APPLICATIONS OF FUNCTIONAL DATA ANALYSIS IN SPORT BIOMECHANICS

Andrew J Harrison

Biomechanics Research Unit, University of Limerick, Limerick, Ireland.

This paper reconsiders the challenge of analysing coordination in human movement with particular emphasis on the application of Functional Data Analysis and Functional Principal Component Analysis. The process of Functional Data Analysis is outlined using examples from biomechanics of sports injuries and coordination in race-walking, jumping and running. The evolution of the spring mass model used to describe jumping and hopping behaviour is examined and it is proposed that this model represents a coordination structure in human movement which may be appropriately analysed using Functional Data Analysis methods.

KEY WORDS: functional data analysis, functional principal components analysis, coordination variability, leg spring stiffness.

INTRODUCTION: In a previous Geoffrey Dyson Lecture, Hamill et al. (2006) considered how well methods of complex analysis had contributed to the understanding of overuse injuries in sport. This work focused on the use of discrete and continuous relative phase techniques to examine the joint and segment coupling coordination structures in running and concluded that those methods could lead to a deeper understanding of the injury process. While traditional methods of analysis in biomechanics often focus on a single joint or segment using discrete measures such as peak displacement, velocity or force, such approaches are inadequate when considering the more complex nature of coordination. Human coordination is a complex process that integrates various functional aspects of muscle action and movement and the multiple degrees of freedom in each of these processes. Coordination may be inherently variable due to the interaction of the multiple degrees of freedom in the components of coordination. To analyse coordination effectively, it is important that variation is considered from a functional perspective rather than the traditionally held view that high levels of variation imply increased noise in the data and low levels of reliability. Therefore, appropriate methods for analysing coordination should ideally consider the modes of variation in the data and be able to distinguish biological variability from measurement error. Various approaches have evolved to analyse the complexity of coordination and most of these methods involve some form of dimension reduction in data. In this paper I consider various examples of the application of Functional Data Analysis (FDA) and Functional Principal Component Analysis (FPCA) in examining time series and coordination data in sports biomechanics.

FUNCTIONAL PRINCIPAL COMPONENT ANALYSIS: Principal Component Analysis (PCA) is a statistical technique, which is ideally suited to dimension reduction and examination of the modes of variation in experimental data. Traditionally PCA has been used to examine and interpret data sets that are discrete in nature rather than continuous time series or families of curves. PCA reduces the dimensionality of an experimental problem by converting a large number of measures into a smaller number of uncorrelated, independent variables, i.e. principal components (PCs) that explain the modes of variation in the experimental data. Currently two distinct approaches have been used to apply PCA to the analysis of biomechanical data sets. These approaches are: PCA of waveforms (Deluzio & Astephen, 2007; Deluzio et al., 2013) or Functional PCA (FPCA) which is generally categorised as part of a larger analysis process, Functional Data Analysis (FDA) which was originally introduced by Ramsay & Dalzell (1991).

In PCA of waveforms, the original curves are re-sampled to ensure equal numbers of records on every waveform and then entered into a large matrix where a Principal Component Score (PC) is derived for each data point on the waveform. While this procedure is relatively easy

to implement using proprietary software applications such as IBM® SPSS® (IBM, New York, USA) or Minitab (Pennsylvania, USA), it has some limitations:

1. Creating data sets of equal length may result in distortion of the time series.
2. Smoothing and calculation of derivatives is carried out separately from PCA procedures resulting in unknown and potentially unwanted sources of variation entering the PCA.
3. Most importantly, in PCA of waveforms, the data points on the curve are assumed to be independent of each other, but in reality we know that any point on a curve is correlated to the data points that precede and follow that point.
4. It may be difficult to relate the waveforms described by each PC to functional patterns of specific subjects in the experimental population.

FDA was devised by Ramsey and Dalzell (1991) in an attempt to rectify some of the limitations of other approaches. The distinctive feature of FDA is that the entire sequence of data points for a measurement are considered as a single entity or function rather than as a series of individual data points (Ryan, et al., 2006). The term functional in FDA and functional principal component analysis (FPCA) refers to the intrinsic nature of measurements frequently obtained in biomechanics experiments. While biomechanical data are often obtained at various regularly spaced time points, these measurements can be assumed to be generated by some relatively smooth underlying function.

Steps in Functional Data Analysis

FDA can incorporate several different procedures, but the following steps are typically used when FDA is applied to the analysis of biomechanical data:

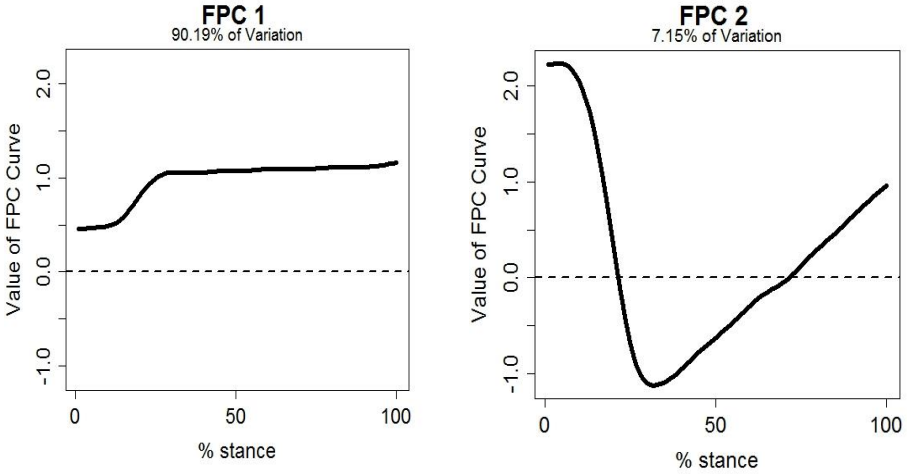
Step 1. Data representation: derivation of smooth functions
Step 2. Registration of data, i.e. time normalisation or landmark registration (optional)
Step 3. Functional Principal Component Analysis
Step 4. Visualising the results of Functional Principal Component Analysis

In step 1 functional data are represented by some form of basis function where the smoothing and derivation of functions are generally linked processes and the decision on the choice of appropriate basis functions is dependent on the nature of the data being analysed. Typically, kinematic and kinetic data can be represented on a functional basis using B-splines. These have been shown to be useful basis functions for smoothing biomechanical data because their structure is designed to provide the smooth function with the capacity to accommodate changing local behaviour (Coffey, et al., 2011). B-splines consist of polynomial pieces joined at certain values of the data series called knots, however other basis functions may be appropriate depending on the nature of the raw data collected.

In Step 2 of FDA, the data may be normalised using land mark registration which ensures the major features of a family of curves such as minima or maxima are aligned. This is an optional step which can be used to reduce the point-wise variation across trials, however, landmark registration is not always suitable to every analysis because landmarks may be missing from certain curves or the timing of landmarks may be ambiguous. Therefore, this step is not always included in FDA and the decision whether to register the data should be made on a case-by-case basis. Ryan and Harrison (2006) demonstrated the effects of landmark registration on kinematic data of the vertical jump in children where the maximum downward displacement of the centre of mass varied between subjects.

Step 3 of FDA generally involves the application of Functional Principal Components Analysis which is an extension of the classical multivariate technique to the functional domain. In this case, eigenfunctions rather than eigenvectors are used to represent the principal components (PC). A major advantage of FPCA is that it produces PCs that are

functions defined in the same domain or state space as the original functional observations and consequently, the functional principal components (FPCs) extracted in the analysis have a definite biomechanical interpretation. The FPC is presented as function within the state space (i.e. it is expressed in the same domain as the original function) and a multiple of this FPC can be added to (or subtracted from) the overall group mean curve to demonstrate the exact effect of the FPC. Figure 1 show two FPCs for the leg abduction-adduction angle (ABD-ADD) during stance phase in running (Donoghue et al., 2008).



A **B**

Figure 1. The first two FPCs for abduction-adduction movement. (Donoghue et al. 2008).

Step 4 of FDA requires the representation of the results in ways that allow insightful interpretation. Ramsay and Silverman (2005) recommend the use of graphs presenting the ensemble mean curve of the original function together with the functions obtained by adding and subtracting a suitable multiple of each FPC function. Figure 2 shows this form of presentation for the first 3 FPCs on the leg abduction-adduction angle data sets of Donoghue et al. (2008). The interpretations of these graphs are relatively simple. Figure 2A shows that high scorers in the first functional principal component (FPC1), illustrated by the plus (+) signs, are characterised by a leg abduction-adduction angle that is higher than the mean function throughout stance. Conversely, low scorers illustrated by the minus (-) signs are characterised by a leg abduction-adduction angle that is lower that the ensemble mean angle throughout stance.

A Mean ± 3×FPC1 **B** Mean ± 3×FPC2 **C** Mean ± 3×FPC3

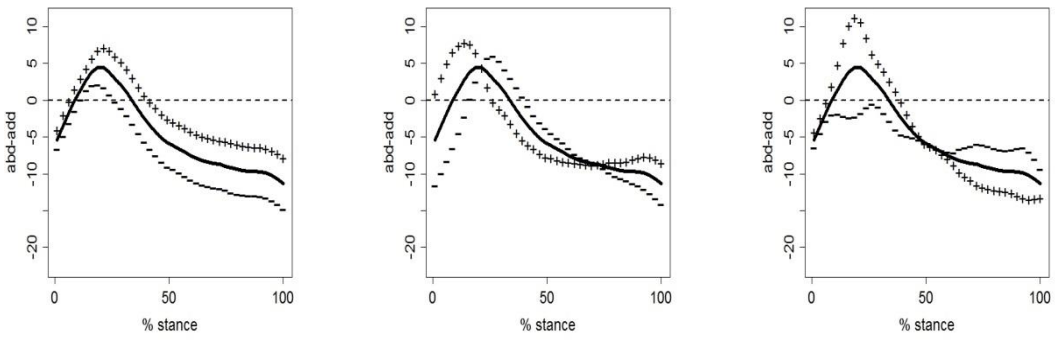


Figure 2. Visualising Functional Principal Components. The graphs show the first 3 principal component of leg abduction-adduction during stance from Donoghue et al (2008). In all cases

the mean function is accompanied by the graphs of high scorers (+) and low scorers (-) on each FPC.

Figure 2B shows the effect of FPC2 relative to the ensemble mean function, with high positive scorers tending to display leg ABD angles that are greater than average at heel-strike, and high negative scorers tending to display smaller leg ABD angles at heel strike. Therefore FPC2 describes the precise leg abduction-adduction movements associated with heel strike. Figure 2C shows FPC3 for the same data set and it is clear that high positive scorers tend to display increased leg abduction-adduction range of motion and negative scorers tend to display decreased leg abduction-adduction range of motion. Therefore FPC3 is representative of leg abduction-adduction range of motion throughout the stance phase.
Donoghue et al. (2008) used FPCA to examine the effects of in-shoe orthoses on the kinematics of the lower limb in subjects with previous Achilles tendon injury compared to uninjured controls. Donoghue et al. (2008) provided evidence using FPCA that in-shoe orthoses appeared to constrain some movement patterns but restored some aspects of variability in other movements.
Coffey et al. (2011) took this analysis further using an extension of FPCA which they called Common FPCA.

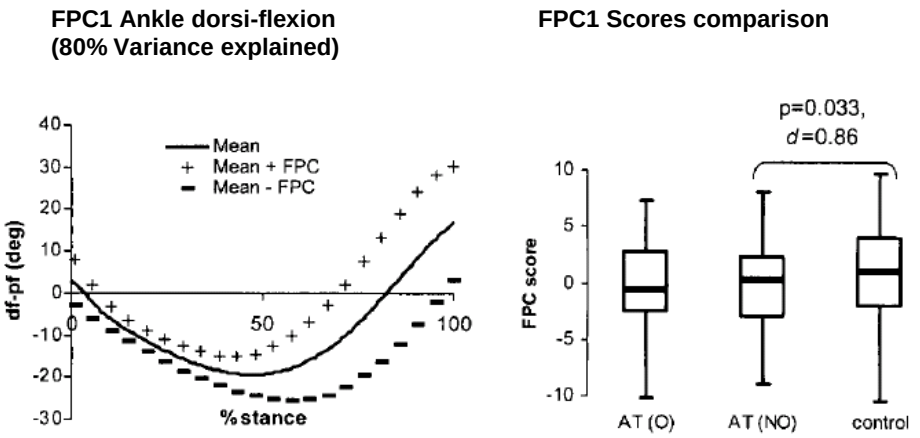


Figure 3 Functional principal component for ankle dorsi-flexion on stance phase in running and analysis of FPC scores for injured subjects wearing orthotics, [AT(O)], without orthotics, [AT(NO)] and uninjured limb [control]. Coffey et al. (2011).

This technique appears to be better suited to analysis of families of curves where repeated measures designs are used. Using Common FPCA on the data of Donoghue et al (2008), Coffey et al. (2011) provided evidence that control subjects had greater levels of variability in lower limb movement patterns than injured subjects.

FPCA also provides and effective means for analysing coordination structures in biomechanical data through the application of bivariate FPCA. This has been used to distinguish differences in kinematic jumping patterns and coordination in groups of children at various stages of development (Harrison, Ryan, & Hayes, 2007; Ryan, et al., 2006). While standard FPCA extracts FPCs from a single variable, in bivariate FPCA each component consists of a vector of FPCs which can be represented on the state space relative to the ensemble mean curve. Figure 4 shows coordination analysis on leg ABD-EV data using bivariate FPCA based on the data set of Donoghue et al. (2008). The analysis shows that the majority of the variation during co-ordination occurs through differences in the leg ABD angle (the vector points almost horizontal indicating little/no change in the EV angle). High positive scorers on this bivariate FPC display increased levels of abduction throughout stance with little relative change in eversion angle.

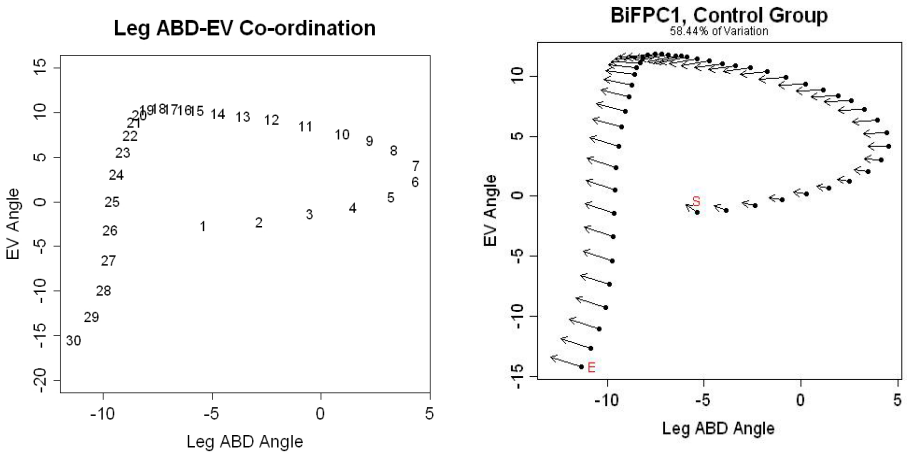


Figure 4. At A, the ensemble average angle-angle plot of the leg ABD vs EV coordination is plotted. B. Shows the bivariate FPC as a vector relative to the ensemble mean function.

An alternative approach by Doná et al. (2009) applied FPCA bilaterally to sagittal knee angle and net moment data in race-walkers of national and international level and found that scatterplots of FPC scores provided evidence of technical differences and asymmetries between the subjects even when traditional analysis (mean ±s curves) was not effective, (see figure 5).

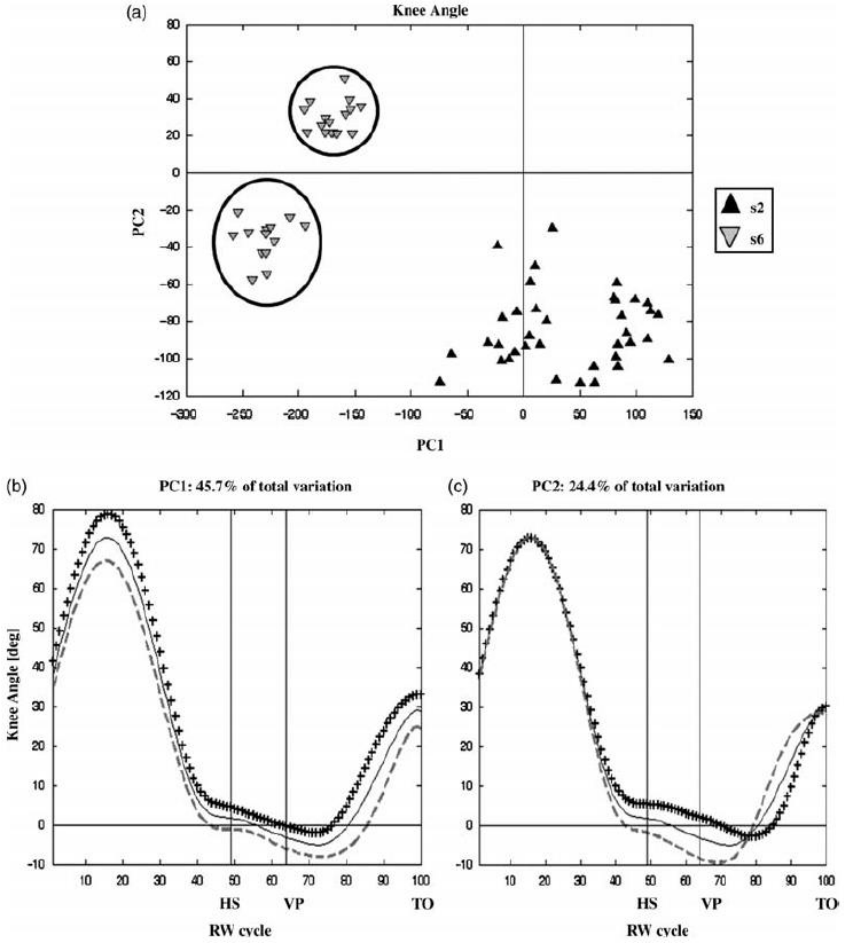


Figure 5. Characterisation of knee joint angle in the sagittal plane. For clarity of representation scatterplots of only two athletes (s2 and s6) are reported: (a) scatterplot of the scores for FPC2

versus FPC1 – the best athlete is represented by black triangles and the lesser performing one by grey triangles. (b, c) the mean knee angle curve is shown with curves created by adding (black plus) and subtracting (grey minus) a multiple of FPC1 (b) and FPC2 (c). Adapted from Doná et al. (2009)

Doná et al. (2009) concluded that FPCA was sensitive enough to detect potentially important technical differences between higher and lower skilled athletes and therefore FPCA might represent a useful and sensitive aid for the analysis of sports movements, if consistently applied to performance monitoring, (see Figure 5). All of the above studies highlight the importance of treating variability in the data as a real, biological phenomenon that has a structure which can be separated from the noise or error information generated by data acquisition. In this respect FPCA appears to be a very useful in examining the modes of biological variability found in biomechanical studies of coordination.

CONCEPTS OF STIFFNESS APPLIED TO RUNNING AND JUMPING: Since the early work of Blickhan (1989), McMahon et al. (1987) and Farley and Gonzalez (1996) running, jumping and hopping in humans has frequently been described using a linear spring-mass model that stores and releases mechanical energy through muscle-tendon structures. Generally this model describes the lower limb as a spring which compresses during landing and returns to full length assisted by muscle contraction before take-off. The overall stiffness of the lower limb consists of a combined stiffness of the joint actions and stiffnesses of the muscle and tendon structures. While much important recent research focuses on the characterisation of stiffness components of isolated muscle and tendon using optic fibre technology (Finni and Komi. 2002) or ultrasound (Couppé, et al., 2003; Lichtwark & Wilson, 2005) the total stiffness of the limb can be estimated using force platform analysis or a combination of force platform and motion analysis. In sprinting, it has been proposed that when an athlete has accelerated to maximum running velocity, (which is attained after 25 to 50 m depending on performance level), then performance may be limited by the ability of the athlete to maintain speed, (Locatelli and Arsac, 1995). Research studies on lower limb stiffness have shown that during fast running, the leg action behaves like a linear spring, (Farley et al., 1991; Farley et al., 1996; McMahon et al., 1987). The stiffness of the leg spring, K_{leg} can be described by the force-displacement relationship of the spring mass model which can be determined by dividing the peak ground reaction force by the change in leg length, ΔL and the vertical stiffness, K_{vert} equals the peak vertical ground reaction force divided by the vertical displacement of the centre of mass, Δy (see Figure 6). In vertical hopping $K_{vert}=K_{leg}$.

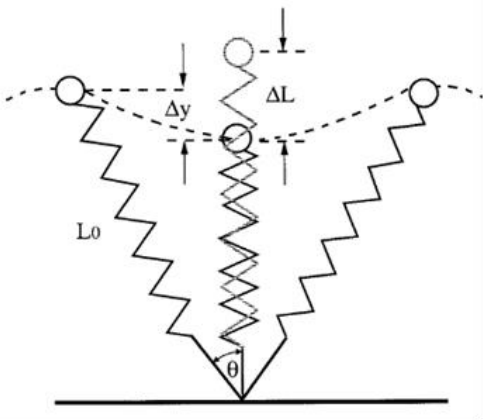


Figure 6: Diagram of the spring-mass model during running L_0, ΔL, Δy, and θ represent the initial length of the leg spring, the maximum compression of the initial length of the leg spring, the maximum vertical displacement of the centre of mass, and the half of the angle swept by

the leg spring during the stance phase, respectively. The spring–mass model drawn with a discontinuous line represents the uncompressed leg spring. The arc drawn with a discontinuous line represents the path of the mass (centre of mass of the subject) during the stance phase. (Adapted from, Chelly and Denis, 2001)

Over the last 30 years, various researchers have demonstrated the importance of leg-spring and vertical stiffness in fast running and hopping activities. Arampatzis et al. (1999) found that leg stiffness increased with increasing running speed and Farley et al. (1991) found that the stiffness of the leg-spring can change as much as twofold to accommodate different hopping frequencies. Chelly and Denis, (2001) showed that leg stiffness values are correlated to maximal running velocity ($r = 0.68$, $P < 0.05$). Similarly, experimental research on well trained athletes has also shown that sprinters have significantly higher leg spring stiffness compared with distance runners (Harrison et al., 2004). It is therefore clear that the capacity to interact with the ground using a stiff spring-like action of the lower limb is critical in encouraging a fast cadence, short ground contact time and fast running speed in the speed maintenance phase of sprinting. While the spring-mass model appears to provide a useful representation of lower limb function in running, hopping and jumping, the ability of this model to discriminate between groups has been somewhat disappointing. For example, Hobara et al. (2011) found no statistically significant effect of stretching on leg stiffness and no significant differences in leg stiffness between males and females at various hopping frequencies (Hobara et al., 2012). Furthermore the reliability of stiffness measures, especially measures of joint stiffness, has been poor. Joseph et al. (2013) demonstrated good reliability for vertical stiffness (ICC=0.87), moderate reliability for leg stiffness (ICC= 0.85) and poor reliability for knee and ankle stiffness (ICC= 0.51 and 0.56 respectively) during a running task. Further inspection of leg and joint stiffness results across various studies consistently shows high levels of variability in stiffness measures with coefficients of variation ranging 10% - 25% for leg stiffness and up to 50% for joint stiffness (Bradshaw, et al, 2013; Hobara et al., 2009). It appears that stiffness reliability may be dependent on the nature of the measure of stiffness used however this is based on an assumption that the reliability of a measure is dependent on its variability. In many studies, stiffness is often defined as a single (discrete) measure rather than a description of the force-displacement relationship during ground contact.
In figure 7B, this measure of stiffness is indicated by the gradient of the line between the origin and the grey circle at the middle of the stance phase. This illustrates the potential oversimplification of stiffness as a discrete point measure derived from a continuously varying signal.

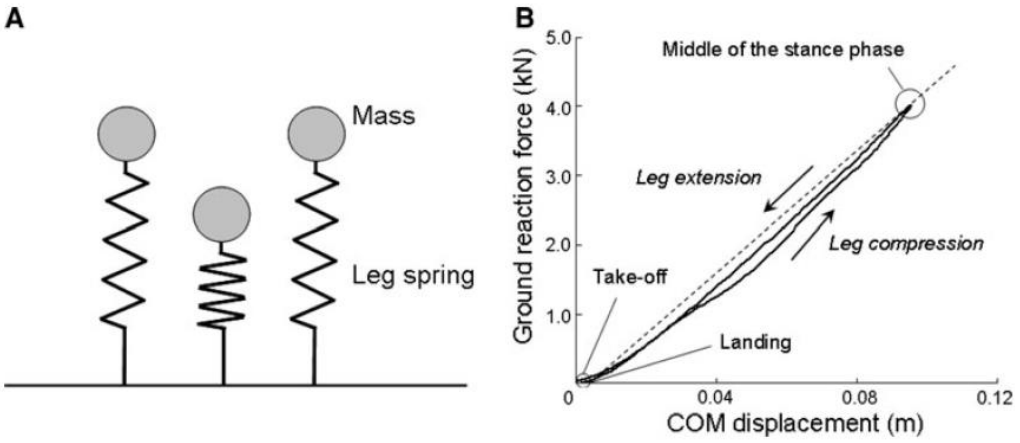


Figure 7: Diagram of the spring-mass model during vertical hopping. At A, the action of the hopping is illustrated at the start, middle and end of the stance phase. In B, a typical vertical

ground reaction force vs centre of mass displacement curve for vertical hopping. Leg stiffness is represented by the slopes (dotted lines) of these curves in the leg compression phase. (Adapted from, Hobara, et al. 2011)

Treating leg and joint stiffness as discrete point measures introduces a similar problem encountered earlier when analysing coordination structures. Furthermore, the high levels of variability observed in measures of leg and joint stiffness may results from the multiple degrees of freedom of the components of stiffness coordination and as such this variability may be an essential biological feature of the stiffness rather than a source of error contaminating the measures. It is recommended that analysis of leg and joint stiffness should be treated as a coordination issue rather than a discrete parameter and as such it is likely that the application of FDA on stiffness coordination may provide improved insight on the role that stiffness control may play in the biomechanics of sport and sports injuries.

CONCLUSION: The question of whether complex analyses, particularly FDA have contributed significantly to the understanding of the biomechanics of sports performance and sports injuries was revisited. Based on the results of various studies it is concluded that FDA and FPCA have provided some useful insights. The potential applications of FDA in sports biomechanics are numerous and varied. FDA appears to provide particular strengths in analysing coordination structures since it presents analysis in the same state space as the coordination structure and this can facilitate clearer insights into specific changes in coordination patterns. Furthermore, the FPCs produced in the analysis relate directly to the functional behaviour of specific subjects in the population being examined and this strengthens the validity of the conclusions drawn from FPCA. While FDA and FPCA appear to provide a promising addition in analysing biomechanical data, some caution is advised to ensure that FDA is used appropriately. As pointed out by Hamill (2006), if a simple analysis method results in the answer to a specific question, then there is no need for a more complex technique. Looking forward, it recommended that sports biomechanics as a discipline should explore and seek to advance the application of FDA, particularly in the analysis of stiffness control in running hopping and jumping and in other coordination related analyses.

REFERENCES:

Arampatzis, A., Bruggemann, G-P. & Metzler, V. (1999) The effect of speed on leg stiffness and joint kinetics in human running *Journal of Biomechanics* 32: 1349-1353

Blickhan (1989) The spring-mass model for running and hopping. *Journal of Biomechanics*, 11: 1217-1227.

Joseph, C.W., Bradshaw, E.J., Justin Kemp, J. and Clark, R.A. (2013) The Interday Reliability of Ankle, Knee, Leg and Vertical Musculoskeletal Stiffness During Hopping and Overground Running *Journal of Applied Biomechanics*, 29: 386-394.

Chelly, S.M. and Denis, C. (2001) Leg power and hopping stiffness: relationship with sprint running performance. *Medicine and Science in Sports and Exercise*, 33: 326-333.

Couppe, C, Kongsgaard, M., Aagaard P., Hansen, P., Bojsen-Moller, J., Kjaer, M. and Magnusson, S,P. (2008) Habitual loading results in tendon hypertrophy and increased stiffness of the human patellar tendon. Journal of Applied Physiology 105: 805–810.

Donà, G., Preatoni, E., Claudio Cobelli, C., Rodano, R. and 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. and Hayes, K. (2008) Functional data analysis of the kinematics of running gait in subjects with chronic Achilles tendon injury. *Medicine and Science in Sports and Exercise,* 40 (7): 1323-1335.

Farley, C.T., Blickhan, R., Saito, J., Taylor, C.R., (1991) Hopping frequency in humans: a test of how springs set stride frequency in bouncing gaits. *Journal of Applied Physiology*. (71): 2127-2132.

Finni, T. and Komi, P.V. (1998) Two methods for estimating tendinous tissue elongation during human movement. *Journal of Applied Biomechanics*, 18: 180-188.

McMahon, T.A. and Cheng, G.C. (1990). The mechanics of running: how does stiffness couple with speed? *Journal of Biomechanics* 23 (Suppl.1), 65-78.

Farley, C.T. and Gonzalez, O. (1996) Leg stiffness and stride frequency in human running. *Journal of Biomechanics*, 29: 1818-186.
Hamill, J., Haddad, J.M. and Van Emmerik, R.E.A. (2006) Overuse injuries in running: Do complex analyses help our understanding? In: Schwameder et al., (Eds) *24th International Symposium on Biomechanics in Sport*: 175-178.
Harrison A.J., Keane S.P. and Coglan J. (2004) Force-velocity relationship and stretch –shortening cycle function in sprint and endurance athletes. *Journal of Strength and Conditioning Research*, 18(3): 473 – 479.
Harrison, A.J., Ryan, W. & Hayes, K. (2007) Functional Data Analysis of joint coordination in the development of vertical jump performance, *Sports Biomechanics*, 6 (2): 196-211.
Hobara, H., Kato, E.,Yoshiyuki, K. (2012) Sex differences in relationship between passive ankle stiffness and leg stiffness during hopping, Journal of Biomechanics, 45: 2750-2754.
Hobara. H., Inoue, K., Kato, E. & Kanosue, K. (2011) Acute effects of static stretching on leg-spring behaviour during hopping. *European Journal Applied Physiology*, 111:2115-2121.
Hobara et al., (2009) Knee stiffness is a majordeterminant of leg stiffness during maximal hopping. *Journal of Biomechanics*, 42: 1768–177
Coffey, N., Harrison, A.J., Donoghue, O.A. and Hayes, K. (2011) Common Functional Principal Component Analysis: A new approach to analysing human movement data. *Human Movement Science*, 30(11): 1144-1166.
Deluzio, K.J., and Astephen, J L. (2007). Biomechanical features of gait waveform data associated with knee osteoarthritis: An application of principal component analysis. *Gait & Posture,* 25 (1), 86–93.
Deluzio, K.J. (2013) Interpreting principal components in biomechanics: Representative extremes and single component reconstruction. *Journal of Electromyography & Kinesiology*, 23 (6),1304-1310.
Lichtwark, G.A. and Wilson A.M. (2005) In vivo mechanical properties of the human Achilles tendon during one-legged hopping. *Journal of Experimental Biology* 208: 4715-4725
Locatelli E. and Arsac, L. (1995) The mechanics and energetics of the 100 m sprint. *New Studies in Athletics,* 10(1): 81-87.
Ramsay, J.O. and Dalzell, C.J. (1991). Some tools for functional data analysis. *Journal of the Royal Statistical Society Series B*, 53: 539-572.
Ramsay and Silverman (2005)
Ryan W., Harrison A.J. and Hayes K. (2006) Functional Data Analysis in Biomechanics: A case study of Knee Joint Vertical Jump Kinematics, *Sports Biomechanics*, 5 (1): 121 -138.