Portuguese Journal of Sport Sciences
11 (Suppl. 2), 2011

# IS PRINCIPAL COMPONENT ANALYSIS MORE EFFICIENT TO DETECT DIFFERENCES ON BIOMECHANICAL VARIABLES BETWEEN GROUPS?

**Giulia Mantovani[1], Mario Lamontagne[1], Daniel Varin[1], Giuliano G. Cerulli[2] and Paul E. Beaulé[3]**

**University of Ottawa, Ottawa ON Canada[1]**
**Let People Move, Perugia Italy[2]**
**Orthopaedic Surgery, Ottawa Hospital, Ottawa, ON Canada[3]**

The biomechanical analysis investigates variables such as angles, inter-segmental forces and moments at the joints. When the relevant parameters (e.g., range of motion, peak values) are selected *a priori* from these variables, they could not perfectly represent the information content of the original dataset. Therefore, in this study we want to validate the efficacy of the Principal Component Analysis (PCA) in overcoming the limitations of the *a priori* selection of the parameters. An application study is reported; the lower-limb joint mechanics between patients operated with two different surgical techniques for a total hip arthroplasty are analyzed with both the traditional analysis and the PCA. The findings from the two methods converged, but the PCA identified new sources of variability not previously detected.

**KEY WORDS:** total hip arthroplasthy, lower-limb, sources of variance, time shift.

**INTRODUCTION:** The biomechanical analysis is defined as an investigation of movement and the forces producing the movement (Lamontagne, Beaulieu, et al., 2009). This is the most comprehensive mean that allows sport practitioners and clinicians to quantify possible functional limitations and the efficiency of the treatments. The angles, the inter-segmental forces and the moments at the joints are estimated by mean of kinematic and kinetic models. How can this information be useful to examine the patients or the athletes? A common technique is the extraction of some relevant discrete parameters, such as the range of motion, the peak and the zero crossing values (Lamontagne, Kennedy, et al., 2009). The statistical analysis is adopted to find significant differences among groups (*e.g.*, male/female, impaired/control). This traditional discrete analysis requires an *a priori* selection of the parameters. It does not analyze the whole waveform but only some relevant points, therefore, a large part of the information is lost. Conversely, the *a priori* selection does not prevent from having correlated parameters. In order to avoid either the redundancy or the loss of information, new methods have been explored in recent years (Chau, 2001a, 2001b).

In this paper we focus on the Principal Component Analysis (PCA), a multivariate linear statistical analysis. The PCA is adopted to derive efficient representation of the original dataset and to retain potentially valuable temporal information. Therefore, PCA is a valid mean to overcome the problems of the *a priori* selection of the parameters. Investigators demonstrated that the PCA can detect significant differences amongst groups of participants and the differences can be related to specific conditions by introducing an interpretation of the new obtained waveforms (Deluzio & Astephen, 2007; O'Connor & Bottum, 2009). In this study we want to validate the efficacy of the PCA in overcoming the limitations of the *a priori* selection of the parameters of the traditional discrete analysis. Our hypotheses are that PCA can extract interpretable variables from the dataset and detect differences between groups of participants in a more efficient way than the traditional discrete analysis.

An application study is reported. The purpose of the study is to use the PCA in order to compare lower-limb joint mechanics between two groups of patients operated for a Total Hip Arthroplasty (THA) with two different surgical approaches. The anterior approach (ANT) spares the gluteus medius and minimus, as opposed to the lateral approach (LAT). Therefore, the initial hypothesis is that ANT patients exhibit fewer differences than LAT patients, if compared to a control group (CON) (Matta & Ferguson, 2005; Mulliken et al., 1998). A traditional discrete analysis has also been applied to same dataset (Varin, 2011) and the differences in the findings are reported in this paper.

**METHODS:** Mathematically, PCA consists of an orthogonal transformation that converts the input variables $\overline{X} = [x_1, x_2, ..., x_m]^T$ into the new uncorrelated variables or principal components (PCs) $\overline{Y} = [y_1, y_2, ..., y_m]^T$ . The transformation is defined by the equation $\overline{Y} = A^T \overline{X}$, where the columns of the matrix $A = [A_1, A_2, ..., A_d]^T$ are the first $d$ eigenvectors (sorted from the largest eigenvalue, $\lambda_i$ ) of the covariance matrix of $\overline{X}$ , with $d < m$ . In doing this, the matrix $A$ contains the majority of the information and the original dataset can be represented by only $d$ values instead of $m$ . Each value of $\overline{Y}$ is named *score coefficient*, and the columns of $A$ are the *loading coefficients* (Jolliffe, 2002). It has been shown that the *loading coefficients* can be interpreted as biomechanical parameters (*e.g.*, range of motion, time shift). Therefore, it is possible to consider these *loading coefficients* instead of the original gait waveform. From *loading coefficients*, $\overline{Y}$ can be calculated and compared among the three groups of participants for each biomechanical waveform by using a statistical analysis.

**Application Study**: The study involved 60 participants: 20 ANT patients, 20 LAT patients and 20 CON participants, matched for age and BMI. Three-dimensional kinematics and kinetics were acquired for each participant while performing three trials of gait, as described in (Varin, 2011).

In the traditional discrete analysis, a series of multiple analyses of variance were used to compare peak angles and range of motion for kinematics, as well as peak moments of force and powers for kinetics (Varin, 2011).

In the PCA approach, the original gait waveforms of all 60 subjects were horizontally linked so that the initial dataset for each of the 21 original variables was a matrix $\overline{X}_{101,60}$, where 101 was the number of points of the gait cycle percentage and 60 was the number of participants (Dona et al., 2009). The first $d$ PCs necessary to reach a percentage of cumulative variance, $\left(\sum_1^i \lambda_i \big/ \sum_1^m \lambda_i\right) \cdot 100$, larger than 85% were retained, where the cumulative variance is the cumulative sum of the eigenvalues ( $\lambda_i$ ) relative to the retained eigenvectors.

Statistics was applied directly to the score coefficients of the 67 PC: one-way ANOVA and Tukey's honestly significant difference criterion were employed to compare the LAT and ANT groups to the CON group. A point system was employed. Each statistically significant difference (95% confidence interval) was worth one point and was added to the group, while if no significant difference was found, no points were added. Therefore, a low score for a patient group indicated a similarity to the CON group.

To help the interpretation of the PCs, they have been represented as suggested by Ramsay (Figure 1), who portrayed the effect of a PC about the mean curve of the original signal by adding (+) and subtracting (-) a multiple $\alpha$ of the PC *loading coefficients* (Ramsay & Silverman, 2002).

**RESULTS AND DISCUSSION:**

The findings from the traditional discrete analysis (Varin, 2011) show that both groups are significantly different from CON for the following 6 parameters: hip and ankle peak angles, range of motion at the hip angle, hip peak abduction and external rotation moments and knee peak moment in flexion. Conversely, the ANT group was found to be the only statistically different group for pelvis obliquity angle, hip peak adduction angle and ankle peak flexion moment. The LAT group showed statistical differences for hip flexion angle at ipsilateral foot-strike, hip internal/external rotation at ipsilateral foot-off and knee peak extension at mid-stance. Therefore, the conclusion of the author was that the study did not demonstrate superior kinematic and kinetic data for either surgical approach.

In the present study, the PCA identified four main sources of variability from the gait waveforms: range of variance (ROV), average (AVG), pattern (PTR) and time shift (TS). When a ROV is identified, it means that one or several participants for a specific variable

present a larger variation compared to the others. The presence of AVG indicated a large variability in the averages of the original data set. PTR was found when original signals had very different patterns while TS were characterized by large time shifts among the original signals. The example in Figure 1 reports two ROV variables. PC2 acts on the range of motion in the late stance phase, while PC1 in the rest of the gait cycle. The relative *score coefficient* values are also reported in the scatter plot (Figure 2). The LAT group tents to have negative *scores* for both PC, meaning that its range of motion is reduced compared to the CON group. The post hoc analysis from the ANOVA confirmed this observation and found that the difference between CON and LAT is significant for both PC1 and PC2 (respectively p=.003 and p<0.001), while the difference between CON and ANT is not significant for any component (p=0.071 and p=0.094).
The same analysis was repeated for all 67 PCs obtained from kinematic and kinetic datasets. Figure 3 summarizes the results for each source of variability (columns) and for each lower-limb joint of the LAT and ANT groups, according to the score system explained in the method section. The findings did not show a net preference toward one of the two surgical techniques. For example, ROV showed an equal score for the LAT and ANT groups. However, when focusing solely on the hip, the range of variance in the ANT group was more similar to the CON group, compared to the LAT group. This observation confirmed the previous findings from traditional analysis methods that reported a better restoration of gait patterns for patients treated with an anterior approach.

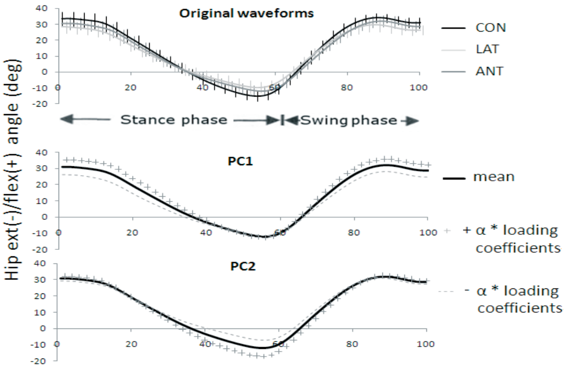


**Figure 1: The first graph reports the averages and the relative standard deviations of the three groups for the flexion/extension angle of the hip. The second and third graphs portrait the effect of the PC1 and PC2 on the mean of the original waveforms. The curve (+) corresponds to the mean plus the loading coefficients adjusted for a constant α. For the curve (-) the adjusted loading coefficients are subtracted to the mean. The second graph shows that the PC1 alters the mean by increasing (+) or decreasing (-) the flexion/extension range of motion at the hip in all the gait cycle but the late stance phase. The opposite observation is done for the PC2.**

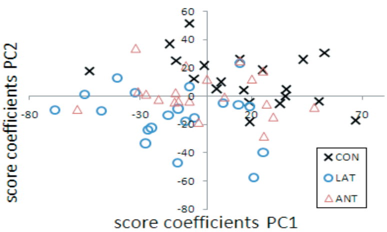


**Figure 2: The scatter plot reports the values of the *score coefficients* (PC1 vs. PC2) for the three groups. Even if the markers of different groups slightly overlap, three clusters are created. This means that the three groups are well distinct for the source of variability represented by the PC1 and PC2.**

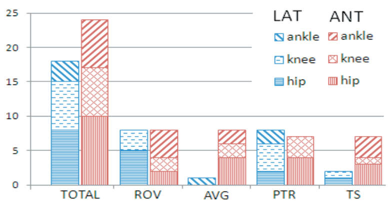


**Figure 3: The columns represent the scores obtained from the statistical analysis. Each point indicates a statistically significant difference between LAT and CON, or ANT and CON groups. The columns are divided into sections relative to the hip, knee and ankle. TOTAL represents the sum of all 67 statistical analyses. ROV: range of variance, AVG: average, PTR: pattern, and TS: time shift.**

The PCs related to PTR were mainly obtained from inter-segmental forces and power. The difference in the score between the two groups was of only one point. However, the point distribution among the joints was different. Specifically, for the ANT group, only the hip and knee patterns were significantly different from the CON group, whereas, for the LAT group, all three lower-limb joints were significantly different. Therefore, the patterns of inter-segmental forces and power were distributed differently between the LAT and ANT groups to compensate for the alterations in the gait patterns caused by the THA.
For the other variables such as the AVG and the TS, the statistical analysis indicated that the ANT group was significantly different from the CON group in much more cases than the LAT group, with a ratio of 8:1 and 7:2 for AVG and TS, respectively. Interestingly, one of the main findings that did not emerge from previous traditional analyses was the time shift in the ANT group signals, particularly in flexion/extension angles and anterior/posterior forces at the hip. While there is no clear-cut reason explaining the differences in the AVG, the results regarding the TS could be attributed to the surgical approach. The space created to expose the acetabulum and the femur required the superficial splitting of the interval between the tensor fasciae latae and sartorius. Since the tensor fasciae latae acts as a hip flexor in the early and second half of the stance phase, its weakening, due to the THA, could explain the delays in flexion/extension angles and anterior/posterior forces, observations which have never been reported by previous studies.

**CONCLUSION:** The findings from the traditional discrete analysis and from the PCA converge to the same conclusion that neither technique restores the normal gait pattern. However, in contrast with the initial hypothesis of the application study, the results from the PCA suggest that the LAT gives a slightly better score than the ANT. This difference is mainly due to the new sources of variability detected by the PCA and not emerged from the traditional discrete analysis. Nevertheless, this method has some limitations. Specifically, each source of variability has the same impact on the final score, even though it does not have the same functional and clinical relevance in the gait. Therefore, the next challenge is to identify the principal components that are more related to gait functionality.

**REFERENCES:**
Chau, T. (2001a). A review of analytical techniques for gait data. Part 1: Fuzzy, statistical and fractal methods. *Gait Posture, 13*(1), 49-66.
Chau, T. (2001b). A review of analytical techniques for gait data. Part 2: neural network and wavelet methods. *Gait Posture, 13*(2), 102-120.
Deluzio, K. J., & Astephen, J. L. (2007). Biomechanical features of gait waveform data associated with knee osteoarthritis: an application of principal component analysis. *Gait Posture, 25*(1), 86-93.
Dona, G., Preatoni, E., Cobelli, C., Rodano, R., & Harrison, A. J. (2009). Application of functional principal component analysis in race walking: an emerging methodology. *Sports Biomech, 8*(4), 284.
Jolliffe, I. (2002). *Principal component analysis*: Springer Verlag.
Lamontagne, M., Beaulieu, M. L., Varin, D., & Beaulé, P. E. (2009). Gait and motion analysis of the lower extremity after total hip arthroplasty: What the orthopedic surgeon should know. *Orthop Clin North Am, 40*(3), 397-405.
Lamontagne, M., Kennedy, M. J., & Beaule, P. E. (2009). The effect of cam FAI on hip and pelvic motion during maximum squat. *Clin Orthop Relat Res, 467*(3), 645-650.
Matta, JM & Ferguson, TA (2005) The anterior approach for hip replacement. *Orthopedics, 28*(9), 927
Mulliken, B. D., Rorabeck, C. H., Bourne, R. B., & Nayak, N. (1998). A modified direct lateral approach in total hip arthroplasty: a comprehensive review. *J Arthroplasty, 13*(7), 737-747.
O'Connor, K. M., & Bottum, M. C. (2009). Differences in cutting knee mechanics based on principal components analysis. *Medicine and Science in Sports and Exercise, 41*(4), 867-878.
Ramsay, J., & Silverman, B. (2002). *Applied functional data analysis: methods and case studies*: Springer Verlag.
Varin, D. (2011). *Kinematics and Kinetics of Total Hip Arthroplasty Patients during Gait and Stair Climbing: A Comparison of the Anterior and Lateral Surgical Approaches*. Masters, Univ. of Ottawa.

*Acknowledgements:*
This grant was partly supporter by the Canadian Institutes of Health Research and Let People Move.