UNIVERSITY OF
BATH

University of Bath

1 **Title**

2 Application of functional principal component analysis in race walking: an emerging

3 methodology

4

5 **Keywords**

6 Functional principal component analysis, kinematics, kinetics, movement variability,

7 sports skills.

8

9 **Author list**

10 Giulia Donà[1], Ezio Preatoni[2,3], Claudio Cobelli[1], Renato Rodano[2], Andrew J.

11 Harrison[4]

12 *[1] Dipartimento di Ingegneria dell'Informazione, Università di Padova, Padova, Italy*

13 *[2] Dipartimento di Bioingegneria, Politecnico di Milano, Milano, Italy*

14 *[3] Dipartimento di Industrial Design, Arti, Comunicazione e Moda (INDACO),*

15 *Politecnico di Milano, Milano, Italy*

16 *[4] Department of Physical Education and Sports Science, University of Limerick,*

17 *Ireland*

18

19 **Contact information of the contact author**

20 Giulia Donà, PhD

21 Dipartimento di Ingegneria dell'Informazione

22 Università di Padova

23 Via Gradenigo 6b

24 35131 Padova, Italy

1   phone: +39 049 8277805

2   fax: +39 049 8277826

3

Abstract

This study considered the problem of identifying and evaluating the factors of individual performance during race walking. In particular, the study explored the use of functional principal component analysis (f-PCA), a multivariate data analysis, for assessing and classifying the kinematics and kinetics of the knee joint in competitive race walkers. Seven race walkers of international and national level participated to the study. An optoelectronic system and a force platform were used to capture three-dimensional kinematics and kinetics of lower limbs during the race walking cycle. f-PCA was applied bilaterally to the sagittal knee angle and net moment data, because knee joint motion is fundamental to race walking technique. Scatterplots of principal component scores provided evidence athletes' technical differences and asymmetries even when traditional analysis (mean ± SD curves) was not effective. Principal components provided indications for race walkers' classification and identified potentially important technical differences between higher and lower skilled athletes. Therefore, f-PCA might represent a future aid for the fine analysis of sports movements, if consistently applied to performance monitoring.

[171 words]

1 Introduction

2

3 A major task in coaching is to devise and implement training programs that facilitate

4 progressive improvement of the athlete's performance. Hence, two of the major

5 challenges in sports biomechanics research are: the identification of the personal

6 performance features for any athlete; and the determination of the most proficient

7 strategy, among the many available, to improve individual technique and maximise

8 performance. Biomechanical studies often attempt to capture an athlete's

9 characteristic kinematics and kinetics, evaluate the correctness and proficiency of

10 their movement and prevent possible injuries.

11

12 When trying to capture the biomechanics of individual technique, research should not

13 merely focus on the best performance of an athlete, but it should attempt to analyse

14 the individual's "typical" mode of the performance. Namely, it should capture the core

15 biomechanical strategy that governs the movement, regardless of the variations that

16 emerge in repeating the same action. Every time that a subject tries to carry out the

17 same movement, a certain amount of variation may be registered between the

18 successive trials: even elite athletes can not precisely reproduce identical movement

19 patterns, after many years of training. Variability in movement patterns plays a

20 fundamental role in sports skills and its influence on the analysis of biomechanical

21 data should be taken into account (Bartlett et al., 2007). The choice of a most

22 representative performance may be arbitrary and results derived from analysis of

23 such performances may be misleading. Hence, a full analysis of an individual's motor

24 behaviour should involve the evaluation of an appropriate number of repetitions

1 (Rodano and Squadrone, 2002; James et al., 2004; James et al., 2007; Preatoni,

2 2007) together with a thorough evaluation of movement variability.

3

4 Biovariability is the consequence of the extreme complexity of the locomotor system

5 and of the redundancy of its degrees of freedom. The neuro-muscular and skeletal

6 system is always subjected to perturbations, that may originate from both internal

7 processes and external influences. In clinical biomechanics, increased variability can

8 be associated with decreased stability in performing a movement (Dingwell et al.,

9 2001; Hausdorff, 2007), while in sports it remains unclear how movement and

10 coordination variability may affect athletes' performance. Some authors indicate that

11 athletes are generally capable of high levels of reproducibility in their performance of

12 skills (Salo and Grimshaw, 1998; Knudson and Blackwell, 2005). Hatze (1995)

13 suggested that the variations occurring in repeated actions could be interpreted as

14 particular realizations of a regular stochastic process with superimposed chaotic

15 excursions. Other studies proposed that variability could indicate a form of instability

16 or loss of ability of the athlete (Bartlett et al., 2007), or a compensatory mechanism

17 for finding an optimal solution to altered situations (Winter, 1984; Hamill et al., 2006).

18

19 Motion analysis of repeated movements generates measurements that do not consist

20 of a single curve, but of a family of curves, each one slightly different from the other.

21 Results of kinematic and kinetic analyses commonly consist of a large number of

22 highly correlated, time-varying variables. There is the need to find a structure in the

23 data, discovering the most characteristic features and predicting whether a pattern is

24 representative for the athlete's skill description or not. Hence, two main needs have

25 to be satisfied: data reduction and data interpretation. The first demand seeks to

eliminate collinearity and to simplify data. The second demand seeks to obtain a meaningful summary of the data, without neglecting the information that movement variability conveys.

Standard data analysis techniques (i.e. determination of mean, standard deviation, etc.) in time series data, summarise individual biomechanics in single patterns (the average behaviour) and show deviations as possible errors (standard deviation bands). This procedure, though useful, reduces the data severely, may discard much important information (Donoghue et al., 2008), and, in particular, does not account for the information that may be inherent in all the variability apparent in the data. Information can therefore, be lost when trials from several subjects are averaged, and the average curve does not closely resemble any of the individual curves (Sadeghi et al., 2003).

In contrast, multivariate statistical analysis has proved to be a powerful tool to eliminate collinearity and to facilitate analysis, presenting only the essential structures hidden in the data (Chau et al., 2005), but again the extent of data loss is a matter of concern. Among multivariate statistical techniques, linear transformations are computationally easier to perform and within linear transformations, the use of functional techniques may provide additional insight into skill differences in kinetics and kinematics patterns (Chau, 2001).

Functional principal component analysis (f-PCA) has been shown to be very effective for the study of human motion in modelling the lactate (Newell et al., 2006) and motion curves (Ormoneit et al., 2005), in identifying hidden combinations and

relationships between biomechanical variables (Daffertshofer et al, 2005), in understanding the motor development process (Ryan et al., 2006), and in analyzing joint coordination data in motor development (Harrison et al., 2007). Principal Component Analysis (PCA) is a multivariate statistical technique, which aims to reduce the dimensionality of high-dimensional data sets. It accomplishes this by computing a new, much smaller set of uncorrelated variables (Principal Components - PCs), which best represent the original data-set. Each new variable is a linear combination of the original ones. The first principal component (PC1) is the linear combination of the original variables which accounts for the maximum amount of variance. The second principal component (PC2) is orthogonal to the first one and accounts for the maximum amount of the remaining variance in the data. All the principal components are orthogonal to each other, so there is no redundant information. All remaining principal components are defined similarly, so that the lowest order components normally account for very little variance and can usually be ignored.

Functional principal component analysis is an extension of the traditional PCA, where the principal components are represented by functions rather than vectors (Ramsay and Silverman, 2002; Ryan et al., 2006; Harrison et al., 2007). The basic philosophy of functional data analysis is the belief that the best unit of information is the entire observed function (a curve within a family of curves) rather than a string of numbers. It is assumed that data are supposed to have an underlying functional relationship governing them. f-PCA demonstrates the way in which a set of functional data varies from its mean, and, in terms of these modes of variability, quantifies the discrepancy from the mean of each individual functional datum. f-PCA allows extracting loadings

and scores. The loadings are the correlation coefficients between the variables and the components. Scores are the contributions of the principal components to each individual variable.

Daffertshofer et al., (2004) suggested that multivariate analysis methods, and PCA in particular, examining the entire waveform data, might accurately identify embedded patterns of complex movements. Recently, Donoghue et al., (2008) used functional data analysis in a clinical study on subjects with a history of Achilles tendon injury. The same technique was used by Ryan et al., (2006) and Harrison et al., (2007) in the study of kinematic vertical jump data, to differentiate some developmental stages of this motor task. These studies demonstrated the potential of Functional Data Analysis (FDA) in providing an insight into movement control and coordination in running and jumping activities. It is clear that much more effort should be spent in the investigation of the use of functional principal component analysis for reducing and interpreting sports motion data, while accounting for their original variability.

Given the ubiquity and health relevance of variability in gait measurements, it is critical that we summarise and compare gait data in a way that reflects the true nature of their variability. There is the need to correctly quantify and assess movement variability, and to relate it with its possible determinants: skills, age, neuro-musculo-skeletal alterations or pathologies, evident or latent injuries. Despite the apparent simplicity of these tasks, if not conducted prudently, the derived results may be misleading, as we will exemplify. In fact, there are to date, many unanswered questions relating to the analysis of quantitative gait data, such as the elusive problem of systematically comparing two families of curves (Chau et al., 2005). For

all these reasons, multivariate statistical approaches as well as the impact of intra-subject variability in the assessment of kinematic and kinetic data need to be further explored.

The aim of the present study was to show how functional principal component analysis could be a valuable tool for studying the kinematics and kinetics of race walking. In particular, the potential of FDA in characterizing race walkers abilities for different performance levels was explored. The attention was focused on the knee motion, which is a fundamental and critical aspect of race walking technique. Hence, the knee flexion-extension angle and joint moment were analysed. Race walking was chosen as the movement of investigation, because it is a motor task that presents peculiar biomechanical and coordinative demands.

Race walking has specific constraints imposed by two rules: first, the race walker must maintain continuous contact with the ground during the progression of steps; second, the knee of the walker's supporting leg must remain straight, without bending, from the moment of the first contact with the ground until the leg is in the vertical position. A failure to observe these rules during the competition implies sanctions or disqualification. These constraints make race walking highly technical and rather stereotyped. The choice of this very repeatable movement provide a good basis for gauging variability among gait curves, faithful to the underlying data distribution and minimally influenced by extraneous observations.

1 **Methods**

2

3 *Participants*

4 Four male (weight 64 ± 2.4 kg, height 1.81 ± 0.09 m) and three female (weight 50.7 ±

5 6.8 kg, height 1.67 ± 0.05 m) race walkers of national and international class were

6 recruited for this study. All the athletes regularly completed at least six training

7 sessions a week. Race walkers were assigned alpha-numeric codes in decreasing

8 order according to their skill levels and named "s1", "s2", etc. The ordering was

9 carried out by considering their competitive results and coach evaluation of their

10 technical ability. Athletes' season best over the most common distances (5, 10 and

11 20 km events) of race walking competitions were considered (Table1). Race walkers

12 were divided in three groups: elite athletes (i.e. international medal winners: s1, s2

13 and s3), very good athletes (i.e. international level competitors: s4 and s5), good but

14 the lesser performing ones (i.e. national rank athletes: s5 and s6).

15 The typical race walking velocities of the athletes could range from 3.34 to 4.17 m/s

16 during competition and from 2.75 to 5 m/s during training. Athletes with any

17 remarkable lower limb injury or dysfunction at the time of the experiments were

18 excluded.

19

20 The experimental procedures complied with university guidelines and were approved

21 by the local institutional review board. Every athlete was informed of the aims of the

22 research, testing procedures, personal data storage procedures and informed that

23 they could withdraw from the experiment at any time. All participants provided

24 written informed consent before participation.

25

1   *Subject preparation*

2   The SAFLo marker-set (Frigo et al., 1998) was used. This consists of a total-body

3   marker-set, with 19 retroreflective hemispherical markers (15 mm diameter) fixed on

4   specific anatomic landmarks, namely: lower prominence of the sacrum, posterior

5   superior iliac spines, lateral femoral condyles, lateral malleoli, fifth metatarsal heads,

6   seventh cervical vertebra and point of maximum kyphosis, acromion processes,

7   lateral epicondyles and styloid processes of the humerus, parieto-occipital areas of

8   the head. Particular care was taken to ensure that sweating or rapid movement did

9   not affect marker position or stability on the skin.

10

11   *Test procedure*

12   Kinematic data were acquired using a photogrammetric motion analysis system

13   (ELITE 2002, BTS, Milan, Italy) which consisted of 8 cameras operating at 100 Hz.

14   The system was calibrated before each experimental session according to

15   manufacturer guidelines and a maximum mean error of 1.0 mm, of a 600 mm rigid

16   bar was tolerated. The cameras were positioned to detect the position of markers

17   placed on both sides of subjects and covered a calibrated acquisition volume of

18   approximately 64 m^3 (8×2×4 metres). The set up also allowed athletes to perform

19   their movement without interference. A force plate (AMTI OR6-7-1000, Watertown,

20   USA) synchronised to the motion analysis system and with a sampling rate of 500 Hz

21   was used to capture the ground reaction force data.

22

23   The subjects performed a standard 20 minute warm up routine and some walking

24   trials before measurements were obtained. A 15 m long walkway allowed them to

25   perform technically correct race walking at a constant speed. The dimensions of the

1 laboratory were big enough to let participants circle continuously and reach an

2 adequate, approximately constant speed over the force platform, where data were

3 acquired. Trials were rejected if the subjects did not place their foot fully on the force

4 platform or there was any obvious evidence of targeting the force platform. Up to 20

5 suitable race walking trials (both for left and right side) were acquired for each

6 athlete. Trials were performed at a self-selected pace and under coach supervision,

7 to ensure the quality and correctness of race walkers performance.

8

9 *Data processing*

10 Specially designed algorithms were used to estimate (Pedotti and Frigo, 1992) and

11 filter (D'Amico and Ferrigno, 1990) the three dimensional coordinates of internal joint

12 centres and their derivatives. The position, velocity and acceleration of the centre of

13 mass were determined from the three dimensional coordinate and anthropometric

14 data. Three dimensional coordinates of internal joint centres, joint angles, velocities

15 and accelerations were calculated and hip, knee and ankle joint moments were

16 estimated by using the Newton-Euler free body dynamic equilibrium equations. Each

17 body segment mass, inertial moments, and gravity centre positions were estimated

18 through Zatsiorsky and Seluyanov regression equations (Zatsiorsky and Seluyanov,

19 1983).

20

21 Fifteen time varying measures were considered: antero-posterior, medio-lateral and

22 vertical ground reaction forces; hip, knee and ankle joint angles in the sagittal plane

23 (bilaterally); hip, knee and ankle joint moments in the sagittal plane (bilaterally); hip,

24 knee and ankle joint powers in the sagittal plane (bilaterally); pelvic tilt, pelvic

25 obliquity and pelvic rotation angles.

1

2   The analysis of the data was carried out using Matlab 7.0.4 (© Elsevier Ltd. The

3   Mathworks Inc, MA, USA). Extension moments and powers of lower limbs were

4   defined as positive. Ground reaction forces were normalised by body weight;

5   moments and powers were normalised by body weight and height (Hof, 1996). Knee

6   kinematic and kinetic variables were reported in this study because they were

7   judged, at this stage, as being among the most important parameters relating to race

8   walking description. The knee flexion-extension is a fundamental of race walking

9   technique. Its control is addressed by one of the two defining rules and,

10 consequently, it is very important for the athlete's performance characterization.

11

12 For every trial the "race walking" cycle was analysed, as the interval from toe-off to

13 the following toe-off of the same foot. The first toe-off was estimated through

14 customised algorithms based on kinematics: this corresponded to the peak of vertical

15 acceleration of the marker applied to the fifth metatarsal of the supporting leg (Hreljac

16 and Marshall, 2000). The second toe-off came from ground reaction force measure

17 and it corresponded to the instant in which the vertical ground reaction force reached

18 the base line.

19 No time normalization process was performed in order to avoid any alteration of

20 temporal patterns. Although the durations of race walking cycle for each trial were

21 different, the observations beyond the end of the movement simply correspond to the

22 final stationary value of the waveform (Epifanio et al., 2008).

23

24 *Functional principal component analysis*

1 The functional data in this experiment is observed at discrete time points t and it is

2 assumed the data contain a certain amount of noise. Therefore, we can think of the

3 observed values as arising from some true, smooth (noiseless) function $x(t)$ plus

4 some additional noise/measurement error. This functional datum (i.e. curve) y_j can be

5 represented by the equation:

6

7
$$y_j = x(t_j) + \varepsilon_j$$
[Eq. 1]

8

9 where $x(t_j)$ is the signal and ε_j is the noise.

10 Following the procedures of Ramsay and Silverman (2002) and Ryan et al. (2006),

11 B-splines were used to smooth the kinetic and kinematic variables. We assume a

12 known set of functions $[\phi_1(t)....\phi_K(t)]$ and we can estimate any unknown function

13 using a linear combination of a sufficiently large number (K) of these functions. We

14 can express our smooth function $x(t)$ via the equation:

15

16
$$x(t) = \sum_{k=1}^{K} c_k \phi_k(t)$$
[Eq. 2]

17

18 where c_k are suitably chosen coefficients and the summation is taken over $k= 1...K$.

19 The estimated basis functions were further smoothed by adding a roughness penalty

20 to the fitting procedure. The extra roughness penalty term, controlled by a smoothing

21 parameter "λ", ensured that the smoothness of each fitted curve was correctly

22 controlled: therefore data fitting was determined not only by its goodness of fit but

23 also by the level of control of its roughness. This was achieved by minimizing the

24 penalized residual sum of squares (PENSSE) term:

1

$$PENSSE = \sum_{l=1}^{N} [y_l - x(t_l)]^2 = \lambda \int x^{''}(t).dt \qquad \text{[Eq. 3]}$$

2

$$\Rightarrow PENSSE = \sum_{l=1}^{N} [y_l - \sum_{k=1}^{K} c_k \phi_k t_l]^2 + \lambda \int x^{''}(t).dt \qquad \text{[Eq. 4]}$$

3

4

5   where N is the number of curves and x''(t) is the second derivative.

6   Therefore λ is a smoothing parameter and

7

$$\lambda \int x^{''}(t).dt \qquad \text{[Eq. 5]}$$

8

9

10   penalizes the curvature of the estimated function. Increasing λ implies a greater

11   emphasis on smoothness and less on fitting (i.e. interpolating) the data. Decreasing λ

12   implies a greater emphasis on fitting the data and less on smoothness and if λ = 0 a

13   least squares fit is used. Cross-validation was used to determine a starting point for

14   possible values of λ before a final subjective choice was made. The smoothing

15   coefficients used to fit the curves for the knee angle and knee joint moment were

16   0.0092 and 0.0116 respectively. This value of λ was very small which means that the

17   solution was very close to interpolation.

18

19   Hence, each subject was represented by a set of time series smoothed through B-

20   splines and arranged in a three-dimensional matrix. Rows corresponded to repeated

21   trials, columns to frames of the race walking cycle, while the third dimension stored

22   the analysed variables. The matrices of all the race walkers were vertically

23   concatenated and f-PCA procedures were applied to data. The choice of how many

1   features to be retained occurred in two stages; firstly, features that cumulatively

2   explained at least 95% of the original data variation were held for further analysis.

3   Secondly, a "scree test" (Cattell, 1966) was applied to these features. A scatterplot

4   analysis was carried out, to inspect the ability of f-PC scores to differentiate athletes.

5   It consisted of a graphical representation of the scores of the first principal

6   components. Points corresponding to trials of the same subject were drawn with the

7   same colour. A biomechanical interpretation was derived by the evaluation of the

8   deviation from the mean caused by these PCs. It consisted of the plot of the mean

9   curve of the variable, with curves created by adding and subtracting a multiple of the

10   principal component (Ramsay and Silverman, 2002). This representation highlighted

11   the race walking characteristics that cause most of the variation in the data.

12

1 **Results**

2

3 The most widely used traditional method to estimate data variability is the standard

4 deviation curve (SD). The SD about the mean ensemble curve of a joint angle

5 represents the possible time-specific variations in the movement of a joint. The

6 standard deviation curve is obtained by adding and subtracting one standard

7 deviation to the mean trend at each point on the mean ensemble curve. The larger

8 the distance between the standard deviation curves of the mean ensemble, the

9 greater the variability in the movement pattern. Standard deviation curves were

10 measured for all the kinetic and kinematic variables. One of the best athletes (s2-

11 female) and one of the less performing ones (s6-male) are considered as examples

12 since they provided the most interesting results. Raw data and SD curves for the

13 knee joint angle and joint moment are shown in Figure 1, where only the right side is

14 presented for sake of clarity.

15

16 The knee angles and knee joint moments of the two race walkers are represented

17 with different colours: black and grey. The race walking action is divided through

18 vertical lines into some principal phases: the front leg support phase, the rear leg

19 support phase and the swing phase. The first phase consists of the swing phase,

20 when the foot has no contact with the ground, swings and then prepares to approach

21 back the ground. The second phase begins at heel strike and ends when the

22 supporting leg passes beyond the vertical projection of the centre of mass. The last

23 part of the race walking cycle begins when the stance leg passes the vertical upright

24 position and ends with toe-off. The standard deviation bands around the sample for

1   the two subjects overlap for most of the race walking cycle, thus revealing

2   qualitatively similar curves at the knee joint.

3

4   For the knee angle, the first four functional components (f-PCs) accounted for most

5   of the variance in the data (95.2%), while for the knee joint moment the first six

6   components accounted for 95.0% of data variance (Figure 2). The first four principal

7   components for the knee angle in the sagittal plane are taken into analysis; Figure 3a

8   presents the scatterplot of the scores on the first and second functional principal

9   components.

10

11   Most of the race walkers scored positively on f-PC1, but s6 had strong negative

12   scores. Moreover, two separate clusters of points (enclosed within circles) could be

13   clearly recognized for s6*,* one with positive and one with negative scores on f-PC2*.*

14   These clusters corresponded to trials evaluated for the right and left lower limbs.

15   Figures 3b-3c show the overall mean curve of all subjects for the knee angle, along

16   with two other curves created by adding (indicated by a plus sign) and subtracting

17   (indicated by a minus sign) a multiple respectively of the first and second functional

18   component to and from the mean curve.

19

20   Inspection of these data shows that f-PC1 described athletes with different mean

21   values for the flexion-extension of the knee throughout the race walking cycle (Figure

22   3b). Athletes scoring positively on f-PC1 flexed their knee more during race walking

23   than the average, while race walkers scoring negatively extended their knee more.

24   The second functional component (f-PC2) was related to a technique variation in late

25   swing and early stance (Figure 3c). Athletes scoring positively on the second

1   functional component maintained an extended knee (knee joint angle near to zero

2   value) during the first part of the stance phase, while athletes with negative scores,

3   tended to hyperextend the knee (negative joint angle) respect to the average curve

4   during the transition through the vertical position. These behaviours were reversed in

5   the last swing phase: negative scores corresponded to a less extended knee while

6   positive ones were related to a more extended leg. One of the best race walkers, s2,

7   had strong negative scores on the second functional component.

8

9   The scatterplot of scores for f-PC3 and f-PC4 of s2 and s6 was reported in Figure 4a.

10   Athletes with negative scores on f-PC3 corresponded to a larger range of knee

11   flexion-extension especially during stance phase compared to participants who

12   scored positively on this functional principal component (Figure 4b). These race

13   walkers tended to hyperextend the knee when passing through the vertical position.

14   Positive scores on f-PC4 were related to a larger knee extension in the approach to

15   the floor and a delay in the knee flexion at the end of the stance phase and during

16   the swing phase (Figure 4c).

17

18   The analysis was supplemented by the investigation of the knee moment in the

19   sagittal plane. Results for the first two most important components are reported in

20   Figure 5: s6 scored positively on f-PC1, while s2 negatively (Figure 5a). Positive

21   scores on f-PC1 were related to a higher knee extension moment respect to the

22   mean curve during the stance phase (Figure 5b). Moreover, athletes scoring

23   negatively anticipated the knee flexion moment, while subjects scoring positively

24   postponed it. Best race walkers had more positive score values for f-PC2 respect to

1 the less performing ones. This was related to a higher knee flexion moment respect

2 to the mean trend just before the heel strike (Figure5c).

3

4 Figure 6a showed the influence of f-PC3 and f-PC4 on knee joint kinetics. One of the

5 less performing athletes, s6, had negative scores on f-PC3, thus revealing (Figure

6 6b) a tendency to postpone the initial knee extension moment and to maintain the

7 delay during the stance phase. On the contrary, the best race walkers, s1 and s2,

8 with positive scores on f-PC3, had an in time and reduced knee extension moment.

9 Hence, scores on f-PC3 were different, while the mean knee moments looked almost

10 identical (Figure 1). Race walkers scoring positively on f-PC4 revealed a larger range

11 of knee flexion-extension moment (Figure 6c).

12

13 A more subtle difference among athletes was pointed out by the analysis of the last

14 two functional components: Figure 7b, in the comparison of "plus" and "minus" lines,

15 revealed f-PC5 to be related to the knee behaving in different ways in the passage

16 through the vertical position. Race walkers scoring positively appeared to have a

17 stronger knee extension moment just after the passage through the vertical position,

18 while fast inverting it into a flexion moment. Athletes scoring negatively showed a

19 smoother trend. The last functional component presented a similar pattern.

20

1   **Discussion and implications**

2

3   In the analysis of race walking technique, particular attention must be paid to the

4   knee joint action in the sagittal plane. The International Federation rules require

5   athletes to keep the supporting leg in a straight position and this makes the knee

6   angle a critical aspect of the motor task. In contrast to normal gait, the knee flexion

7   that normally occurs in the first phase of stance disappears and an extension angle is

8   maintained from about 25% to 75% of contact time.

9   Hence, the knee angle in the sagittal plane and the knee flexion extension moment

10   were analysed both with a traditional (SD curves) and functional principal component

11   analysis. The results demonstrated the potential benefits of functional data analysis

12   in providing greater insight into subtle differences in kinematic and kinetic patterns for

13   athletes of different skill levels. The knee joint angle, SD curves did not reveal

14   evident differences between one of the best athletes (s2) and one of the less

15   performing ones (s6). However, the scatterplots for the first functional principal

16   components clearly separated scores related to these two race walkers: s6 scored

17   negatively while s2 positively on f-PC1. This result was related to a more detailed

18   investigation of pattern differences through f-PCA, with an extension of the knee

19   being either higher or lower with respect to the mean curve throughout the race

20   walking cycle. f-PC2 represented a technical behaviour related to the effort exerted

21   by knee extensor muscles to keep that joint straight in stance as the rules impose.

22   Some race walkers maintained an extended knee during the stance phase, while

23   others tended to hyperextend the knee respect to the average curve. f-PC2 can be

24   considered the consequence of the IAAF rules on the movement pattern.

25

1   A similar pattern was found also in the analysis of the knee flexion-extension

2   moment. A flexion moment occurred in the mean trend at the time corresponding

3   closely to the heel strike, to compensate the external hyperextension moment.

4

5   Studies on race walkers have shown that a small degree of knee flexion sometimes

6   occurs immediately before heel strike, presumably as a protective mechanism

7   against the stress of landing on a hyperextended knee (White and Winter, 1985).

8   Moreover, the knee flexion moment is interpreted by some authors (Chau, 2001;

9   Murray et al., 1983) as the outcome of passive structures (posterior capsule and

10 ligaments) rather than active muscular forces. These results are due to different ways

11 to perform the so called "functional lengthening". This consists in an increased stride

12 length: at the rear limb the ankle plantar flexes; at the forward limb the ankle dorsi-

13 flexes and the knee extends. The process of hyperextension for some athletes might

14 put stress on the posterior structures of the knee joint (Murray et al., 1983).

15 Generally, the lesser performing race walkers appeared to have difficulties in

16 maintaining a correct knee flexion moment just after the passage through the vertical

17 position. Hence, failing to reduce the knee stress at impact, prior to and during heel

18 contact, may be injurious to the ligaments. This may be particularly dangerous in

19 race walking, where heel strikes are more intense than in normal walking and the

20 athlete walks for many kilometres every day. The answer to this problem may be

21 explored by future prospective studies.

22

23 A further result obtained through functional data analysis was the identification of two

24 separate clusters of scores, clearly recognizable in the scatterplots of the lesser

25 performing athletes. These clusters corresponded to trials evaluated for the right and

left lower limbs, thus implying in these race walkers an asymmetric movement of their legs. In literature, the asymmetrical behaviour of the lower limbs during able-bodied ambulation has been found to reflect natural functional differences between the lower extremities (Sadeghi et al., 2000). It could be hypothesized that this asymmetry in race walking is related to the contribution of each limb in carrying out the tasks of propulsion and control. Longitudinal studies appear to be necessary to confirm this hypothesis. Lesser performing race walkers showed knee hyperextension and asymmetry between right and left limb, thus revealing a possible risk of injury.

This study showed how functional principal component analysis can help in a quantitative analysis of high level race walkers. The main potential advantage is that unique factors of athletes can be identified by evaluating movement variability, which, is often considered as a source of error in traditional analysis. f-PCA is very effective for identifying specific locomotion characteristics, by considering the principal component coefficients from each trial of each subject.

Inter- and intra-subject variability may be effectively taken into account and a thorough analysis of differences within and between athletes may be performed. Therefore, PCA provides a useful analytical toolkit for analysing sports biomechanics data, ensuring that important data are not sacrificed in the analysis and that important trends in the kinematic and kinetic patterns are not missed through limitations in the statistical analysis procedures. Moreover f-PCA provided a potentially valuable means of performance analysis and skill characterization. The biomechanical interpretation of the functional components allowed the unique and peculiar characteristics of the highly stereotypical movement patterns in race walking

to be separated among athletes. Athletes' peculiar motor strategies were discovered even for subjects associated by similar skill levels. It is likely that such subtle characteristics might not be found through traditional data analysis techniques. The results obtained by f-PCA revealed an improvement in the sensitivity to differentiate among performance styles. Thus, f-PCA may allow researchers: (i) to identify subtle differences that may be crucial for successful performance at elite levels; (ii) to identify differences that have no effect on performance and separate these from other differences which are performance related. This potential to separate stylistic features from performance characteristics could be an important aid to coaches, providing them with evidence of the movement characteristics which are most important.

f-PCA possesses promising advantages in terms of: (i) solution of the movement variability problem in the data analysis; (ii) reduction of data dimensionality; (iii) functional interpretation of movement variability; (iv) fine characterisation of the individual biomechanics. However, further efforts must be made to: (i) make the method more familiar for practitioners; (ii) find relations between measures and underlying phenomena; (iii) define which factors are truly important for performance and/or injury prevention. Longitudinal experimental designs will be necessary to address these issues.

1 **Conclusion**

2

3 The results of this study showed the potential benefits of functional principal

4 component analysis in detecting unique technique features for a sample of elite race

5 walkers. The basic philosophy of functional data analysis is to consider each function

6 fitted to a set of data as a single observation. It is assumed that data have an

7 underlying functional relationship governing them. Hence, functional data analysis

8 appears to be inherently suitable for analysing biomechanical data. f-PCA fulfilled two

9 main aims: (i) it objectively reduces the large quantity of data that is used to describe

10 gait waveforms, and (ii) it extracts discriminatory principal components that

11 characterise and functionally interpret different race walking patterns.

12

13 Race walking time series data were represented as a set of scores and components

14 providing important information. f-PCA scores and scatterplots gave immediate visual

15 evidence of the main differences amongst athletes. Moreover, functional principal

16 components allowed the interpretation of these differences by identifying the portion

17 of the race walking cycle in which they occurred.

18

19 f-PCA was demonstrated to be a useful tool in detecting the potential deficiencies of

20 some athletes in performing movement. Moreover, it managed to discover that two

21 athletes, classified as belonging to the same level in agreement with their competition

22 results and with trainers information, were performing the race walking movement

23 with significantly different motor strategies. In traditional gait analysis, these

24 differences would probably have been missed in the evaluation of the mean curve

25 trend.

1

2 Functional PCA might be used quantitatively to support individual training procedures

3 and inferred information could be inserted in a graphical interface to present an

4 immediate feedback of the athletes' motor peculiarities. Further studies focusing on

5 within-subject variability and its evolution over time are recommended to investigate

6 the role of variability in injury prevention and performance optimisation.

7

1 **References**

2

3 Bartlett, R.M., Wheat, J., and Robins, M. (2007). Is movement variability important for

4 sports biomechanics?. *Sports Biomechanics*, 6(2), 224–243.

5

6 Cattell, R.B. (1966). The scree test for the number of factors. *Multivariate Behavioral*

7 *Research*, *1*, 245–276.

8

9 Chau, T. (2001). A review of analytical techniques for gait data. Part 1: fuzzy,

10 statistical and fractal methods. *Gait & Posture*, *13*, 49–66.

11

12 Chau, T., Young, S., and Redekop, S. (2005). Managing variability in the summary

13 and comparison of gait data. *Journal of Neuroengineeringand Rehabilitation, 2*(22).

14

15 Daffertshofer, A. Lamoth, C.J.C., Meijer, O.G., and Beek, P.J. (2004). PCA in

16 studying coordination and variability: a tutorial. *Clinical Biomechanics*, *19*(4), 415–

17 428.

18

19 D'Amico, M., and Ferrigno, G. (1990). Technique for the evaluation of derivatives

20 from noisy biomechanical displacement data using a model-based bandwidth-

21 selection procedure. *Medical and Biological Engineering and Computing, 28*(5), 407-

22 415.

23

1 Dingwell, J.B., Cusumano, J.P., Cavanagh, P.R., and Sternad, D. (2001). Local

2 dynamic stability versus kinematic variability of continuous overground and treadmill

3 walking. *Journal of Biomechanical Engineering*, 123(1), 27–32.

4

5 Donoghue, O.A., Harrison, A.J., Coeffey, N., and Hayes, K. (2008). Functional Data

6 Analysis of Running Kinematics in Chronic Achilles Tendon Injury. *Medicine &*

7 *Science in Sports & Exercise, 40*(7), 1323-1335.

8

9 Epifanio, I., Ávila, C., Page, Á., and Atienza C. (2008). Analysis of multiple

10 waveforms by means of functional principal component analysis: normal versus

11 pathological patterns in sit-to-stand movement. *Medical and Biological Engineering*

12 *and Computing*, *46*, 551–561.

13

14 Frigo, C., Rabuffetti, M., Kerrigan, D.C., Deming, L.C., and Pedotti, A. (1998).

15 Functionally oriented and clinically feasible quantitative gait analysis method. *Medical*

16 *and Biological Engineering and Computing*, *36*(2), 179–185.

17

18 Hamill, J., Haddad, J., and Van Emmerik, R.E.A. (2006). Overuse injuries in running:

19 Do complex analyses help our understanding? *Proceedings of the XXIV ISBS*. H.

20 Schwameder, G. Strutzenburger, V. Fastenbauer, S. Lindinger, E. Mueller (Eds). pp.

21 27-32.

22

23 Harrison, A.J., Ryan, W., and Hayes, K. (2007). Functional data analysis of joint

24 coordination in the development of vertical jump performance. *Sports Biomechanics*,

25 *6*(2), 199–214.

Hatze, H. (1995). The extended transentropy function as a useful quantifier of human motion variability. *Medicine and Science in Sports and Exercise*, 27(5), 751-759.

Hausdorff, J.M. (2007). Gait dynamics, fractals and falls: Finding meaning in the stride-to-stride fluctuations of human walking. *Human Movement Science*, *26*, 555-589.

Hof, A.L. (1996). Scaling gait data to body size. *Gait & Posture*, *4*(3), 222–223.

Hreljac, A., and Marshall, R.N. (2000). Algorithms to determine event timing during normal walking using kinematic data. *Journal of Biomechanics*, *33*, 783-786.

James C.R. (2004). Innovative analysis of human movement, Human Kinetics, Champaign (IL).

James, C.R., Herman, J.A., Dufek, J.S., and Bates, B.T. (2007). Number of trials necessary to achieve performance stability of selected ground reaction force variables during landing. *Journal of Sports Science and Medicine*, 6, 126-134.

Knudson, D.V., and Blackwell, J.R. (2005). Variability of impact kinematics and margin for error in the tennis forehand of advanced players. *Sports Engineering*, 8, 75-80.

1   Murray, M.P., Guten, G.N., Mollinger, L.A., and Gardner, G.M. (1983). Kinematic and

2   electromyographic patterns of olympic race walkers. *American Journal of Sports*

3   *Medicine*, *11*(2), 68–74.

4

5   Newell, J., McMillan, k., Grant, S., and McCabe, G. (2006) Using functional data

6   analysis to summarise and interpret lactate curves. *Computers in Biology and*

7   *Medicine*, 36, 262–275.

8

9   Ormoneit, D., Black, M.J., Hastie, T., and Kjellström, H. (2005). Representing cyclic

10   human motion using functional analysis. *Image and Vision Computing*, *23*, 1264–

11   1276.

12

13   Pedotti, A., and Frigo, C. (1992). Quantitative analysis of locomotion for basic

14   research and clinical applications. *Functional Neurology, 7*(S4), 47-56.

15

16   Preatoni E. (2007). Innovative Methods for the Analysis of Sports Movements and for

17   the Longitudinal Monitoring of Individual Motor Skills. PhD thesis, Politecnico di

18   Milano, Italy.

19

20   Ramsay, I.O., and Silverman, B.W. (2002). *Applied functional Data Analysis*. Inc.,

21   New York.

22

23   Rodano, R. and Squadrone, R. (2002) Stability of selected lower limb joint kinetic

24   parameters during vertical jump. *Journal of Applied Biomechanics* 18, 83-89.

25

1   Ryan, W., Harrison, A.J., and Hayes, K. (2006). Functional data analysis in

2   biomechanics: a case study of knee joint vertical jump kinematics. *Sports*

3   *Biomechanics*, *5*, 121-138.

4

5   Sadeghi, H., Allard, P., Prince, F., and Labelle H. (2000). Symmetry and limb

6   dominance in able-bodied gait: a review. *Gait & Posture, 12*, 34–45.

7

8   Sadeghi, H., Mathieu, P.A., Sadeghi, S., and Labelle, H. (2003). Continuous curve

9   registration as an intertrial gait variability reduction technique. IEEE Transactions On

10   Neural Systems And Rehabilitation Engineering, 11(1), 24-30.

11

12   Salo, A., and Grimshaw, P.N. (1998). Examination of Kinematic Variability of Motion

13   Analysis in Sprint Hurdles. *Journal of Applied biomechanics*, 14(2), 211-222.

14

15   White, S.C., and Winter, D. (1985). Mechanical power analysis of the lower limb

16   musculature in race walking. *International Journal of Sport Biomechanics*, 1,15-24.

17

18   Winter, D. A. (1984). Kinematic and kinetic patterns in human gait: Variability and

19   compensating effects. *Human Movement Science*, 3(1-2), 51–76.

20

21   Zatsiorsky, V., and Seluyanov, V. (1983). The mass and inertia characteristics of the

22   main segments of the human body. In H. Matsui and K. Kobayashi (Ed.),

23   *Biomechanics*, *Volume* 4B of *VIII* (pp. 1152–1159). International Series on

24   Biomechanics, Human Kinetics.

25

Table 1

subject	gender	5 km	10 km	20 km
s1	M	0:19:58.00 (4.17)	0:40:56.74 (4.07)	1:25:39.0 (3.89)
s2	F	0:22:55.20 (3.64)	0:46:38.53 (3.57)	-
s3	M	0:21:03.68 (3.96)	0:42:22.59 (3.93)	-
s4	M	0:20:06.61 (4.14)	0:42:59.95 (3.88)	-
s5	F	0:23:25.60 (3.56)	0:48:34.43 (3.43)	1:39:47.0 (3.34)
s6	M	0:21:56.33 (3.80)	0:44:24.97 (3.75)	1:33:06.0 (3.58)
s7	F	0:24:04.61 (3.46)	-	-
mean (speed)		3.82	3.77	3.60
SD (speed)		0.28	0.24	0.28

1 **TABLE CAPTION**

2

3 Table 1: Athletes' season best over the most common distances of race walking

4 competitions. The performances achieved over the 5, 10 and 20 km events are

5 reported. Data are presented in the following format: *h:mm:ss.cc*, where *h* stands for

6 hours, *m* for minutes, *s* for seconds and *c* are decimal places. Dashes mean that the

7 athlete did not compete over that distance. Average progression speed (m/s) is

8 reported between brackets.

9

1 **Figure 1**

2
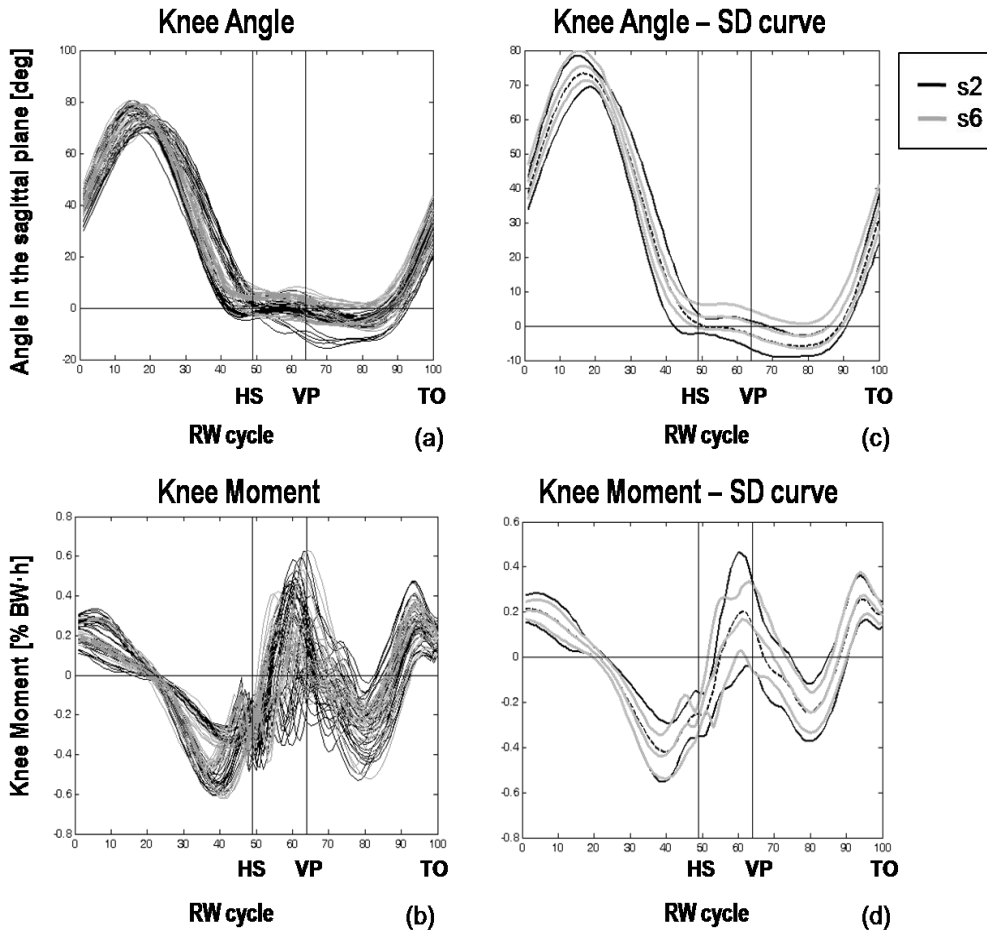
3

4

1   **Figure 2**
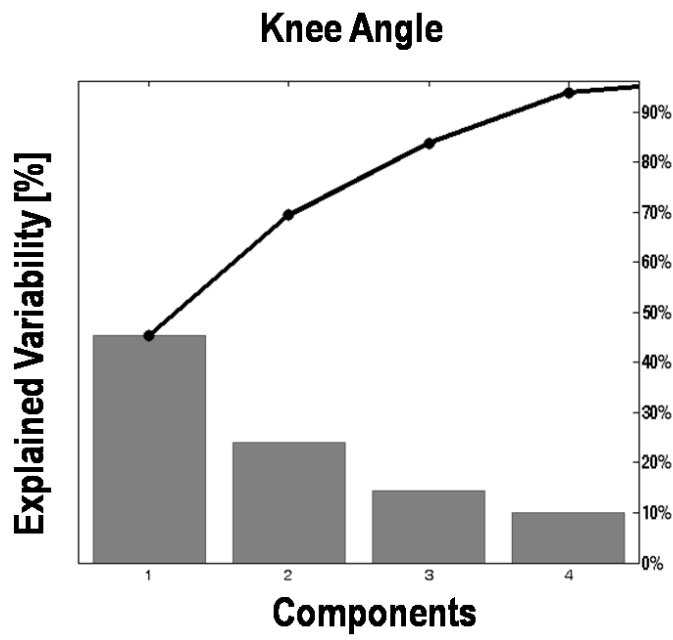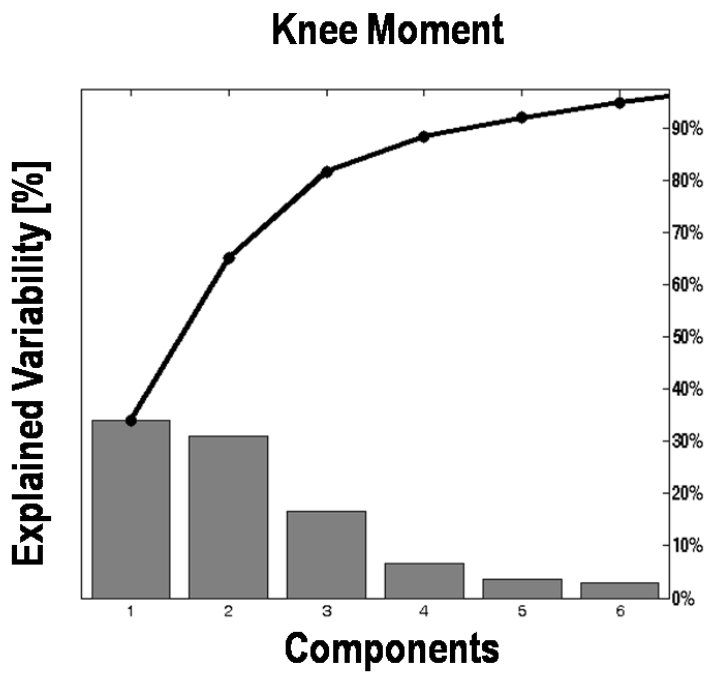
Knee Angle

2

3   (a)

Knee Moment

4

5   (b)

6

1   **Figure 3**

Knee Angle

(a)

(b)   (c)

2

1   **Figure 4**

Knee Angle

(a)

(b)   (c)

2

Figure 5

Knee Moment

(a)

(b)

(c)

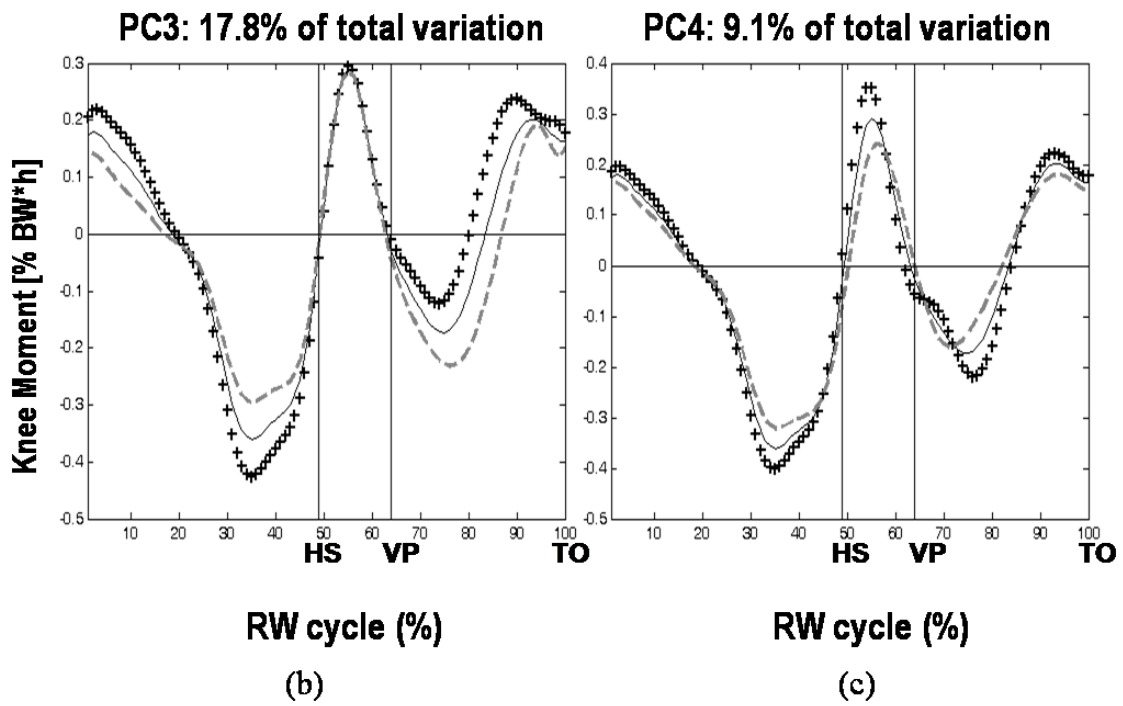# Knee Moment

(a)

(b)   (c)

Figure 7

Knee Moment
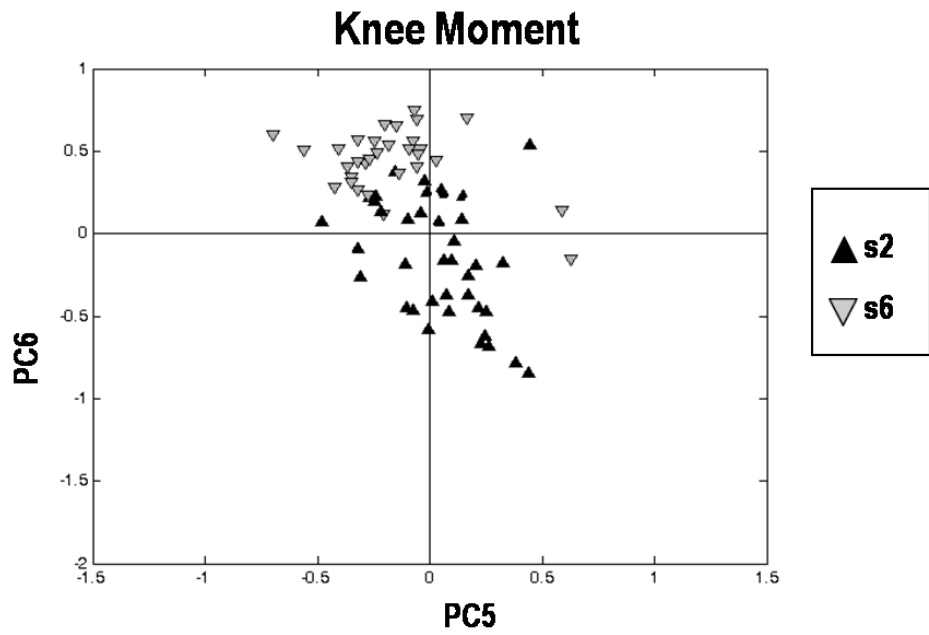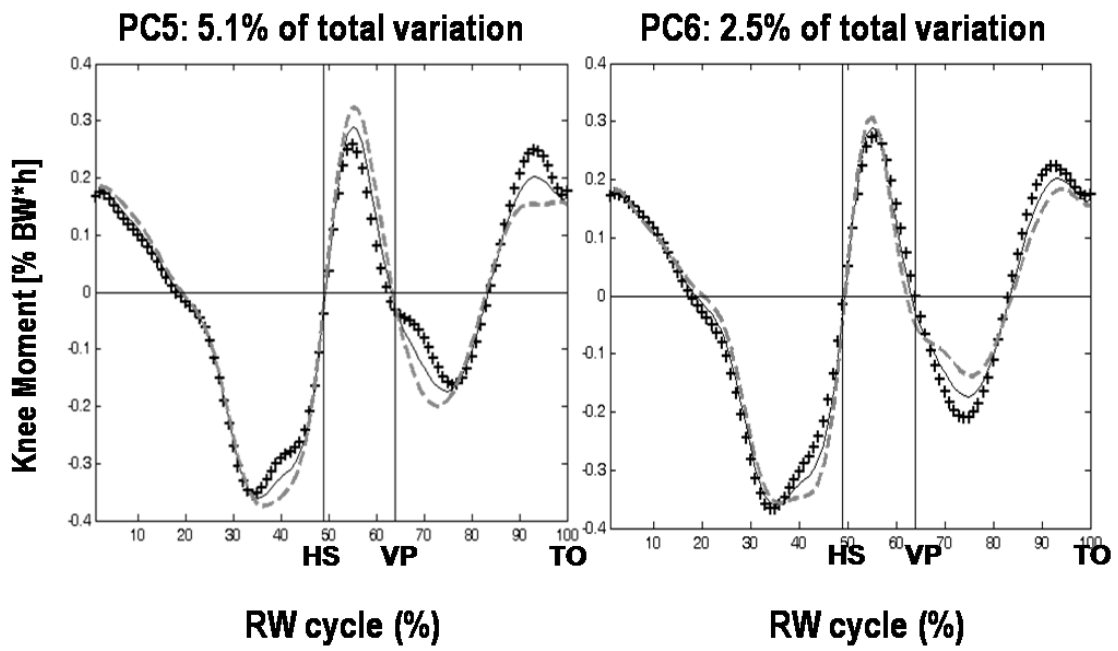

(a)

(b)

(c)
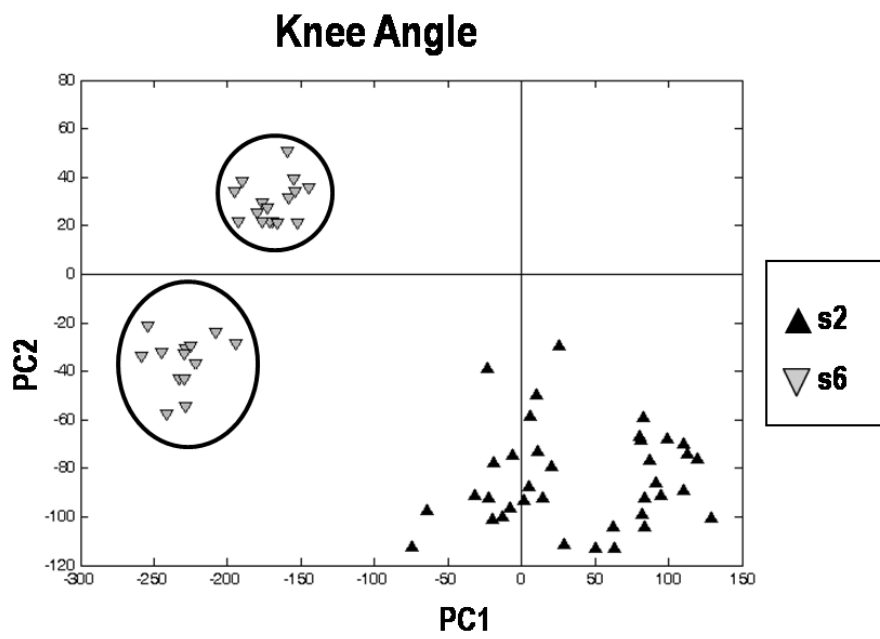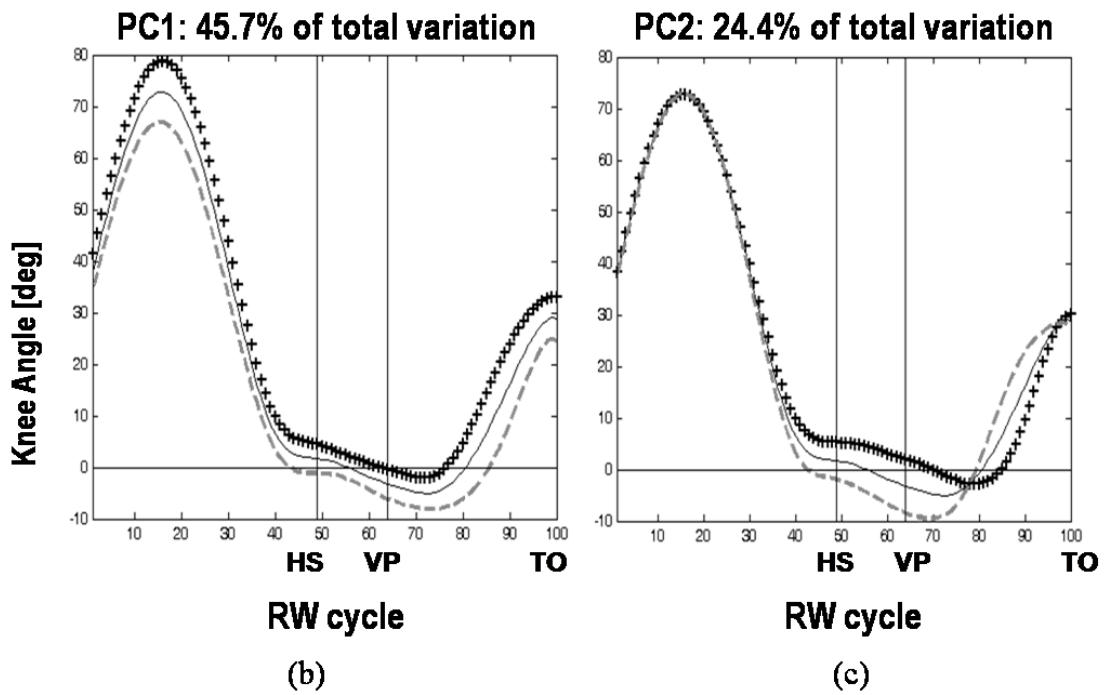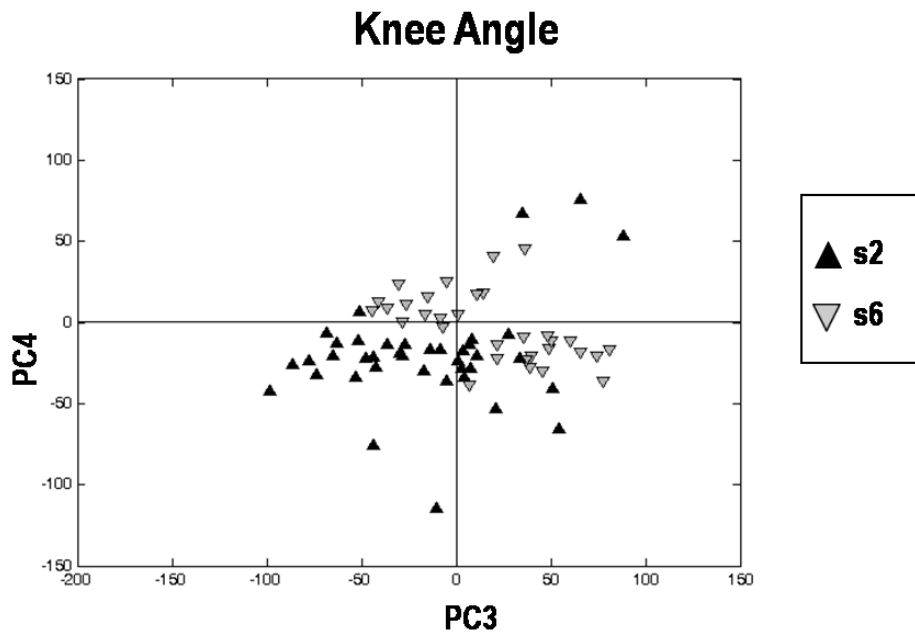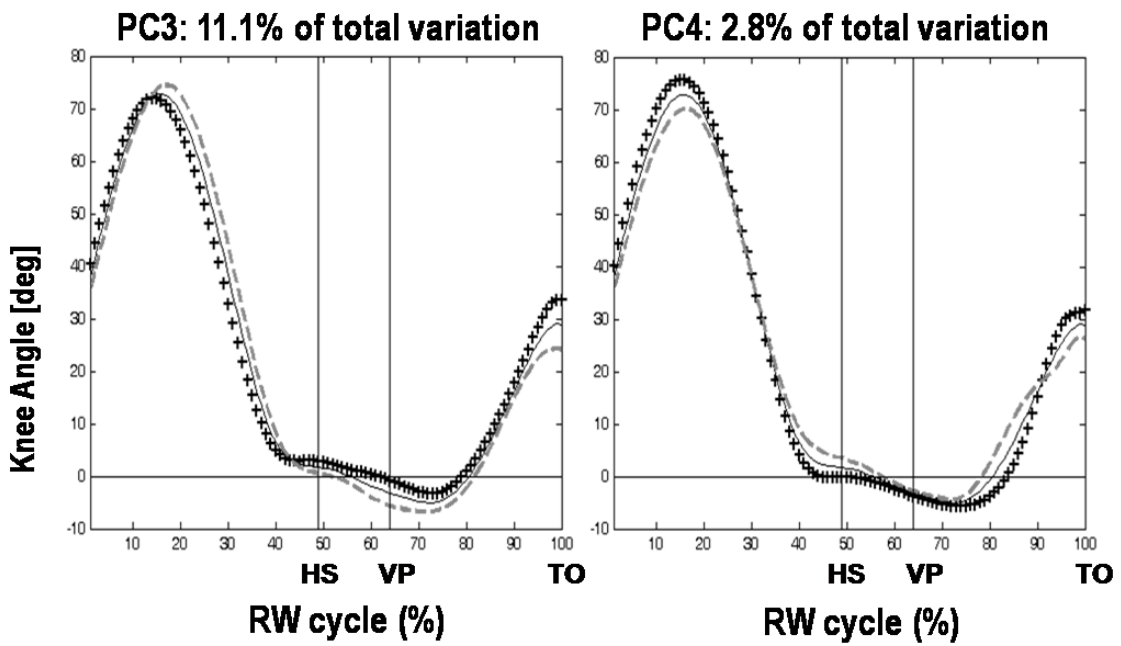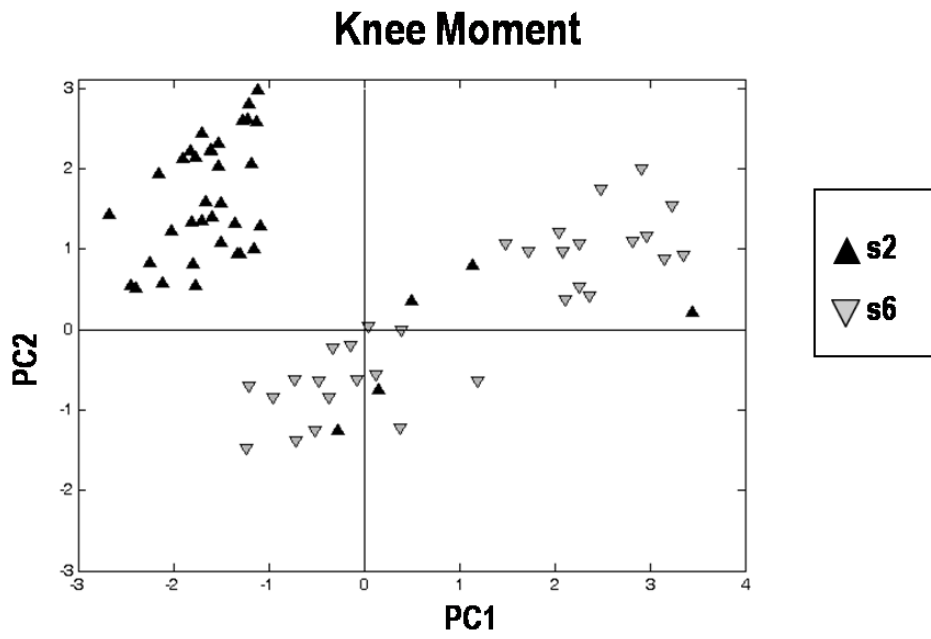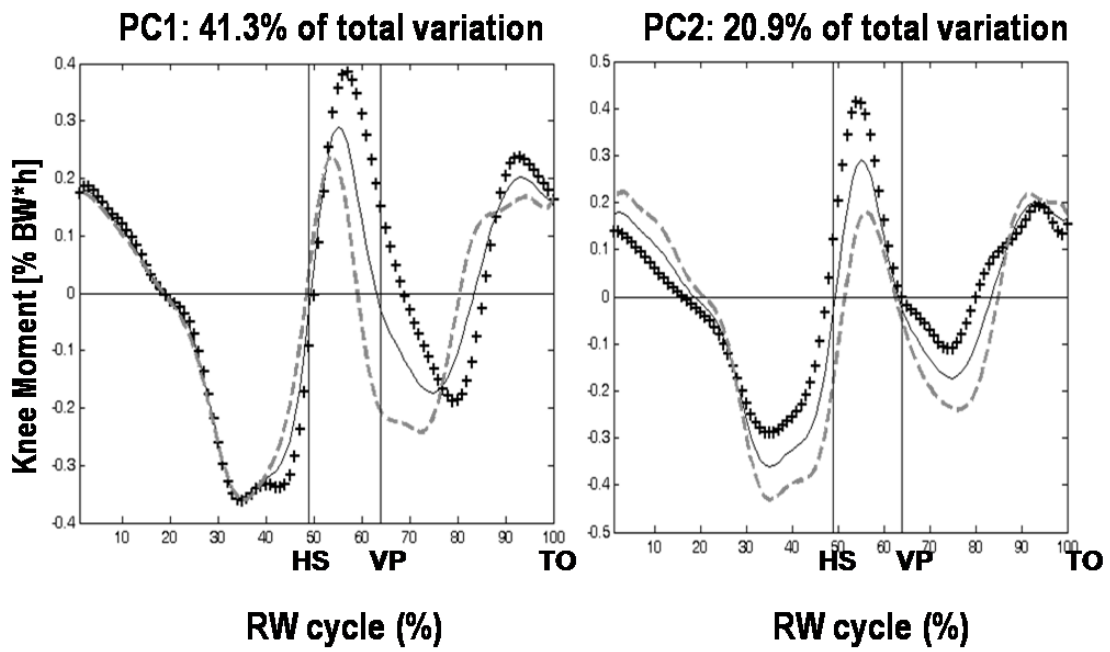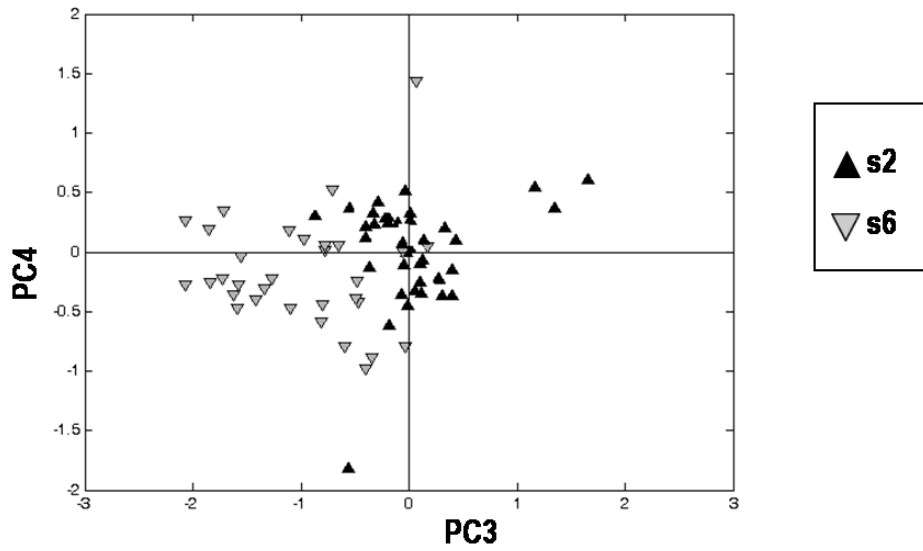
FIGURE CAPTIONS

2

3 **Figure 1**: Knee angle *(a)* and ankle moment *(b)* patterns of the right leg for two out of

4 the seven race walkers. The bunch of individual trials of s2 (black) and s6 (grey) is

5 reported. Knee angle and ankle moment SD curves are represented in *(c)* and *(d)*.

6 Upper and lower bands are obtained by adding and subtracting one standard

7 deviation (SD) to the mean trend (dashed line). The vertical lines divide different race

8 walking phases: "TO" stands for the toe-off; "HS" is the heel-strike; "VP" represents

9 the passage of the stance leg through the vertical upright position (vertical projection

10 of the centre of mass).

11

12 **Figure 2**: Variance explained by the first functional principal components for the knee

13 angle (a) and angular moment (b) in the sagittal plane. Each bar represents the

14 variance explained by the corresponding f-PC; the line above the bars shows the

15 cumulative percentage.

16

17 **Figure 3**: Characterisation of knee joint angle in the sagittal. For clarity of

18 representation scatterplots of only two athletes (s2 and s6) are reported: (a)

19 scatterplot of the scores for f-PC2 versus f-PC1 - the best athlete is represented by

20 "▲", and the lesser performing one by "▼"; (b, c) the mean knee angle curve is

21 shown with curves created by adding (black plus) and subtracting (grey minus) a

22 multiple of f-PC1 (b) and f-PC2 (c).

23

24

1 **Figure 4**: Characterisation of knee joint angle in the sagittal plane for s2 and s6: (a)

2 scatterplot of the scores for f-PC4 versus f-PC3; (b, c) the mean knee angle curve is

3 shown with curves created by adding (black plus) and subtracting (grey minus) a

4 multiple of f-PC3 (b) and f-PC4 (c).

5

6 **Figure 5**: Characterisation of knee joint moment in the sagittal plane for s2 and s6:

7 (a) scatterplot of the scores for f-PC2 versus f-PC1; (b, c) the mean knee moment

8 curve is shown with curves created by adding (black plus) and subtracting (grey

9 minus) a multiple of f-PC1 (b) and f-PC2 (c).

10

11 **Figure 6**: Characterisation of knee moment in the sagittal plane for s2 and s6: (a)

12 scatterplot of the scores for f-PC4 versus f-PC3; (b, c) the mean knee moment curve

13 is shown with curves created by adding (black plus) and subtracting (grey minus) a

14 multiple of f-PC3 (b) and f-PC4 (c).

15

16 **Figure 7**: Characterisation of knee moment in the sagittal plane for s2 and s6: (a)

17 scatterplot of the scores for f-PC6 versus f-PC5; (b, c) the mean knee moment curve

18 is shown with curves created by adding (black plus) and subtracting (grey minus) a

19 multiple of f-PC5 (b) and f-PC6 (c).

20