ISB 2013 BRAZIL
XXIV CONGRESS OF THE INTERNATIONAL SOCIETY OF BIOMECHANICS
XV BRAZILIAN CONGRESS OF BIOMECHANICS

Identification of an optimal principal components analysis threshold to describe jump height accurately using vertical ground reaction forces

[1,2] Chris Richter, [1]Kevin McGuinness, [1]Leonardo Gualano, [1]Noel E. O'Connor and [2]Kieran Moran
[1] CLARITY: Centre for Sensor Web Technologies
[2] Applied Sports Performance Research, School of Health and Human Performance, Dublin City University, Dublin, Ireland

SUMMARY
In functional principal component analysis (fPCA) a threshold is chosen to define the number of retained principal components, which corresponds to the amount of preserved information. A variety of thresholds have been used in previous studies and the chosen threshold is often not evaluated. The aim of this study is to identify the optimal threshold that preserves the information needed to describe a dependent variable accurately. To find an optimal threshold, a neural network was used to predict jump height from vertical ground reaction force curve measures generated by a fPCA at different thresholds. The findings indicate that a threshold from 99% to 99.9% (6-11principal components) is optimal for describing jump height, as these thresholds generated significantly lower jump height prediction errors than other thresholds.

INTRODUCTION
The majority of studies in biomechanics have relied almost exclusively on a discrete point analysis that examines discrete measures (e.g. maximums, minimums, overall duration). Significant limitations in discrete point analysis are the pre-selection of parameters and the possible loss of extremely important information [1,2,3]. In recent years, functional principal component analysis (fPCA) has been proposed to avoid these limitations by examining continuous waveforms [4,5]. fPCA reduces the dimensionality of a data set by generating a number of principal components that preserve the information needed to fully describe a data set [6]. When applying fPCA, a threshold (*x*% of the total variance in the data) is chosen by the user, which defines the amount of information preserved and determines the number of retained principal components. A scree plot[1] can be used to estimate the optimal number of principal components. While a variety of thresholds are used, a 95% threshold appears to be the most frequent in recent biomechanical studies [3,7,8]. Principal components beyond the threshold of 95% are often discarded as they have very little influence on the data [1]. However, the captured influence of principal components in this context is assessed only in relation to the data rather than a dependent variable, which is extremely important in biomechanical analyses. To date, no bio-mechanical studies appear to have examined if there is an optimal threshold that retains sufficient information to best describe a dependent variable. The aim of this study is to: (a) identify the optimal threshold for jump height prediction and, (b) test if a principal component beyond the 95% threshold can have significant influence on the dependent variable.

METHOD
A feed-forward back-propagation neural network[2] with a single hidden layer containing 20 hidden units was used to identify the optimal threshold for inferring a dependent variable from an input matrix. A neural network was used to access the optimal fPCA threshold, as it is able to find any existing input-target relationship [9].

Dependent variable: The jump height of a counter-movement jump (CMJ) was chosen as the dependent variable because it is fully captured by vertical ground reaction force (force) generated during the propulsion phase of the jump. Force curves of 42 athletes were captured during CMJs. All athletes were free from any injury at the time of data capturing and were experienced in performing a CMJ. The University Ethics Committee approved the study and all subjects were informed of any risk and signed an informed consent form before participation. Prior to data collection, every subject completed a standard warm-up routine. The subjects performed 15 maximum effort CMJs without an arm swing, standing with each foot on a force platform and rested for 30 seconds between the trials. Two force plates (BP-600900, AMTI, MA, USA), each with a frequency of 250Hz, recorded the produced force. Jump height was calculated by the impulse momentum relationship. Based on jump height, the best jump performance of each subject was used for data analysis. All curves were normalized to body mass (N/BM) and only the propulsion phases were used for analysis.

Input matrix: fPCA [5] was performed to generate principal components for a given threshold using the captured force curves. fPCA was used because it does not require a linear time normalization, which can alter the data [2]. The generated principal components were VARIMAX rotated to optimize their interpretability [5,6]. Principal component scores were calculated to reflect the degree to

[1] A scree plot is a line that shows the ratio of the influence of a principal component on the data's total variance (Figure 1).

[2] MATLAB neural network toolbox implementation

which a subject is affected by a principal component over the whole function [5].

The fPCA threshold can be seen as a parameter of the jump height's prediction model. In machine learning the optimal value for such parameters is typically chosen using cross-validation [10]. A leave-one-out[3] cross validation was performed due to the relatively small sample size. The network was trained using the principal component scores as input data and jump heights as target data. After training, the principal component scores of the test sample were input into the network to predict jump height. The absolute difference between predicted and actual jump height (absolute error) was calculated to measure the accuracy of the network, and averaged over each round of cross validation. Cross validation was performed for fPCA thresholds from 75–100% by increasing the number of retained principal components. The entire process was repeated 25 times using different random initial weights in the network to achieve a repeatable measure of the expected accuracy.

A repeated measurement ANOVA (Bonferroni adjustment for multiple comparisons) was performed to examine the effect of the threshold on the absolute error of the network. The significance level was set at $\alpha = 0.05$. Data processing was performed in MATLAB and statistical analyses were performed using SPSS 20.

RESULTS

Visual inspection shows clear differences across the generated absolute errors (Figure 1). Thresholds smaller than 90% (up to 2 PCs) show the largest magnitude and spread in absolute errors, thresholds smaller than 99% (up to 5 PCs) and greater than 99.9% (more than 11 PCs) show moderate absolute errors and a wide spread of the absolute errors, while thresholds between 99% (from 6 PCs) to 99.9% (to 10 PCs) show the smallest magnitude and variation in absolute errors.

The statistical analysis found significantly lower ($p < 0.001$) absolute errors for the thresholds between 99% and 99.9% compared to other thresholds.

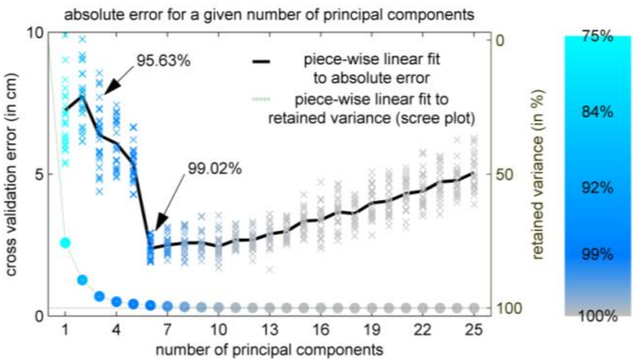

Figure 1: Absolute error (cross validated) of the used network in predicting jump height from principal component scores. Each point is the average accuracy from a complete run of cross-validation.

[3] Leave-one-out cross-validation uses one sample as test data and retains the other samples as training data. This process is repeated until each sample is used once as the test data.

DISCUSSION

The visual and statistical findings show that a fPCA threshold between 99-99.5% (optimal threshold) is most effective in describing jump height, generating significantly lower absolute error than other thresholds. Thresholds below the optimal threshold generated significantly higher absolute errors, indicating that the performed fPCA did not preserve enough information for the neural network to find the relationship between input data and target data accurately. Thresholds above the optimal threshold generated significantly higher absolute errors than the optimal threshold. Thresholds above the optimal threshold preserved unnecessary information (such as noise) that decreased the power of the input data to explain the target data. Further**,** visual inspection of the generated absolute error shows higher variation in absolute error below or above the optimal threshold, highlighting again either a lack of information or too much information retained.

The findings indicate that principal components beyond the frequently used threshold of 95% should be considered in experiments that use a fPCA threshold without performing cross validation. This is because principal components beyond the threshold of 95% can have a large influence on a dependent variable [e.g. 7]. Principal components beyond the threshold of 95% decreased the absolute error significantly and reduced the variation in absolute error in this experiment. However, principal components beyond a threshold of 99.9% significantly increase absolute error in this experiment and should be discarded.

In addition, the number of principal components suggested by the scree plot (4 PCs) differs from the optimal number identified by the network. The suggestion by the scree plot underestimates the number of principal component needed to describe jump height.

CONCLUSIONS

An optimal fPCA threshold to describe a dependent variable (jump height) accurately is within 99-99.9%. A scree plot should not be used for biomechanical purposes to choose the number of principal components, because principal components with a small influence on the data can have a large influence on a dependent variable.

ACKNOWLEDGEMENTS

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

REFERENCES

1. Donoghue OA, et al. *Medicine & Science in Sports & Exercise*, **40**:1323-1335, 2008.
2. Donà G, et al. *Sports Biomechanics* **8**:284-301, 2009
3. Richter et al. Proceedings of ISBS XXX, Melbourne, Australia, Proceeding 30, 2012.
4. Chau T. *Gait Posture,* **13**:49-66, 2001
5. Ramsay JO. *Functional data analysis*, Springer Verlag**,** New York, USA, 2006
6. Jolliffe IT. *Principal component analysis*, Springer Verlag**,** New York, USA, 2002
7. Harrison A, et al. *Sports Biomechanics*, **6**:199-214, 2007.
8. Mantovani et al., *Portuguese Journal of Sport Sciences*, **11**:911-914, 2011.
9. Hornik K, et al. *Neural Networks* **2**:359-366, 1989
10. Friedman J, et al., *The elements of statistical learning*, Springer Verlag**,** New York, USA, 2002.