

Analysis of Multiple Waveforms by means of Functional Principal Component Analysis: Normal vs Pathological Patterns in Sit-to-Stand Movement

Irene Epifanio⁽¹⁾, Carolina Ávila⁽²⁾, Álvaro Page⁽²⁾ and Carlos Atienza⁽²⁾

(1) Departament de Matemàtiques, Universitat Jaume I

(2) Instituto de Biomecánica de Valencia, Universidad Politécnica de Valencia

Corresponding address:

Álvaro Page del Pozo

Instituto de Biomecánica de Valencia

Universidad Politécnica de Valencia

Camino de Vera, s/n. 46022 Valencia (SPAIN)

Tel: + 34 963879160 Fax: +34 963879169 email: afpage@ibv.upv.es

Abstract

This paper presents an application of Functional Principal Component Analysis (FPCA) to describe inter-subject variability of multiple waveforms. This technique was applied to the study of sit-to-stand movement in two groups of people, osteoarthritic patients and healthy subjects. Although STS movement has not been much applied to the study of knee osteoarthritis, it can provide relevant information about the effect of osteoarthritis disease on knee joint function.

Two waveforms, knee flexion angle and flexion moment, were simultaneously analysed. Instead of using the common multivariate approach we used the functional one, which allows working with continuous functions without neither discretization nor time scale normalization.

The results show that time-scale normalization can alter the FPCA solution. Furthermore, FPCA presents a better discriminatory power compared to the classical multivariate approach. Then, this technique can be applied as a functional assessment tool, allowing the identification of relevant variables to discriminate heterogeneous groups, such as healthy and pathological subjects.

Keywords: Functional assessment, Sit-to-stand movement, Functional data, Principal component analysis.

1. Introduction

Principal Component Analysis (PCA) is a multivariate statistical technique which allows describing a set of correlated variables. PCA has been widely used in the field of Biomechanics to describe continuous waveforms (1), (2). Their applications includes fields such diverse as gait analysis (1, 3-7), equilibrium control (8), coordination of thumb joints (9), analysis of lifting techniques (10, 11) or analysis of EMG signals (12, 13), to name a few examples.

According to the objective of the analysis and the source of variability analyzed, we can consider two different approaches to use PCA for describe waveforms.

The first one is focussed on the determination of individual patterns of a movement from a set of time variables that describe it. Thus, each time series is used as a variable whose observations are the recorded values at each time for one or more trials (2, 6, 8, 9), obtaining a set of principal components interpreted as the principal modes of each individual movement performance, or “eigenpostures” (7).

This paper is focused in the second approach, whose objective is the analysis of the inter-subject variability of a set of continuous waveforms. After time-scale normalization, a set of variables can be defined, one for each percentage of the performed movement, obtaining a reduced set of Principal Components (PC's) that quantifies the differences of the analysed time series across subjects (3-5, 10, 11). These PCs can be used to compare groups (patient and controls, for example) (1, 10, 11) or as input variables in a classification system (3).

The applications of PCA to inter-subject variability analysis are usually focused in single time series. An attempt to simultaneously analyze a set of waveforms and discrete variables is developed in (3). This approach generates an enormous

number of discrete variables (almost a thousand), which then needs a high number of PC's to explain a representative percentage of the original variance. On the other hand, in these applications, PCA is used from a multivariate perspective: a finite set of discrete variables are obtained from one or several continuous time series by sampling at arbitrary time intervals. This way, the functional nature of the original variables, is lost.

The specific statistical analysis of continuous functions is dealt with by a new statistical family of techniques called Functional Data Analysis (FDA) (14). The theoretical basis of (FDA) and its differences and advantages with respect to multivariate data analysis (MDA) are described and discussed in the work of Ramsay and Dazell (15). FDA has developed several procedures to extend some classical statistical techniques to the field of functional data. The functional version of PCA provides functional principal components from a set of waveforms without any discretization process. Other features of multivariate PCA, such as the rotation of PCs, have their equivalence in the functional version (16).

An important question in the PCA applications is the way in which the time variable can be handled. In the multivariate approach, the analysis is made from variables defined after a process of linear normalization of the time scale. In the functional one, this kind of normalization is not necessary. The normalization of the time scale is a widely used practice, although it is questionable in the context of PCA, because there is some evidence that the linear normalization can alter the shape of temporal patterns, even increasing the variability of some variables (17). The objective of this work is double. First, we intend to show the effectiveness of functional PCA in defining and interpreting movement patterns from several waveforms. Second, we aim to analyze the effect of the time normalization on the

results of the PCA, and on its capacity to differentiate normal from abnormal patterns.

Therefore, an application of knee flexion angle and flexion/extension moment in sit-to-stand (STS) movement is analyzed for a group of osteoarthritic patients and a control group to show the technique and to illustrate the way to interpret the results.

2. Material and methods

2.1. Experimental setup

Two groups of subjects were considered in this study: the group of osteoarthritic patients, which included 21 volunteers (12 women and 9 men) between 60 and 85 years old with severe knee osteoarthritis and prescribed for a total knee joint replacement; and a control group with 10 healthy volunteers (5 women and 5 men) of similar age, height and weight to the patients' group. Recruiting of osteoarthritic patients was performed by the Orthopaedic Surgery Department of Hospital de la Ribera (Alzira, Spain). All subjects were informed of the purpose and procedures of the experiment and their informed consent was obtained.

Because the variability of STS movement performance among osteoarthritic patients is very high compared to the healthy subjects, it was decided to use a larger sample size for the patients' group.

Most of the osteoarthritic patients had serious difficulties to rise from a standard chair requiring knee flexion higher than 70 degrees. Thus, all subjects were studied while performing sit-to-stand (STS) movement from an armless high chair with the seat tilt forward. This kind of chair was used in order to allow the STS movement performance for every subject.

A total of three trials were performed by each subject. Every individual was asked to cross his or her arms and hold them against the chest during the rising manoeuvre to avoid covering the lower limb markers. Motion was recorded from the time the subjects were told to rise until they reached a relaxed standing position. Motion velocity was not controlled, allowing every subject to stand up at a self-selected speed. The feet positions were not restricted either, so as to facilitate the easiest rising strategy for every subject. The purpose of the study was to evaluate the natural kinematics and kinetics in every subject, considering all possible sources of variability for later analysis and comparison between both groups.

Motion of one lower limb (the most affected one in the case of patients) was recorded with a two-camera video-based motion tracking system KINESCAN/IBV at a frame rate of 50 Hz and obtained through a stereo-photogrammetry technique. Three reflective spherical markers were attached to the lateral side of each lower limb segment using a rigid frame. In addition, ground reaction forces on the studied leg were recorded by a force plate DINASCAN/IBV, synchronized with the motion analysis system (Figure 1). A three-segment model of the lower limb was employed in the analysis, where each segment was assumed to be a rigid body and the ankle and knee joints were considered to be spherical joints. Full kinematics and kinetics of the knee joint were obtained by an in-house developed kinematics software and an inverse dynamic analysis model of the lower limb respectively. Three dimensional joint angles and displacements were calculated by means of Euler angles between the anatomical frames of the different segments of the leg. The anatomical frames were defined from anatomical landmarks located on each segment. The 3D position and orientation of anatomical frames was estimated through matrix

transformation from the local technical frames defined by the markers.

Relationship between technical and anatomical frames was obtained through a previous calibration of anatomical landmarks with regard to technical markers by stereo-photogrammetry. The positions of the joint centers were estimated relative to the position of the anatomical landmarks and the knee joint centre was located at the midpoint of a line between the centers of the lateral and medial femoral condyles.

Resultant external forces and moments at the knee were calculated through the Newton-Euler formulation, based on the joint angles, body segment anthropometric data and the reaction forces obtained from the force plate.

Total duration of STS movement was obtained from recorded kinematics and kinetics waveforms. Initial time instant was set from the ground reaction waveform, at the time when compression force value was above a certain minimum threshold, indicating the beginning of load transfer from the chair to the floor and consequently, on the lower limb. End of the STS movement was obtained from the knee flexion angle waveform, at the time when the subject reached a steady low flexion-extension value after rising from the chair.

Although a complete set of 3D data was obtained, just sagittal plane kinematics and kinetics will be addressed here, and particularly, knee flexion angle and net external knee flexion/extension moment. There are several reasons for this selection: Flexion is the main degree of freedom of the knee and has the largest range of motion of all the kinematical variables studied at this joint. It is also the main motion of the STS movement. Flexion-extension moments are the most relevant moments at the knee and play a significant role in identifying functional differences or changes at this joint when studying the execution of many different daily activities (18-21). Moreover, previous studies (22) have shown that variables

associated to knee flexion angles and moments are the most relevant variables for identifying differences between osteoarthritic patients and healthy subjects in STS movement performance; so it is interesting to analyze the same variables from a functional point of view. Last but not least, artifacts associated to skin and soft tissues can significantly affect the accuracy of motions in coronal and frontal planes (medial-lateral motions and internal-external rotations), whereas motion in the sagittal plane as well as the related kinetic variables are more consistent or less affected by this type of measuring error (23).

In order to allow comparison among different subjects regardless of body size, knee moments were normalized and reported as a percentage of subject's body weight times shank's length.

2.2. Data processing

Data have been processed in three different ways. Firstly, a set of ten discrete measures (each one specified by a single value) was extracted from the waveforms obtained in each trial. These variables are described in Table 1. For each subject, the average measures of the three trials were computed. Therefore, the first approach considers an analysis matrix of size 31x10.

Selection of discrete variables was done according to previous literature related to STS movement analysis (18-22). Most authors choose peak flexion moment, initial and final flexion angles, range of flexion and total movement duration as the main scalar parameters to be considered in the study of knee biomechanics during STS movement. Since knee flexion-extension moment pattern observed in patients was quite variable and rather different from the flexion moment pattern typical of healthy subjects, other descriptive parameters like peak extension

moment as well as mean and final moments were also considered in the present study.

The second and third approaches used the two waveforms introduced in Section 2.1. In the second approach, waveform measures were time normalized: each waveform was defined by 101 values, one for each percent of the movement, i.e. each waveform is considered in terms of its individual STS movement duration, and it is sampled at each 1% from 0% to 100%. Afterwards, the three trials of each subject were averaged. Therefore, the second approach considers two (time normalized) waveforms for each one of the 31 subjects.

In the third approach, no time normalization process was performed in order to avoid a loss or alteration of temporal patterns. The STS movements were observed during the time interval $[0, 3.5]$ sec (3.5 was chosen because at this time all subjects had completed the movement). Although, the durations of STS movement for each trial were different, the observations beyond the movement end simply correspond to the final stationary value of the waveform. The three trials of each subject were also averaged. Therefore, the third approach considers two (raw) waveforms for each one of the 31 subjects.

The second and third approaches deal with waveforms. Although they are recorded discretely (the waveform x_i might consist of n_i pairs $(t_{ij}, x_{ij}), j=1, \dots, n_i$), a continuous curve or function $(x_i(t))$ with $t \in [0, 100]$ and $t \in [0, 3.5]$ for the second and third approach respectively) lies behind these data. In order to convert the waveform observations into a true functional form, we approximate (smooth) each curve by a weighted sum (a linear combination) of 60 B-spline basis functions of order 6 and determine the coefficients of the expansion by fitting data by least squares, as done in (24). Each curve is, therefore, completely determined by the coefficients in this basis, and $x_i(t)$ is computable for any desired argument value t .

Let us see how to apply PCA familiar in MDA in this infinite dimensional domain. A short answer would be that summations change into integrations, but details are given in the following section.

2.2.1. PCA for functional data

In order to see how PCA works in the functional context, let us recall PCA for MDA. In MDA, principal components are obtained by solving the eigenequation

$$V\xi = \rho\xi \quad (1)$$

where V is the sample variance-covariance matrix, $V=(N-1)^{-1} X'X$, where, in turn, X is the centred data matrix, N is the number of individuals observed, and X' indicates the transpose of X . Furthermore, ξ is an eigenvector of V , and ρ is an eigenvalue of V .

In the functional version of PCA, vectors are not considered any more, but PCs are replaced by functions or curves. Let $\{x_1(t), \dots, x_N(t)\}$ be the set of observed functions. The mean function can be defined as the average of the functions point-

wise across replications ($\bar{x}(t) = N^{-1} \sum_{i=1}^N x_i(t)$). Let us assume that we work with

centred data (the mean function has been subtracted), and define the covariance

function $v(s,t)$ analogously by $v(s,t) = (N-1)^{-1} \sum_{i=1}^N x_i(s)x_i(t)$. As explained in

(16), the functional counterpart of equation (1) is the following functional eigenequation

$$\int v(s,t)\xi(t)dt = \rho\xi(s) \quad (2)$$

where ρ is still an eigenvalue, but now $\xi(s)$ is an eigenfunction of the variance-covariance function, rather than an eigenvector. Now, the principal component score corresponding to $\xi(s)$ is computed by using the inner product for functions

$$s_i = \int x_i(s)\xi(s)ds \quad (3)$$

Note that for multivariate data, the index s is not continuous, but a discrete index j

replaces it: $s_i = \sum_j x_{ij}\xi_j$

There are several strategies for solving the eigenanalysis problem in equation (2).

In order to retain the continuity of the original functional data and to reduce the amount of information, we have used the approach proposed in [16]. Instead of using a lot of variables obtained by discretizing the original functions, this type of analysis works with the coefficients of the functions expressed as a linear combination of known basis functions (B-splines in our case). Functional PCA can be carried out easily by using the free library FDA for MATLAB available at <http://www.functionaldata.org>. For a complete review of computational methods for functional PCA, see (16).

Regarding the problem of how many PCs can be computed, let us note that in the functional context, “variables” now correspond to values of t , and there is no limit to these. Therefore, a maximum of $N - 1$ components can be computed. However, if the number of basis functions K (60 B-splines in our case) defining the waveforms is less than N , K would be the maximum. Nevertheless, the first components usually explain most of the variation, and it is not necessary to compute all components.

We can think of principal components as a set of orthogonal basis functions or curves constructed so as to account for as much variation at each stage as possible. Once we have a set of orthogonal components spanning as much variation as we want, we can always rotate these orthogonally to get a new set spanning the same space. The advantage is that rotated components may be easier to interpret. We have used the VARIMAX rotation method (the sum of the

variances of the squared coefficients within each eigenvector is maximized). Again, the procedure can be found in (16) and it is available at the library FDA.

2.2.2. Functional PCA with multiple waveforms

Human movement analysis implies often multiple kinematic or kinetic variables; therefore, several waveforms must be simultaneously described and analyzed. In our case, we work with two functional data: flexion knee angles and flexion-extension knee moment.

Functional PCA can deal with two functional observations per individual, two curves $x(t)$ and $y(t)$. Let $\{(x_1(t), y_1(t)), \dots, (x_N(t), y_N(t))\}$ be the set of pairs of observed functions. Two mean functions $(\bar{x}(t), \bar{y}(t))$ and two covariance functions $(v_{XX}(s, t), v_{YY}(s, t))$ can be computed for each kind of function respectively. Furthermore, we can calculate the cross-covariance function of the centred data

$$\text{by: } v_{XY}(s, t) = (N - 1)^{-1} \sum_{i=1}^N x_i(s) y_i(t).$$

A typical PC is defined by a two-vector $\xi = (\xi_X, \xi_Y)$ of weight functions (two curves). They are solutions of the eigenequation system $V\xi = \rho\xi$, which in this case can be written as

$$\begin{aligned} \int v_{XX}(s, t) \xi_X(t) dt + \int v_{XY}(s, t) \xi_Y(t) dt &= \rho \xi_X(s) \\ \int v_{XY}(s, t) \xi_X(t) dt + \int v_{YY}(s, t) \xi_Y(t) dt &= \rho \xi_Y(s) \end{aligned} \quad (4)$$

Now, the PC score for the i -th bivariate function $(x_i(t), y_i(t))$ is computed by $s_i = \int x_i \xi_X + \int y_i \xi_Y$ because the inner product between bivariate functions is defined by the addition of the inner products of the two components. This amounts to stringing two functions together to form a composite function.

To solve the eigenequation system, each function $x_i(t)$ and $y_i(t)$ is replaced by a vector of values or basis coefficients, and a single synthetic function is built by

joining them together. When PCs have been computed, we separate the parts belonging to each coordinate. Again, this procedure is implemented on the FDA library and is completely explained in (16). Analogously, the varimax rotation method has been used to improve the ease of interpretation.

2.3. Statistical Analysis

In order to analyze the information provided by the three considered approaches, data have been summarized by descriptive measures, and differences between control and patients' groups analyzed, according to each approach. For the first approach, this difference has been quantified by means of the nonparametric test (Mann-Whitney U test (25)), that has been applied to each one of the ten variables. For the second and third approach, we have computed only the first four components, since in both cases they explain more than 95% of the variation. The proportion of variance explained by each eigenfunction is computed as in the multivariate case, by each eigenvalue ρ divided by the sum of all eigenvalues. Furthermore, for each PC, the variation accounted for each original curve $x(t)$ and $y(t)$ is given by $\int \xi_x(s)\xi_x(s)ds$ and $\int \xi_y(s)\xi_y(s)ds$ respectively, because their sum is one by definition; in this way, we can know the weight of each curve, flexion knee angles and flexion-extension knee moment, in the PCs obtained.

The PC scores of each individual on each component have also been described, and, subjected to the Mann-Whitney U test for testing differences between control and patients groups.

Finally, a discriminant analysis has been carried out for each approach and the effectiveness of each approach in classifying both groups was analyzed. The misclassification error rates for each approach were calculated by a cross-

validation procedure, “leave-one-out” method, so each individual was classified from the functions obtained with its own exclusion from the sample.

3. RESULTS

3.1. CLASSICAL SCALAR APPROACH

Statistical analysis of discrete variables reveals that osteoarthritic patients and healthy subjects are significantly different in initial flexion angle, final flexion angle and range of flexion values. Osteoarthritic patients have a lower initial knee flexion angle and higher final flexion, consequently having a smaller range of knee flexion (Table 2).

Regarding the flexion-extension moment, both groups are significantly different in the maximum flexion moment (M_{min}), which is lower for the osteoarthritic group, and the mean moment (M_{mean}), which has negative values for the control group and positive values or near to zero for the patients. There are not significant differences regarding the maximum extension moment (M_{max}) neither the final moment (M_{final}).

Concerning the time variables, there are no statistically significant differences between groups in any of the three variables considered. Total duration of movement is generally longer in osteoarthritic patients; however, the difference with healthy subjects is not significant. Also the time of the M_{min} event seems to happen first in the control group, but this difference is not statistically significant either.

In general, the patients' group has a higher variability in most of the variables considered (initial and final flexion angle, as well as in the maximum, mean and final moment and in the total duration of motion).

3.2. FUNCTIONAL PCA OF NORMALIZED WAVEFORMS

The first four principal components obtained from the analysis of normalized waveforms explain 97.8% of the whole variance associated to the two functional variables (Table 3). Variances of nPC1 and nPC2 are mainly explained by the normalized knee moment (99.3% and 97.9%, respectively), whereas variance of nPC4 is mainly associated with the flexion angle. Finally, variance of nPC3 is associated to both variables, although it is more related to the moment than to the knee angle (79.1% and 21.9%, respectively) (Table 3).

The first PC reveals significant differences in the knee flexion moment between osteoarthritic patients and control group (Table 3). Healthy subjects are characterized by a flexion moment pattern, represented by the negative scores on nPC1 and the negative peak of corresponding waveform of figure 2b. On the other side, osteoarthritic patients show a predominantly extension moment pattern while rising from a chair, related to positive scores on nPC1 and a positive peak on the moment waveform.

Differences in flexion angle waveforms are irrelevant in this PC, since it mainly depends on moments.

The component nPC2 also reveals a difference between the moment patterns of both groups, especially on the second half of the STS movement; although it is not statistically significant because of the great variability among patients (Table 3). Positive scores, mainly related to osteoarthritic patients, correspond to moment values higher than the mean in the second half of the movement and a positive final moment, which means a final extension moment; whereas negative scores, mainly linked to control group, denote moments lower than the mean and a final flexion moment pattern (figure 2d).

The principal component nPC3 reveals significant differences between both groups not only regarding the moment pattern, but also the knee flexion range.

Patients' group has positive mean values in this PC, in contrast to the negative mean value of the control group; this difference being statistically significant (Table 3).

This PC is related to both, angle (21% of variance) and knee moment (79% of variance). Osteoarthritic patients (positive scores waveforms) are characterized by a lower initial flexion angle and consequently by a smaller range of knee flexion than the control group (figure 2e). With regard to moments, osteoarthritic and healthy groups show a different moment pattern in the first part of the movement as well as divergence in the magnitude and location of the moment peaks (figure 2f). Osteoarthritic patients begin the movement with an extension moment, characterized by a first positive peak moment, not shown by control subjects. Besides, patients have a much lower flexion moment than controls, represented by the lower negative peak moment in their waveform. Similar to nPC1 results, control group moves with a flexion moment pattern throughout the whole movement.

To finish, the principal component nPC4 is mainly related to flexion angle. Positive scores are representative of both higher initial and final flexion angle and subsequently, a flexion angle higher than the mean all over the curve (figure 2g). Nevertheless, differences between healthy subjects and patients corresponding to this last component are not statistically significant.

3.3. FUNCTIONAL PCA OF RAW WAVEFORMS

As in the previous approach, the first four principal components explain most of the total variance (95.8%) of the whole set of non-normalized waveforms (Table 4).

The component rPC1 explains 43% of the total variance and is mostly associated to differences in the flexion/extension moment. Those recordings with positive

scores in rPC1 have a positive or extension final moment, while negative scores are related to a negative or flexion final moment. Moreover, this PC reveals differences in the shape of the moment waveform; positive scores are characterized by a clearly marked flexion moment peak that later decreases and finally reaches a final extension moment, whereas those moment waveforms with negative scores do not have such an evident peak, but an increase in the flexion moment that then remains practically invariable until the end of movement (figure 3b).

Differences between the control and patients' scores in rPC1 are not statistically significant; hence this principal component must be associated to individual differences that are not produced by osteoarthritis. However, there is a noticeable difference between the standard deviations of both groups, with the osteoarthritic patients having a much higher variance than the healthy subjects.

The component rPC2 explains 31.9% of total variance and reveals different moment patterns between osteoarthritic patients and healthy subjects; this difference being statistically significant (Table 4). Osteoarthritic patients show an extension moment pattern, characterized by a positive and smaller moment peak corresponding to positive score waveforms (figure 3d). Control group is distinguished by a flexion moment pattern, with a higher negative peak on their moment waveforms (negative scores in rPC2).

The angle component of rPC2 also shows different patterns of motion.

Osteoarthritic group has a lower initial flexion angle and a slower motion performance (figure 3c).

The principal component rPC3 explains 10.5% of the total variance and is representative of the variability in the angle waveforms (Table 4). The meaning of this PC is related to the final flexion value and therefore to the differences in the

range of flexion. Osteoarthritic patients have positive scores, which means a certain knee flexion at the end of the movement (figure 3e). In contrast, control subjects achieve a complete knee extension at the standing posture. As in the case of rPC2, the difference between both groups related to rPC3 is statistically significant.

The component rPC4 is also mainly related to flexion angle. However, it is not related to pathology differences, since mean values of patients and controls' scores do not differ significantly. Positive scores are representative of a higher initial flexion angle and a time gap with regard to those curves having negative scores (figure 3g). This delay is not due to a lower motion velocity but to the larger range of flexion.

3.4. DISCRIMINANT ANALYSIS

To conclude, the discriminatory power between control and patients' groups of the variables used in every one of the studied approaches was compared. Ten variables were included in the traditional approach based on discrete variables and the scores in four principal components in each of the functional PCA methods. A discriminant analysis was performed in each case. Misclassification error rates for each approach are gathered in Table 5. The best results in classification are those of the functional PCA on the raw waveforms without time normalization (6.5% of errors). The discriminatory power is a bit worse when applying the PCA to time-normalized waveforms (9.7% of errors). The traditional approach using 10 discrete variables to represent the whole waveform gives the worst result (16.1% of errors), in spite of using more than twice the number of variables than the functional approaches (10 compared to 4).

4. DISCUSSION

This study shows the possibilities of functional PCA as a useful technique for analysing multiple waveforms associated to human movements, as well as for identifying continuous temporal patterns from this kind of data. We have applied this technique to analyze the STS movement. STS movement has been previously used to characterize the differences between healthy and pathologic people, but the published papers on this subject do not use a functional approach (18, 20-22, 26, 27).

PCA has been widely applied to identify individual patterns of motion by extracting principal modes of variation of a set of time variables. In this approach, each time series is used as a variable whose observations are the recorded values at each time for one or more trials (2, 6, 8, 9). The obtained principal modes do not quantify the variability across subjects, although aggregated patterns of movement or classification processes can be performed from the individual principal modes, implicitly assuming that the structure of such eigenpostures is the same for all subjects (7).

Our work is focused in a different approach whose objective is to describe inter-subject variability of a set of time variables. This approach has been used in previous papers in a multivariate way: a finite set of discrete variables are obtained from one or several continuous time series by sampling at arbitrary values of time. This way, the functional nature of the original variables, is not considered. For this application, PCA is usually applied to single waveforms after a linear normalization of the time scale. When more than one variable is considered (joint angles, velocities, kinetic variables) a separate PCA is performed for each variable and one set of principal components (PCs) for each functional variable is obtained (1, 4, 10, 11). This approach increases the number of PCs;

moreover, these PCs can be correlated because they have been obtained from the same movement by means of separated analyses of related variables. In this study we have worked with the raw data, without applying any time-scale normalization, and we simultaneously analyze two continuous waveforms corresponding to kinematic (flexion angle) and kinetic (knee moment) variables. This approach is similar to that developed by Astephen and Deluzio (3), although the use of PCA in that paper is not functional but multivariate, since the original waveforms are decomposed into a set of variables after a linear normalization of the time scale. Therefore, the continuous representation of the original variables in (3) implies the use of a thousand intermediate variables. In contrast, we have used in this paper functional methods in which the original waveforms are not discretized but are described by means of B-spline bases. This maintains the continuous nature of functional data without the need to use so many intermediate variables.

The application of functional PCA to the study of STS movement in two samples of people (patients with knee osteoarthritis and control group) shows how kinematic and kinetic variables can be simultaneously analyzed to define normal and pathological temporal patterns. From the original waveforms, we obtained four principal components explaining most of the total variance. Two of these components are related to differences in the movement patterns of osteoarthritic patients and healthy group.

The first rPC is associated to variability in the knee moment pattern. Positive scores correspond to waveforms with a marked peak of knee flexion moment and a final extension moment while the negative ones correspond to a more flat pattern with some degree of final flexion moment. This component could be associated to individual differences not related with osteoarthritis disease, since

the differences between control group and patients are not statistically significant. However, the related dispersion is higher in the patients group.

The second component indeed reveals differences between controls and patients. It is related to the more or less dynamic nature of the movement and to differences in the knee moment pattern. Waveforms with a flexion moment pattern are associated to the control group and distinguished from those with an extension pattern, representative of the patients' group. This is in agreement with results obtained by Su et al (22). They studied the STS movement of osteoarthritic patients and compared it with function of normal subjects and patients after successful total knee replacement. They observed a lower maximum flexion moment in osteoarthritic patients than in normal subjects. Moreover, flexion angle patterns are also different, patients having a smaller range of motion, lower initial flexion angle and slower motion performance than controls.

The third rPC also shows significant differences between both groups and is related to the variance of the flexion angle waveforms. Osteoarthritic group is characterized by a smaller range of motion, related to some remaining knee flexion after reaching the final standing posture. This remaining flexion might be caused by some degree of knee flexion contracture, characteristic of patients with knee osteoarthritis. Concerning the range of knee flexion, no other published studies have found a significant difference between osteoarthritic and healthy subjects while performing the STS movement. Nevertheless so far, the only published study on the topic of the effect of osteoarthritis in knee function during the STS movement performance is the one by (22). They just considered the maximal knee flexion angle as kinematic comparative variable between groups. Most published studies have mainly focused on function recovery after total knee arthroplasty (18, 27-29) or the effect of age on STS movement performance (20,

21, 26) . Jevsevar et al (18) found that patients after TKA had a smaller range of knee flexion and lower maximum flexion moments than healthy subjects.

Finally, the fourth component is not related to the disease either, but it is associated with individual differences in the range of flexion angle, as well as with phase differences during motion performance.

Thus, important differences were found in the knee mechanical behaviour of both groups. The biomechanical interpretation of these differences would be that patients develop a motion strategy with compensatory mechanisms in order to decrease or avoid the flexion moment on their diseased knee, since higher flexion moments entail higher joint compression forces as well as a higher force demand for the quadriceps and, subsequently, a pain increase. Su et al (22) have described some of these compensatory manoeuvres to reduce peak knee flexion moments like increasing horizontal anterior displacement and velocity of the body centre of mass, leaning more forward, increasing hip flexion angle or reducing knee flexion angle.

On the other side, knee moment and angle patterns observed in the control group are very similar to those obtained in other studies of STS movement in normal subjects (19, 30).

The scores of the four rPCs are used as input variables in the discriminant analysis. This analysis successfully separates the controls and patients patterns with a misclassification error rate of 6.5%. This rate is similar to the one obtained in (3) to classify controls and osteoarthritic gait patterns from 25 PCs obtained by means of conventional multivariate PCA, which was 6%. This result suggests that the main differences between movement patterns could be captured by means of a reduced set of kinematic and kinetic waveform measures: adding many more

variables could bring only redundant information that does not improve the discriminatory power of the variables set.

The results of this study reveal important differences between knee osteoarthritic patients and healthy subjects related to flexion-extension moment patterns as well as range of flexion angle during STS movement performance. This means that STS movement analysis can be applied as a functional evaluation tool of knee osteoarthritis. The application of functional PCA in this particular case allowed the identification of two biomechanical features useful for detection and assessment of knee osteoarthritis. That is, calculation of rPC scores from knee flexion angle and moment waveforms would allow classification of new subjects as well as evaluation of their function according to their proximity to each one of the considered groups.

A relevant question in the application of functional PCA is the role of the time normalization scale. Usually, the original waveforms are normalized in order to obtain a 0-100 time scale. This normalization can greatly affect the variability of original data and consequently to the results of PCA. The results of our study confirm this hypothesis and different PCs are obtained from the same data depending on whether time scale is normalized or not. It can be questioned which of the kinds of analyses represents better the original information. Both of them result on four PCs and the amounts of variability captured by both set of factors are very similar. Nevertheless, the normalization changes the timing information and can alter the interpretation of the meaning of PCs. For example, the slope of angle waveforms is directly related to the speed of motion in the raw data but have no physical meaning after time normalization. Because the magnitude of peaks of kinetic variables can be associated to the speed of motion (throughout the inertia forces) it could be more appropriate to work with raw data, at least when

kinetic and kinematic variables are included in the same analysis. A comparison between rPC2 and nPC1 (figures 2 and 3) shows that both of them are associated to different patterns in knee moment and that negative scores imply marked peak of flexion moment. Nevertheless, the angle component of rPC2 shows how the higher flexion peak is associated to a faster motion performance. This information is lost in the angular component of nPC1.

On the other hand, the time normalization process does not improve the discriminatory power of the PCs. In fact, results of table 5 show that the rate of misclassification errors is a little smaller when rPCs are used as input variables in the discriminant analysis instead the normalized ones nPCs.

Functional PCA can be used as a first exploratory analysis useful to select discrete variables with clinical interest. For example, in our study only rPC2 and rPC3 reveal significant differences between controls and patients. Comparing, in those factors, the patterns associated to the patients (positive scores) and the controls (negative scores), it is possible qualitatively to identify some discrete variables useful for a later comparison between groups: initial knee flexion angle (figure 3c), final flexion angle (figure 3e), range of flexion, total duration of movement (figure 3c) and peak flexion moment (figure 3d). All these variables display significant differences between groups, except the total duration for which the dispersion is very large (see table 2). This way it is possible to improve the a priori selection of variables, a process that sometimes is carried out in a subjective way and that usually provides many correlated variables (3).

In any case, the information provided by functional PCA is more complete than that obtained from a conventional analysis by means of a finite set of discrete variables. The results of table 5 show that the classification achieved by using 10

discrete variables is worse than that obtained with only 4 PCs, both when using normalized waveforms as well when using raw data.

Finally, it is necessary to point out that, although functional PCA can be applied to a large number of functional variables, it is advisable to be conservative when selecting the variables to be introduced into the model. The use of a large set of time series usually does not provides more relevant information, but makes more difficult to interpret the results. In fact, functional PCA presents two difficulties that are common to any PCA problem: deciding on the number of components to extract in the analysis, and interpreting the components, which is not always an entirely straightforward matter.

ACKNOWLEDGEMENTS

This study has been partially supported by Spanish Government Grant DPI2006-14722-C02-01 (cofinanced by EU FEDER funds), CICYT MTM2005-08689-C02-02, TIN2006-10134 and Fundació Caixa Castelló P11B2004-15.

References

1. Deluzio KJ, Wyss UP, Zee B, Costigan PA, Serbie C. Principal component models of knee kinematics and kinetics: Normal vs. pathological gait patterns. *Human Movement Science*. 1997;16(2-3):201-17.
2. Daffertshofer A, Lamoth CJ, Meijer OG, Beek PJ. PCA in studying coordination and variability: A tutorial. *Clin Biomech (Bristol, Avon)*. 2004 May;19(4):415-28.
3. Astephen JL, Deluzio KJ. Changes in frontal plane dynamics and the loading response phase of the gait cycle are characteristic of severe knee osteoarthritis application of a multidimensional analysis technique. *Clin Biomech (Bristol, Avon)*. 2005 Feb;20(2):209-17.

4. Deluzio KJ, Wyss UP, Costigan PA, Sorbie C, Zee B. Gait assessment in unicompartamental knee arthroplasty patients: Principal component modelling of gait waveforms and clinical status. *Human Movement Science*. 1999;18(5):701-11.
5. Sadeghi H, Sadeghi S, Prince F, Allard P, Labelle H, Vaughan CL. Functional roles of ankle and hip sagittal muscle moments in able-bodied gait. *Clin Biomech (Bristol, Avon)*. 2001 Oct;16(8):688-95.
6. Raptopoulos LSC, S Dutra M, Castro Pinto FAN, Carlos de Pina Filho, A. Alternative approach to modal gait analysis through the Karhunen–Loève decomposition: An application in the sagittal plane. *J Biomech*. 2006;39(15):2898-906.
7. Troje NF. Decomposing biological motion: A framework for analysis and synthesis of human gait patterns. *Journal of Vision*. 2002;2(5):371–387.
8. Mouchnino L, Mille ML, Martin N, Baroni G, Cincera M, Bardot A, et al. Behavioral outcomes following below-knee amputation in the coordination between balance and leg movement. *Gait Posture*. 2006 Aug;24(1):4-13.
9. Li ZM, Tang J. Coordination of thumb joints during opposition. *J Biomech*. 2007;40(3):502-10.
10. Wrigley AT, Albert WJ, Deluzio KJ, Stevenson JM. Differentiating lifting technique between those who develop low back pain and those who do not. *Clin Biomech (Bristol, Avon)*. 2005 Mar;20(3):254-63.
11. Wrigley AT, Albert WJ, Deluzio KJ, Stevenson JM. Principal component analysis of lifting waveforms. *Clin Biomech (Bristol, Avon)*. 2006 Jul;21(6):567-78.
12. Ivanenko YP, Poppele RE, Lacquaniti F. Five basic muscle activation patterns account for muscle activity during human locomotion. *J Physiol (Lond)*. 2004;556(1):267-82.
13. Lamoth CJ, Daffertshofer A, Meijer OG, Lorimer Moseley G, Wuisman PI, Beek PJ. Effects of experimentally induced pain and fear of pain on trunk coordination and back muscle activity during walking. *Clin Biomech (Bristol, Avon)*. 2004 Jul;19(6):551-63.

14. Ramsay JO, Silverman BW. Applied functional data analysis: Methods and case studies. Springer; 2002.
15. Ramsay J, Dalzell C. Some tools for functional data analysis. Journal of the Royal Statistical Society. Series B (Methodological). 1991;53(3):539-72.
16. Ramsay JO, Silverman BW. Functional data analysis. Springer; 2005.
17. Page A, Epifanio I. A simple model to analyze the effectiveness of linear time normalization to reduce variability in human movement analysis. Gait Posture. 2007 Jan;25(1):153-6.
18. Jevsevar DS, Riley PO, Hodge WA, Krebs DE. Knee kinematics and kinetics during locomotor activities of daily living in subjects with knee arthroplasty and in healthy control subjects. Phys Ther. 1993;73(4):229.
19. Rodosky MW, Andriacchi TP, Andersson GB. The influence of chair height on lower limb mechanics during rising. J Orthop Res. 1989;7(2):266-71.
20. Ikeda ER, Schenkman ML, Riley PO, Hodge WA. Influence of age on dynamics of rising from a chair. Phys Ther. 1991;71(6):473.
21. Schultz AB, Alexander NB, Ashton-Miller JA. Biomechanical analyses of rising from a chair. J Biomech. 1992 Dec;25(12):1383-91.
22. Su FC, Lai KA, Hong WH. Rising from chair after total knee arthroplasty. Clin Biomech (Bristol, Avon). 1998 Apr;13(3):176-81.
23. Leardini A, Chiari L, Della Croce U, Cappozzo A. Human movement analysis using stereophotogrammetry. part 3. soft tissue artifact assessment and compensation. Gait Posture. 2005 Feb;21(2):212-25.
24. Page A, Ayala G, Leon MT, Peydro MF, Prat JM. Normalizing temporal patterns to analyze sit-to-stand movements by using registration of functional data. J Biomech. 2006;39(13):2526-34.
25. Conover WJ. Some methods based on ranks. Practical Nonparametric Statistics, 3rd ed. John Wiley & Sons, New York. 1998:269-467.

26. Hughes MA, Myers BS, Schenkman ML. The role of strength in rising from a chair in the functionally impaired elderly. *J Biomech.* 1996 Dec;29(12):1509-13.
27. el Nahass B, Madson MM, Walker PS. Motion of the knee after condylar resurfacing--an in vivo study. *J Biomech.* 1991;24(12):1107-17.
28. Mizner RL, Snyder-Mackler L. Altered loading during walking and sit-to-stand is affected by quadriceps weakness after total knee arthroplasty. *J Orthop Res.* 2005 Sep;23(5):1083-90.
29. Saari T, Tranberg R, Zugner R, Uvehammer J, Karrholm J. The effect of tibial insert design on rising from a chair; motion analysis after total knee replacement. *Clin Biomech (Bristol, Avon).* 2004 Nov;19(9):951-6.
30. Arborelius UP, Wretenberg P, Lindberg F. The effects of armrests and high seat heights on lower-limb joint load and muscular activity during sitting and rising. *Ergonomics.* 1992 Nov;35(11):1377-91.

TABLE 1. Variables selected for analysis: waveforms and associated discrete measures.

WAVEFORM	DISCRETE MEASURES
<p>FLEXION-EXTENSION MOMENT:</p> <p>(Positive values mean extension moment while negative values mean flexion moment)</p>	<p>M_{max} (%BWxShL): Maximum moment of the whole curve, equivalent to maximum extension moment</p> <p>$t_{M_{max}}$ (sec): time instant of maximum moment</p> <p>M_{min} (%BWxShL): Minimum moment of the whole curve, equivalent to maximum flexion moment</p> <p>$t_{M_{min}}$ (sec): time instant of minimum moment</p> <p>M_{mean} (%BWxShL): Mean moment (mean value of the curve)</p> <p>M_{final} (%BWxShL): Final moment (moment in the last instant)</p>
<p>FLEXION-EXTENSION ANGLE</p> <p>(Positive values mean flexion while negative values mean hyper-extension of the knee)</p>	<p>i_{flex} (degree): Initial flexion (flexion at the first point of the curve)</p> <p>f_{flex} (degree): Final flexion (flexion in the last point of the curve)</p> <p>$R_{o_{flex}}$ (degree): Range of flexion (difference between initial and final flexion values)</p>
<p>BOTH</p>	<p>t_{tot} (sec): Total duration of STS movement</p>

TABLE 2. Results of statistical analysis from discrete variables extracted from individual waveforms. First two columns show mean values and standard deviation for control and patients' groups. Last column shows the results of the Mann-Whitney U test to quantify differences between both groups (bold in this column indicates variables with significant differences).

VARIABLE (units)	Mean (std)		p-value
	CONTROLS	PATIENTS	
i_flex (degree)	61.7 (9.9)	47.7 (13.4)	0.007
f_flex (degree)	0.9 (4.0)	7.8 (9.0)	0.005
R_o_flex (degree)	60.7 (9.1)	40.0 (8.8)	0.000
Mmin x10 ⁻² (%BWxShL)	-14.7 (4.9)	-4.2 (4.2)	0.000
Mmax x10 ⁻² (%BWxShL)	2.3 (1.8)	3.7 (3.7)	0.398
Mmean x10 ⁻² (%BWxShL)	-5.2 (2.2)	0.1 (4.1)	0.001
Mfinal x10 ⁻² (%BWxShL)	1.6 (1.4)	1.7 (4.0)	1.000
t_Mmin (sec)	0.34 (0.09)	0.51 (0.47)	0.237
t_Mmax (sec)	0.99 (0.43)	0.81 (0.38)	0.353
t_tot (sec)	1.25 (0.2)	1.49 (0.64)	0.331

TABLE 3. Results of functional PCA made from normalized knee flexion angle and flexion/extension moment waveforms. Principal components obtained from time-normalized waveforms are denoted as nPC_i (i=1 to 4), in order to distinguish them from the ones obtained from raw waveforms (rPC_i). The first row describes the variances explained by the first four nPCs after varimax rotation. The importance of knee angle or moment variables within each nPC is quantified by the percentage of explained variance (second and third rows). The differences between control and patients groups (represented by the differences between the means of the PC scores for each group) are described in the last rows and they have been quantified by means of the Mann-Whitney U test.

Functional PC	nPC1	nPC2	nPC3	nPC4
Explained variance (%)	47.1	24.1	17.2	9.4
% explained by angle variable	0.7	2.1	20.9	96.3
% explained by moment variable	99.3	97.9	79.1	3.7
CONTROLS: mean (std)	-175.4 (163.4)	-74.9 (92.7)	-123.5 (77.7)	0.9 (54.1)
PATIENTS: mean (std)	83.5 (149.2)	35.7 (145.4)	58.8 (82.4)	-0.5 (99.9)
p-value	0.001	0.066	0.000	0.720

TABLE 4. Results of functional PCA made from raw knee flexion angle and flexion/extension moment waveforms, that is, without time normalization. Principal components obtained from raw waveforms are denoted as rPC_i (i=1 to 4). The first row describes the variances explained by the first four rPCs after varimax rotation. The importance of knee angle or moment variables within each rPC is quantified by the percentage of explained variance (second and third rows). The differences between control and patients groups (represented by the differences between the means of the PC scores for each group) are described in the last rows and they have been quantified by means of the Mann-Whitney U test.

Functional PC	rPC1	rPC2	rPC3	rPC4
Explained variance (%)	43.0	31.9	10.5	10.4
% explained by angle variable	3.0	6.3	93.5	94.0
% explained by moment variable	97.0	93.7	6.5	9.0
CONTROLS: mean (std)	-1.2 (10.2)	-25.5 (14.5)	-10.2 (5.8)	1.8 (7.6)
PATIENTS: mean (std)	0.6 (34.0)	12.2 (17.8)	4.9 (14.2)	- 0.9 (16.2)
p-value	0.882	0.000	0.002	0.513

TABLE 5. Results of the discriminant analysis of each approach. Comparison of corresponding misclassification error rates.

APPROACH	Misclassification error rate (%)		
	Patients	Controls	Total
Discrete Variables	9.5	30.0	16.1
Functional PCA (time normalized)	4.8	20.0	9.7
Functional PCA (raw waveforms)	0.0	20.0	6.5

Figure 1. Picture of the experimental set-up. Only technical markers are shown.

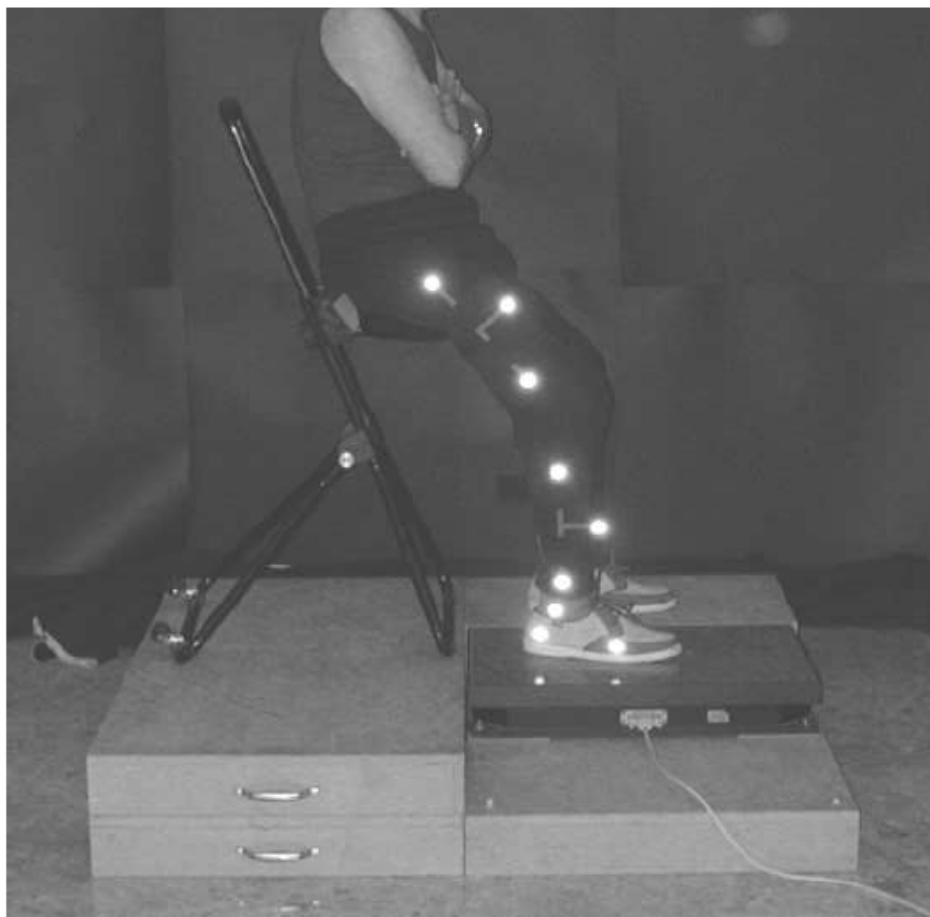


Figure 2. Effect of positive and negative scores of nPC_i on the shape of flexion angle (left column) and flexion-extension moment (right column) curves. Solid line represents the mean of time normalized waveforms; (+) line represents the effect on the mean curve of adding a standard deviation of nPC₁ scores (plots a and b), nPC₂ scores (plots c and d), nPC₃ scores (plots e and f) and nPC₄ scores (plots g and h), respectively; (-) line represents the mean minus a standard deviation of nPC_i scores.

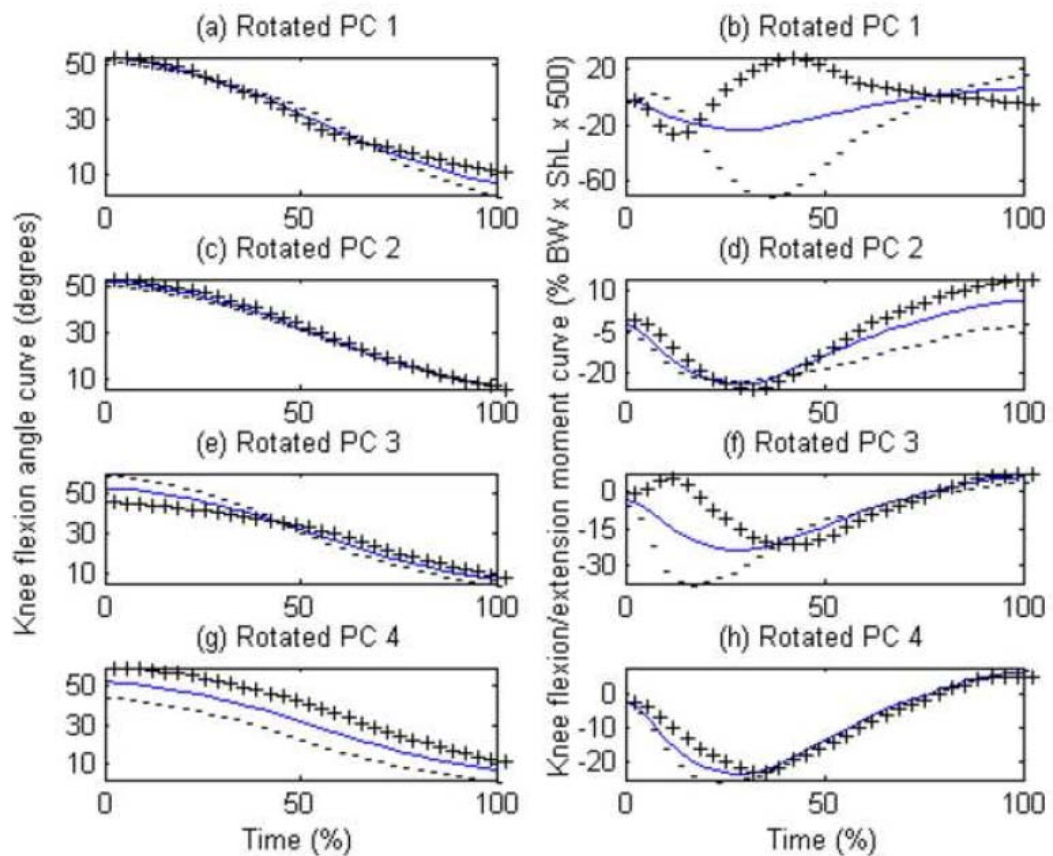


Figure 3. Effect of positive and negative scores of rPCi on the shape of flexion angle (left column) and flexion-extension moment (right column) curves. Solid line represent the mean of raw waveforms; (+) line represents the effect on the mean curve of adding a standard deviation of rPC1 scores (plots a and b), rPC2 scores (plots c and d), rPC3 scores (plots e and f) and rPC4 scores (plots g and h), respectively.

