within the analysis of biomechanical kinematic data. However, the use of a crude measure such as root mean square difference (RMSD), or discrete data point (i.e. mean/max etc.) analysis which reduces an entire series to one point, is not desired and four alternatives are presented. The results of this study showed that CRP, VC, FDA and CC all have possibilities for use in data dimensionality reduction. A key component of each of the selected techniques is the maintenance of the temporal element. Graphical displays illustrate the time series element of each method, and thus show how each of the variables within each data reduction technique may be entered into an ANN. The use of CRP, especially within lower limb kinematics, is preferred as an input as it maintains the trends within two joints by compressing displacement and velocity of both joints into one time series variable. Each technique brings its own benefit and the biomechanist must outline the demands of the study, the desired impact of the data reduction, and the future use of the reduced data before deciding on a particular method.

REFERENCES:

Daffertshofer, A., Lamoth, C. J. C., Meijer, O. G. and Beek, P. J. (2004) PCA in studying coordination and variability: a tutorial. *Clinical Biomechanics. 19 (4)*: 415-428.

Donoghue, O., Harrison, A.J., Coffey, N. and Hayes, K. (2006). Functional data analysis of the kinematics of gait in subjects with a history of Achilles tendon injury. In. *Proceedings of the XXIVth International Society of Biomechanics in Sport Conference*, Salzburg: pp 175-178.

Hamill, J. (2006). Overuse injuries in running: Do complex analyses help our understanding? In. *Proceedings of the XXIVth International Society of Biomechanics in Sport Conference*, Salzburg: pp. 27-32

Hamill, J., Haddad, J. M. and McDermott, W. J. (2000) Issues in quantifying variability from a dynamical systems perspective. *Journal of Applied Biomechanics. 16*: 407-418.

Hamill, J., van Emmerik, R. E. A., Heiderscheit, B. C. and Li, L. (1999) A dynamical systems approach to lower extremity running injuries. *Clinical Biomechanics. 14*: 297-308.

Heiderscheit, B. C., Hamill, J. and van Emmerik, R. E. A. (2002) Variability of stride characteristics and joint coordination among individuals with unilateral patellofemoral pain. *Journal of Applied Biomechanics. 18*: 110-121.

O'Halloran, J. & Anderson, R. (2006). The validation of an Artificial Neural Network to predict power output from rowing kinematics. In. Pro*ceedings of the XXIVth International Society of Biomechanics in Sport Conference*, Salzburg: pp. 95-98

Sparrow, W. A., Donovan, E., van Emmerik, R. and Barry, E. B. (1987) Using relative motion plots to measure changes in intra-limb and inter-limb coordination. *Journal of Motor Behaviour. 19*: 115-129.

Wheat, J. S. and Glazier, P. S. (2006). Measuring coordination and variability in coordination. In. K. Davids, S. Bennett, K. Newell (Eds.) *Movement System Variability* (pp. 167-181). Champaign, IL: Human Kinetics.