CLUSTERING VERTICAL GROUND REACTION FORCE CURVES PRODUCED DURING COUNTERMOVEMENT JUMPS

Chris Richter[a,b,c,d,*], Noel E. O'Connor[a,b], Brendan Marshall[a,c,d], Kieran Moran[a,c]

[a] *Insight Centre Data Analytics, Dublin, Ireland*
[b] *CLARITY: Centre for Sensor Web Technologies, Dublin City University, Dublin, Ireland*
[c] *Applied Sports Performance Research, School of Health and Human Performance, Dublin City University, Dublin, Ireland*
[d] *Sports Surgery Clinic, Santry Demense, Dublin 9, Ireland*

Abstract

The aim of this study is to assess and compare the performance of commonly used hierarchical, partitional (k-means) and Gaussian model-based (Expectation-Maximization algorithm) clustering techniques to appropriately identify subgroup patterns within vertical ground reaction force data, using a continuous waveform analysis. In addition, we also compared the performance across each technique using normalized and non-normalization input scores. Both generated and real data (one hundred-and twenty two vertical jumps) were analyzed. The performance of each cluster technique was measured by assessing the ability to explain variances in jump height using a stepwise regression analysis. Only k-means (normalized scores; 82 %) and hierarchical clustering (normalized scores; 85 %) were able to extend the ability to describe variances in jump height beyond that achieved using the group analysis (i.e. one cluster; 78 %). Further, our findings strongly indicate the need to normalize the input data (similarity measure) when clustering. In contrast to the group analysis, the subgroup analysis was able to identify cluster specific phases of variance, which improved the ability to explain variances in jump height, due to the identification of cluster specific predic-

[*] Corresponding author
Email address: chris.richter@dcu.ie (Chris Richter)

tor variables. Our findings therefore highlight the benefit of performing a subgroup analysis and may explain, at least in part, the contrasting findings between previous studies that used a single group level of analysis.

Keywords: clustering, vertical ground reaction force, analysis of characterizing phases, countermovement jump

1. Introduction

The countermovement jump (CMJ) is an important task in a number of sports (e.g. volleyball, basketball) and its biomechanics have been frequently studied [16]. However, identified features that relate to the performance outcome (jump height) are often inconsistent [28]. For example, maximum vertical ground reaction force (vGRF) is reported in some studies as a performance related factor [4, 8, 30], while it is not in others [19, 21, 24]. This makes it difficult to conclude which neuromuscular capacities or movement techniques should be altered to enhance jump height, the criterion performance outcome in CMJs. Recently, we have shown that some of the contrasting findings across studies may be due to the use of discrete point analysis [28]. An alternative to discrete point analysis is a continuous waveform analysis (e.g. functional principal component analysis or analysis of characterizing phases) which has grown in popularity within many disciplines, including biomechanics, and has been reported to provide a better insight than discrete point analysis [6, 7, 9, 11, 20, 26, 28, 29].

An additional reason for the inconsistencies across studies however, may be inter-subject variability. Vertical ground reaction curves generated during a CMJ can differ significantly in shape across subjects (e.g. non-modal, uni-modal or bi-modal), which could imply that different movement strategies are being employed, which may in turn have different performance related factors. This might explain some of the contrasting findings, since previous studies generally employed a single group analysis which can mask performance related factors if different shapes have different performance related factors [1, 32, 33]. An alternative to a single group analysis is a subgroup analysis, which classifies similar patterns (curve shapes or movement strategies) into subgroups; so called clusters. An optimal clustering maximizes the ability to predict the dependent variable (e.g. jump height) of a data set [10]. To the authors' knowledge it appears that none of the previous CMJ studies have used a subgroup analysis, while subgroup analyses have been frequently performed in studies that examine human gait [2, 15, 22, 23, 34, 35, 37].

A challenge in subgroup analysis is that a variety of clustering techniques exists that may result in different clusters [12, 13, 18, 39]. Additionally, while the number of studies that have used continuous waveform analysis in the area of biomechanics is increasing, little is known about the performance of different clustering techniques with continuous waveform analysis in biomechanics. The computed continuous features aim to represent the pattern of a curve over multiple phases of the movement cycle and can be highly collinear, which may influence results of some clustering techniques. Clustering approaches differ in their underlying assumptions and can be divided broadly into hierarchical, partitional and probabilistic clustering [12, 18, 39]. The advantage of hierarchical clustering techniques is that they provide a highly interpretable description of the hierarchy within the data (i.e. dendrogram) and do not require the number of clusters to be chosen prior to the analysis. However, the assignment of samples into clusters requires the generation of inter-point distances of the input data (where different approaches can give very different results) and imposes a hierarchical structure within the examined data [12, 18, 39]. In contrast, partitional clustering (e.g. k-means) can be performed without calculating inter-point distances, it is commonly used and is usually more suitable for large data sets [18]. However, k-means clustering also requires the user to choose the number of clusters (prior to analysis) and the construction of a dendrogram is computationally prohibitive [12, 13, 18, 39]. In addition, both hierarchical and partitional clustering techniques follow a deterministic process where the generated clusters and their members are somewhat dependent on the ordering of samples [39]. Consequently, a third method, model-based clustering might be more appropriate for classifying biomechanical data. Model-based clustering techniques assign individuals into clusters based on their fit to a given mathematical model. An often used model is the Gaussian mixture model [10], which assigns subjects into clusters based on the nature of the statistical inference, might be more appropriate for classifying movement strategies. Due to the variation in clustering approaches, and the relative novelty of classifying continuous biomechanical data / features, it is important to identify which clustering technique has the greatest ability to recognize and appropriately separate patterns within multiple curves.

The primary aim of this study is to assess and compare the performance of commonly used hierarchical, partitional and probabilistic clustering techniques to appropriately identify patterns within a sample of self-created curves (manipulated data set) and a sample of vGRF curves captured during countermovement jumps (real data set), using a continuous waveform analysis. A secondary aim is to examine if there are benefits to performing a subgroup analysis compared to the commonly used single group analysis when identifying vertical ground reaction vGRF factors related to jump height.

2. Methods

2.1. Data Set

Manipulated Data Set A random vGRF curve from the real data set (see below) was selected and used to create a sample of 100 manipulated curves, which contained three clusters to reflect some of the general shapes of the vGRF curve. Curves in the first cluster (n = 41) were manipulated to have a unimodal shape, where the peak value occurred from 25-30 % of the cycle. Curves in the second cluster (n = 9) were manipulated to have a unimodal shape, where the peak value occurred from 70-75 % of the cycle. Curves in the third cluster (n = 50) were manipulated to have a bimodal shape, where the peak value occurred from 75-80 % of the cycle (Figure 1). To generate the manipulated data set the randomly selected curve was transformed into a function, using seven coefficients and a b-spline basis system [5, 25]. The third (cluster 1 and 3) and fifth (cluster 2 and 3) coefficients were multiplied with a random factor between one and two, while the fourth coefficient (Cluster 3) was multiplied with a random number between minus one and zero. After altering the coefficients, manipulated curves were generated by solving the altered coefficients to 101 points. Subsequently, the peak position of each curve was shifted randomly in time, using a dynamical time warping approach, within a random range of -2.5 and 2.5 %. The used distribution was created *ipso facto* to model a realistic distribution, accounting for low frequent modal shapes.

Real Data Set One-hundred-and-twenty-two male athletes (age = 22.4 ± 4.2 years; mass = 71.1 ± 9.4 kg; height = 1.82 ± 0.1 m), who were physically active, experienced in performing the countermovement jump (based on the sports they played: Gaelic football, hurling and basketball), and free from lower limb injury participated in this study. The University Ethics Committee approved the study and all participants were informed of any risk and signed an informed consent form before participation.

Prior to data collection, every participant performed a standard warm-up routine consisting of low intensity jogging, stretching and ten sub-maximal and five maximal effort countermovement jumps. Each participant performed 15 jumps without an arm swing, standing with each foot on a separate force platform. Participants rested for 30 seconds between trials. Two force plates

(BP-600900, AMTI, MA, USA) recorded the vGRF (1000Hz). Based on jump height, the best jump performance of each subject was identified and used for analysis[1]. Jump height was calculated using the center of mass velocity at takeoff, with take-off determined when the vGRF fell below 5 N [28]. The position of the center of mass was calculated using a motion analysis system (Vicon 512 M, Oxford Metrics Ltd, England) to record the position of twelve reflective markers (250Hz), in combination with anthropometric data [38]. Reflective markers were attached bilaterally, using double sided tape, on the following anatomical landmarks: fifth metatarsal joint, posterior calcaneus (in line with the fifth metatarsal joint), lateral malleolus, lateral femoral epicondyle, greater trochanter and the glenohumeral joint. All curves were normalized to body mass and only the vGRF-time curve during the propulsion phase was analyzed because it holds the information needed to fully describe jump height. The start of the propulsion phase was identified from the power-time curve of the body's centre of mass, when the power became positive.

2.2. Data Clustering

To generate scores that capture the patterns within the continuous waveforms, an Analysis of Characterizing Phases was performed [28]. Analysis of Characterizing Phases detects phases of variation (key phases) within the

[1] The best jump was used because it is a well-defined criterion and avoids taking an average of multiple curves which may have distorted the data.

sample of curves, which are used to generate participants' scores (similarity score). Similarity scores were computed for key phases using the magnitude domain. The number of similarity scores extracted for each waveform is equal to the number of identified key phases. Similarity scores were determined by calculating the area between a participant's curve (p) and the mean curve across the data set (q) for every point (i) within the key phases (Equation 1)[2].

$$similarity\ score = \int p_i - q_i \tag{1}$$

Key phases were identified using the information generated by the principal components needed to describe 99.5 % of the variances in the data [27]. To increase the interpretability of the retained principal components a VARIMAX rotation was performed [11, 26]. For further explanation of Analysis of Characterizing Phases the reader is referred to a previous paper by the authors [28]. Given that Analysis of Characterizing Phases generates just a few similarity scores to describe a complex waveform, it was necessary to insure that the generated scores preserve the information needed to cluster curves with similar patterns (shapes). The quality of the preserved information was estimated, for only the manipulated data set, by a subjective visual inspection of the generated similarity scores and was judged sufficient

[2] The used equation can result in a similarity score close or equal to zero when a subject and the reference signal are opposite or when a signal oscillates above and below the references signal. In the present study, the shape of the reference and subject curve followed a similar pattern within the key phases.

since a clear linear relationship exists for curves within each cluster (Figure 2). The reader should note that the calculation of subject score within the present paper differs slightly from Richter et al. [28] to overcome a dependency of the finding on the reference signal chosen. In Richter et al. [28] the best jump was selected as reference signal because the subject score calculation used absolute values to measure similarity. This approach assumes that altering a curve towards the reference signal has a positive effect on the dependent variable. However, this might not be true as other movement strategies might represent a better movement solution. The score generation approach used in the present paper overcomes this limitation and findings are not dependent on the reference signal. The overall mean was selected as the reference signal because it is commonly used and easy to relate to when interpreting the findings.

To classify the manipulated and real data sets the computed similarity scores were input into a hierarchical clustering algorithm (hierarchical clustering), a k-means approach (partitional clustering) and an Expectation-Maximization algorithm (model-based clustering). Due to the linear relationship between similarity scores within a cluster, where clusters could overlap in space possibly hampering the ability of the hierarchical and the k-means clustering, the hierarchical and the k-means clustering were also performed using normalized similarity scores (as suggested in Jain et al. [13]). The normalization was performed by transforming the similarity scores into their correlation matrix (Equation 2), to quantify numerically the relationship between the similarity scores, which cannot be described by distances of the generated similarity scores. The correlation matrix ($\hat{P}; \hat{P} \in \mathbb{R}^{122x122}$) was created by calculating the Pearson's r-value ($corr$) utilizing the similarity scores (SS) of the curves i ($i = 1, 2, \ldots,$ number of curves) and j ($j = 1, 2, \ldots,$ number of curves).

$$[\hat{P}]_{(i,j)} = corr_{(i,j)} = \frac{1}{N-1} \sum_{k=1}^{N} \frac{(SS_{i,k} - \mu_i) * (SS_{j,k} - \mu_j)}{\sigma_i * \sigma_j} \tag{2}$$

where μ is the average and σ the standard deviation for curve i and j of their corresponding similarity scores, which were calculated using the identified key phases ($k = 1, 2, \ldots, N$, where N is the number of identified key phases).

The hierarchical algorithm calculated pairwise distances using Euclidean distance, and created a hierarchical cluster tree using the nearest distance [18]. The quality of the hierarchical clustering was measured by calculating the cophenetic correlation coefficient between the hierarchical cluster tree and the pairwise distances [18, 31]. Hierarchical clustering properties were changed if the cophenetic correlation coefficient was less than 0.7, which indicates a low or medium correlation between the hierarchical cluster tree and the pairwise distances[3] [3]. The k-means clustering technique used the squared Euclidean distance as the distance measure and the Expectation-

[3] All generated hierarchical cluster trees and the pairwise distances generated a cophenetic correlation coefficient above 0.7

Maximization algorithm was applied using the Gaussian mixture model [18].

For the manipulated data, the performance of each clustering technique was assessed by the percentage of accurately classified curves, assessed by counting how often the assigned membership and the actual membership of a curve matched. To examine the benefits of using a subgroup analysis, key phases were identified using both a single group and a subgroup analysis, and directly compared. The number of clusters in the subgroup analysis was set at three clusters due to the contained number of general shapes (three shapes).

For the real data set, the performance of each clustering technique was measured by assessing the ability to explain variances in jump height (dependent variable) across generated clusters. This approach was based on the assumption that an appropriate grouping of vGRF curve shapes (or similar movement strategies) does not mask performance related factors and hence enhances the ability to describe variances in jump height. To assess the ability to explain variances in jump height for a given number of clusters the average r^2-value of a stepwise regression analysis was computed across these clusters. The clustering technique with x clusters that generated the highest ability to explain variances in jump height was considered the most appropriate clustering technique for the captured vGRF curves. Input variables for the regression model were similarity scores measured solely over the key phases of a cluster. During the clustering process two problems can occur for a given cluster solution: (a) the regression analysis does not identify a predictor variable and, (b) only one subject is assigned to a cluster. If the stepwise regression analysis was not able to identify any predictor variables, the highest r^2-value computed during the correlation analysis between the generated similarity scores and jump height was used (irrespective of whether it was statistically significant or not)[4]. If a given cluster solution assigned only one participant to a cluster, the cluster and its member were considered as an outlier and removed from the analysis.

If the stepwise regression analysis was not able to identify any predictor variables within a cluster, the highest r^2-value (irrespective of its significance) computed during the correlation analysis (between the generated similarity scores and jump height) was used. If a cluster technique assigned only one participant to a cluster, the cluster was discarded.

To examine the benefits of a subgroup analysis over a single group analysis both the key phases and the predictor variables were compared when calculated for the whole data set (single group) to the key phases the predictor variables selected within each of the generated clusters (subgroup analysis). The number of clusters was set to increase from one to ten clusters. All statistical analyses were performed using MatLab (R2012a, MathWorks Inc., USA).

[4] It should be noted that for the cases where no predictor variable was identified by the regression analysis, the sample size of the corresponding cluster was low and the correlation of an independent variable to the dependent variable was not high enough to reach a significant correlation.

3. Results

3.1. Manipulated Data Set

For the manipulated data set, the accuracy of the clustering techniques was (from high to low): hierarchical clustering utilizing normalized scores (98 % accuracy), k-means clustering utilizing normalized scores (97 % accuracy), Expectation-Maximization algorithm (95 % accuracy), hierarchical clustering utilizing similarity scores (67 % accuracy) and k-means clustering utilizing similarity scores (61 % accuracy).

Key phases differ between the single group and subgroup analysis. Key phases for the whole group analysis were identified at 20-30 %, 45-57 % and 72-82 % of the movement cycle. The key phases for each cluster, examined using a subgroup analysis were identified at 22-36 % and 82-91 % for cluster 1, 55-67 % and 78-87 % for cluster 2, and 60-68 %, and 81-89 % of the movement cycle for cluster 3.

3.2. Real Data Set

For the real data set, predictor variables (similarity scores computed from key phases), identified by the stepwise regression analysis, were able to explain 78 % of the variances in jump height ($r^2 = 0.78$). Hierarchical clustering (normalized scores) best described jump height using four clusters (85 %) and k-means (normalized scores) performed best using four clusters (83 %). The Expectation-Maximization algorithm, hierarchical clustering (similarity scores) and the k-means (similarity scores) were not able to increase the ability to describe jump height over that achieved using the single group analysis (Figure 3).

Hierarchical (normalized scores) clustering explained most accurately the variances in jump height but generated two clusters with sample sizes less than ten members (Cluster 1 = 7; Cluster 3 = 6). For the clusters with small sample sizes, the regression analysis was not able to identify predictor variables. Hence, k-means (normalized scores) clustering was selected for further analysis, as it had almost the same ability to describe variance in jump height with larger sample sizes and better-balanced cluster sizes. Visual inspection of the mean curves of the generated k-means (normalized scores) clusters indicates four distinct vGRF curve shapes: (cluster 1) unimodal with high initial vGRFs where peak vGRF occurs shortly after the start of the concentric phase, (cluster 2) unimodal with low initial vGRF where peak vGRF occurs at about 70 % of the movement cycle, (cluster 3) bimodal with high initial vGRFs where peak vGRF occurs shortly after the start of the concentric phase, and (cluster 4) bimodal with initial vGRFs similar to both the first and second maxima where peak vGRF could occur either before 15 % or around 80 % of the movement cycle (Table 1; Figure 4). No significant difference exists in jump height across the clusters.

Key phases and identified predictor variables differed between the single group and subgroup analysis, while the strongest relation to jump height occurred at around 85 % across both subgroup and single group analysis (Figure 5). All predictor variables were identified by the stepwise regression analysis. The reader should note that the subgroup analysis was able to increase the ability to describe jump height, while using fewer data points (a smaller percentage) of the movement cycle.

4. Discussion

4.1. Clustering Technique Comparison

The examined clustering techniques differed in their performance in both the manipulated and real data sets. Using the manipulated data, the hierarchical clustering utilizing normalized scores, k-means clustering utilizing normalized scores, and Expectation-Maximization algorithm performed best. Using the real data set, only k-means (normalized scores) and hierarchical clustering (normalized scores) extended the ability to describe variances in jump height beyond that achieved using the group analysis (e.g. one cluster). With respect to the Expectation-Maximization algorithm, it was not able to generate clusters with a higher ability to describe variances in jump height than that achieved at a single group level (i.e. one cluster). While the Expectation-Maximization algorithm was successful for the generated data set, it failed to successfully classify the real data. A possible reason for this contrasting performance lies in the nature of both data sets. The manipulated data set holds clear distribution patterns where peak vGRF differed across curves within a cluster by only ± 5 %. The real data set, however, has much more variation and the probability distribution does not differ as clearly across clusters (Figure 6).

4.1.1. Benefits of Normalizing Data

Normalizing similarity scores (transformation of scores into their correlation matrix) had a significantly positive effect on the performance of both hierarchical and partitional clustering techniques, indicating that differences in magnitude between similarity scores are not as effective as their quantified numerical relationship at maximizing the ability to predict a dependent variable. The same effect is likely to occur when discrete points are used for clustering individuals. To the best of our knowledge, previous studies that aimed to identify movement patterns by clustering discrete kinematic and kinetic variables did not normalize their input variables, which may have reduced their ability to recognize movement patterns [2, 15, 17, 22, 34]. To date, no study has compared clustering approaches using biomechanical waveforms, which makes it difficult to control the effect of normalizing the input data. For this reason we applied k-means clustering to a publicly available data set (The Berkeley Growth Data: Tuddenham and Snyder [36]). The Berkeley Growth Data has been used to measure the accuracy of k-means clustering (e.g. Jaques and Preda [14]) and, similar to vGRF curves, the shapes of the sample of curves might hold the information needed to classify the data correctly. Applying k-means to the Berkeley Growth Data using non-normalized and normalized similarity scores resulted in clustering accuracies of 74.2 % and 94.6 %, respectively. In the experiment of Jaques and

Preda [14][5], the highest accuracy of k-means was 66.7 %. The increase in accuracy of k-means in the present work is due to the effect of normalization (accounting for ± 20.4 %) and the use of similarity scores (accounting for ± 7.5 %). The contrasting findings between non-normalized and normalized scores for hierarchical and partitional techniques (for the manipulated, real and Berkeley Growth data) strongly suggest that input variables should be normalized when classifying curves where the curve shape might hold important information. It should be noted, however, that other normalization approaches (e.g. Euclidian distance) may lower the ability to recognize shape pattern.

4.2. Benefits of Subgroup Analysis

With respect to the benefit of performing a subgroup analysis, the subgroup analysis alone was able to capture key phases, which reflect specific characteristics of each cluster, resulting in different locations of key phases and predictor variables across clusters. These differences (Figure 5) resulted in a greater ability of the subgroup analysis to describe variances in jump height over a group level analysis (on average +8.3 %). In addition to this increased ability to describe variances in jump height, the subgroups required less information (less % of the data) to predict jump height (on average 17 % less of the movement cycle). While previous CMJ studies have not exam-

[5] Jaques and Preda [14], assessed the ability of k-means using non-normalized data (whole discrete curve, 20 spline coefficients and functional principal component scores)

ined the effectiveness of a subgroup analysis, gait studies have also shown its appropriateness over a single group analysis [2, 15, 22, 34].

The subgroup analysis was able to identify four distinct vGRF curve shapes. The characteristics of these clusters strengthen the idea that different individuals may have different performance related factors [1, 32, 33]. The combination of the knowledge of general curve shapes and the location of performance related factors gives a further insight into inconsistencies in respect to maximum vGRF reported in some discrete point analysis studies as a performance related factor [4, 8, 30], while not in others [19, 21, 24]. In light of the subgroup findings, maximum vGRF represents different neuromuscular capacities across each cluster. For cluster 1 and 2 (shapes with low initial vGRFs), maximum vGRF represents the ability to generate vGRFs at the end of the movement cycle as the ankle, knee and hip joint extend towards full extension; while it represents the ability to generate vGRFs quickly (1-15 %) after the start of the concentric phase for cluster 3 and 4. Consequently, maximum vGRF cannot be compared using a single group analysis because even if an analysis of peak vGRF accounts for different modalities of a vGRF curve, it can fail to examine comparable neuromuscular capacities. The present work indicates that classifying a sample of individuals into multiple clusters can overcome limitations of a group analysis and hence enhances the understanding of the underlying neuromuscular movement's strategies during a movement task.

5. Conclusion

K-means clustering utilizing normalized subject scores appears to be the most suitable technique for clustering vGRF curves, while hierarchical clustering also showed a high level of suitability. Further, when clustering curve shapes, it is extremely important to normalize subject scores, by transforming them into their correlation matrix, before using a clustering technique. The subgroup analysis should be used in preference to a single group analysis because it explained greater variances in the dependent variable (jump height), indicating different movement strategies for which some different performance determining factors were evident. These findings may explain, at least in part, the contrasting findings between previous studies that examined vGRF during vertical jumping at the single group level of analysis.

6. Conflict of interest statement

The authors declare that no conflict of interest is associated with the present study.

7. Acknowledgements

This work is supported by Science Foundation Ireland under grant 07 / CE / I114.

8. References

[1] Bates, B. T., 1996. Single-subject methodology: an alternative approach. Medicine & Science in Sports & Exercise 28 (5), 631–638.

[2] Carriero, A., Zavatsky, A., Stebbins, J., Theologis, T., Shefelbine, S. J., 2009. Determination of gait patterns in children with spastic diplegic cerebral palsy using principal components. Gait & Posture 29 (1), 71–75.

[3] Cohen, J., 1988. Statistical power analysis for the behavioral sciencies. Routledge.

[4] Cormie, P., McBride, J. M., McCaulley, G. O., 2009. Power-time, force-time, and velocity-time curve analysis of the countermovement jump: impact of training. The Journal of Strength & Conditioning Research 23 (1), 177–186.

[5] De Boor, C., 1978. A Practical Guide to Splines. Vol. 27 of Appl. Math. Sci. Springer-Verlag., New York.

[6] Dona, G., Preatoni, E., Cobelli, C., 2009. Application of functional principal component analysis in race walking: an emerging methodology. Sports Biomechanics 8 (4), 284–301.

[7] Donoghue, O. A., Harrison, A. J., Coffey, N., Hayes, K., Jul. 2008. Functional data analysis of running kinematics in chronic Achilles tendon injury. Medicine and Science in Sport and Exercise 40 (7), 1323–35.

[8] Dowling, J. J., Vamos, L., 1993. Identification of kinetic and temporal factors related to vertical jump performance. Journal of Applied Biomechanics 9 (2), 95–110.

[9] Godwin, A., Takahara, G., Agnew, M., Stevenson, J., 2010. Functional data analysis as a means of evaluating kinematic and kinetic waveforms. Theoretical Issues in Ergonomics Science 11 (6), 489–503.

[10] Han, J., Kamber, M., Pei, J., 2006. Data mining: concepts and techniques, 2nd Edition. Morgan kaufmann.

[11] Harrison, A. J., Ryan, W., Hayes, K., May 2007. Functional data analysis of joint coordination in the development of vertical jump performance. Sports biomechanics / International Society of Biomechanics in Sports 6 (2), 199–214.

[12] Hastie, T., Tibshirani, R., Friedman, J., 2001. The elements of statistical learning. Springer, New York.

[13] Jain, A., Murty, M., Flynn, P., 1999. Data clustering: a review. ACM computing surveys (CSUR) 31 (3), 264–323.

[14] Jaques, J., Preda, C., 2013. Functional data clustering: a survey. Tech. rep., Research Centre Lille Nord Europe, Lille, France.

[15] Kienast, G., Bachmann, D., Steinwender, G., Zwick, E. B., Saraph, V., 1999. Determination of gait patterns in children with cerebral palsy using cluster analysis. Gait & Posture 10 (1), 57.

[16] Klavora, P., 2000. Vertical-jump Tests: A Critical Review. Strength and Conditioning Journal 22 (5), 70.

[17] Marshall, B., 2010. Can a pre-training biomechanical pathway identify the most effective exercise to enhance a given groups, subgroups or individuals countermovement jump height ? . Phd, Dublin City University.

[18] Martinez, W., Martinez, A., Solka, J., 2004. Exploratory data analysis with MATLAB. CRC Press.

[19] Morrissey, M. C., Harman, E. a., Frykman, P. N., Han, K. H., 1998. Early phase differential effects of slow and fast barbell squat training. The American Journal of Sports Medicine 26 (2), 221–30.

[20] Newell, J., McMillan, K., Grant, S., McCabe, G., Mar. 2006. Using functional data analysis to summarise and interpret lactate curves. Computers in Biology and Medicine 36 (3), 262–75.

[21] Newton, R., Kraemer, W., Häkkinen, K., 1999. Effects of ballistic training on preseason preparation of elite volleyball players. Tech. Rep. 2, School of Exercise Science & Sport Management, Southern Cross University, Lismore, NSW, Australia. rnewton@scu.edu.au.

[22] O'Byrne, J. M., Jenkinson, A., O'Brien, T. M., 1998. Quantitative analysis and classification of gait patterns in cerebral palsy using a three-dimensional motion analyzer. Journal of Child Neurology 13 (3), 101–108.

[23] O'Malley, M. J., Abel, M. F., Damiano, D. L., Vaughan, C. L., 1997. Fuzzy clustering of children with cerebral palsy based on temporal-distance gait parameters. Rehabilitation Engineering, IEEE Transactions on, 5 (4), 300–309.

[24] Petushek, E., Garceau, L., Ebben, W., 2010. Force, velocity, and power adaptations in response to a periodized plyometric training program. In: ISBS-Conference Proceedings Archive. pp. 262–265.

[25] Ramsay, J. O., 2006. Functional data analysis. John Wiley & Sons.

[26] Ramsay, J. O., Silverman, B. W., 2002. Applied functional data analysis: methods and case studies. Springer,, New York.

[27] Richter, C., McGuiness, K., Gualano, L., O'Connor, N. E., Moran, K., 2013. Identification of an optimal principal components analysis threshold to describe jump height accurately using vertical ground reaction forces. In: XXIV Congress of the International Society of Biomechanics. Natal, Brazil.

[28] Richter, C., O'Connor, N. E., Marshall, B., Moran, K., 2013. Analysis of Characterizing Phases on Waveforms An Application to Vertical Jumps. Journal of applied biomechanics (in press).

[29] Ryan, W., Harrison, A., Hayes, K., 2006. Functional data analysis of knee joint kinematics in the vertical jump. Sports Biomechanics 5 (1), 121–138.

[30] Sheppard, J. M., Chapman, D. W., Gough, C., McGuigan, M. R., Newton, R. U., 2009. Twelve-month training-induced changes in elite international volleyball players. The Journal of Strength & Conditioning Research 23 (7), 2096–2101.

[31] Sokal, R. R., Rohlf, F. J., 1962. The comparison of dendrograms by objective methods. Taxon 11 (2), 33–40.

[32] Stergiou, N., 2004. Innovative Analyses of Human Movement , 1st Edition. Human Kinetics, Leeds, U.K.

[33] Stergiou, N., Scott, M., 2005. Baseline measures are altered in biomechanical studies. Journal of Biomechanics 38 (1), 175–178.

[34] Stout, J. L., Bruce, B., Gage, J. R., Schutte, L., 1995. Joint kinetic patterns in children with spastic hemiplegia cerebral palsy. Gait & Posture 3 (4), 274.

[35] Toro, B., Nester, C. J., Farren, P. C., 2007. Cluster analysis for the extraction of sagittal gait patterns in children with cerebral palsy. Gait & Posture 25 (2), 157–165.

[36] Tuddenham, R. D., Snyder, M. M., 1954. Physical growth of California boys and girls from birth to eighteen years. Publications in child development. University of California, Berkeley 1 (2), 183.

[37] von Tscharner, V., Enders, H., Maurer, C., Jan. 2013. Subspace identification and classification of healthy human gait. PloS one 8 (7), e65063.

[38] Winter, D. A., 2009. Biomechanics and motor control of human movement, 4th Edition. Wiley & Sons.

[39] Witten, I. H., Frank, E., 2005. Data Mining: Practical machine learning tools and techniques, 3rd Edition. Morgan Kaufmann.