USING FUNCTIONAL DATA ANALYSIS TO INVESTIGATE SIGNIFICANT LOAD INCREMENT

Azmin Rambely,[1] Wan Rozita Wan Din[2] and Abdul Jemain[2]

School of Mathematical Sciences, Faculty of Science & Technology, Universiti Kebangsaan Malaysia, 46300 Bangi, Malaysia [1]
Centre of Foundation Studies and Applied Sciences, Universiti Pertahanan Nasional Malaysia, 57000 Kem Sg Besi, Kuala Lumpur, Malaysia [2]

The purpose of the study is to investigate the significant load increment using Functional Data Analysis (FDA) technique during walking with load carriage activity. Ten military personnel were chosen as subject and ground reaction force data are collected. FDA method mainly B-spline basis formation, smoothing, curves registration and permutation tests were applied to analyze the GRF data. FDA proves to be one of the best methods in analyzing biomechanics data and the study shows that a 10% BW load increment is significant to Malaysian military.

KEY WORDS: Smoothing, curve registration, permutation test, load carriage.

INTRODUCTION: Malaysian military personnel are smaller build up size with height of around 168 cm and weight around 70 kg. They without exception have to carry excessive loads as heavy as 50 kg or more during training and on battle ground. Excessive load can cause spine injury (Parkinson & Callaghan, 2009) or perhaps more serious problems when they are getting older.
FDA has been applied to various researches such gene classification (Song, Deng, Lee, & Kwon, 2008), plankton analysis (Ikeda, Dowd, & Martin, 2008), lactate curves (Newella et al., 2006), gait analysis (Sadeghi, Allard, Shafie, Mathieu, Sadeghi, Prince, & Ramsay, 2000) and jaw movement (Crane et al., 2010). Other methods like one-way ANOVA (Pal et al., 2009), one-way MANOVA (Birrell, Hooper, & Haslam, 2007), principal component analysis (Lee, Roan, & Smith, 2009) and many more have been used in analyzing load data but not FDA. Thus this paper aims to investigate the significant load increment of 10% of the body weight (BW) during load carriage activity using Functional Data Analysis (FDA) technique.

METHODS: Ten healthy male military soldiers free from any injuries, age 31 ±6.2 years old, weigh 71.6 ±10.4 kg with height of 166.3 ±5.9 cm were chosen as subjects. Consent was obtained from all participants prior to the experiment. The subject wore tight outfit and bare-footed while carrying a normal military backpack with different loads of 0% (control),10, 20, 30 and 40% of each individual participant body weight (BW). In each condition, the movements were recorded by using a Vicon 1.4 Motion Analysis System which comprises of seven infrared cameras. Thirty-nine reflected markers were placed on the bony landmarks of the subject's body. The subject walked at his comfortable speeds few times before the movements were recorded. They were required to walk on a 6 m track at their own comfortable speed where Kirstler Force Platforms Type 9281C was embedded on the floor to detect the Ground Reaction Forces (GRF). The GRF data taken were normal walking without load (control) and with load of 10%, 20%, 30% and 40% of body weight (BW).

Figure 1: A subject with body markers attached carrying load of 0%, 10%, 20%, 30% and 40% BW.

Data Analysis: FDA is a relatively new method for analyzing data. Its main advantage lies in its ability to transform data to functional forms for further analysis. Besides normal statistical verification of data, FDA also enables researchers to use information from derivatives of the data. Other important features available in FDA are the ability to smooth and interpolate the data, to remove phase lag and to align curves according to specified peak and trough which are not possible by other methods. Statistical data analysis normally starts by finding an appropriate distribution to represent the data being investigated. Instead, FDA starts by converting data to functions and uses sets of basis functions, together with their coefficients, to construct functional data objects. B-spline basis expansion is selected to represent the curve in the form of functional data object which takes the form of,

$$x(t) = \sum_{k=1}^{K} c_k \phi_k(t)$$

where c_k is the corresponding coefficient and $\phi_k(t)$ is a set of B-spline basis function (Ramsay & Silverman, 2005).

The polynomials segments of basis function were jointed at specified knots or break points which were strictly increasing sequence and smoothed. Degrees of freedom for these particular curves were the number of interior knots excluding lower and upper limits plus the order of polynomials.

Smoothing parameters chosen were roughness penalty and penalizing fourth derivatives, with lambda 1e-12 to be applied to estimated functional parameters.

Next step was landmark or curve registration whereby all the curves were aligned according to their peaks and troughs which vary in amplitude. The process started by finding the precise points for peaks and valleys for all the curves where the registration was to be done. The function used timings of these features to estimate a nonlinear transformation of the argument continuum for each functional observation called warping function (Ramsay, Hooker, & Graves, 2009). For warping and registration purposes, the B-spline functions constructed were polynomials of ten basis functions with order six. Registration was the most important process in the analysis sequence since it could align all the prominent features of the curves which previously are not possible through other methods beside FDA. Only by applying this process, curves of different loads could be distinguished easily.

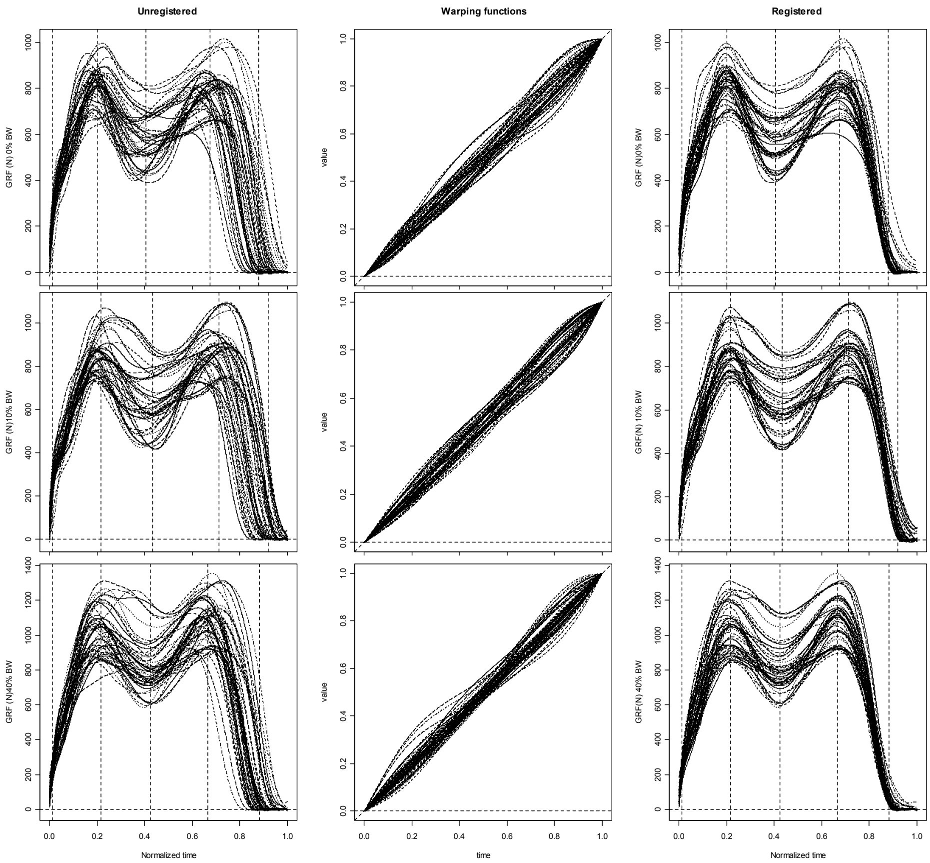
Permutation test was done to test the significance of load increment. It is usually as powerful as or more powerful than alternative approaches (Good, 1994) like bootstrapping for instance. Permutation tests were carried out on two different purposes. Firstly, to test the significance of increasing the load every 10% BW and secondly to test the effect of curve registration in the overall process flow. Therefore, it was done twice, after smoothing and again after curve registration. The test done was the *t*-statistics calculated point-wise and the test was based on the maximal value with default of 101 equally-spaced points. The *t*-statistics equation is

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}},$$

where $\bar{X}$ is the sample mean, S is the standard deviation and n is the sample size for every load calculated individually. Two adjacent groups of loads were compared i.e. 0% vs. 10% BW, 10% vs. 20% BW, 20% vs. 30% BW and 30% vs. 40% BW. The programming of statistical analyses was done in R.

RESULTS AND DISCUSSION: Walking gait cycle time period varies for different loads. Time normalization should be able to standardize the gait cycle to a common value hence reducing phase shifting at the end of the cycle. However, it is observed that time normalization is not sufficient to compare the result. By using FDA registration and warping function approach, alignment of the curves is done at three main points where maximum and minimum are located.

Registration is applied to shift every curve such that every one of them starts at the same moment resulted in zero phase lag. Figure 2 depicts the result of registration and warping functions. First column is the smoothed data without time normalization and registration, while the second column is the respective warping functions to be used for registration purposes. The third column is the result of the data after smoothing, time normalization and registration. By comparing the first and third columns, registration does tremendously improve the data whereby phase shifting has been eliminated resulting in only amplitude variation for comparison. The highest and lowest peaks and valleys are clearly visible. In order to verify the difference in amplitude variation, mean for every load is calculated.



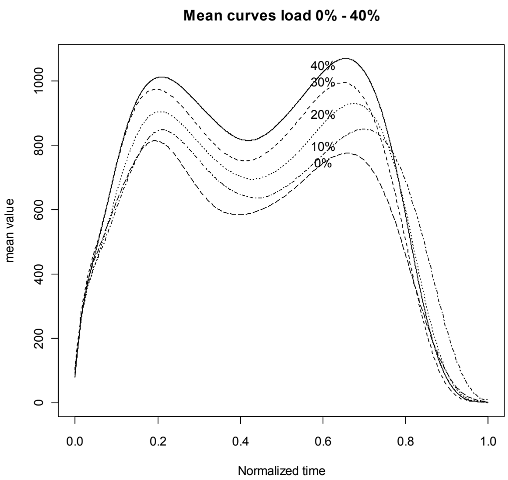


Figure 3 Mean GRF curves after time normalization and registration for each load condition.

Figure 2 Functional GRF curves for unregistered data (1st col.), warping functions (2nd col.) and registered data (3rd col.) for loads 0%, 10%, and 40% BW, respectively.

Figure 3 displays the mean GRF for the five different load conditions. It is observed that there exists a certain amount of differences in the magnitude of GRF for every load increment. Thus a permutation test was performed to evaluate the significant of 10% BW load increment to Malaysian military. The test was done by comparing each of the 10% difference in BW interval (0% to 10% BW, 10% to 20% BW etc). Results show that each 10% BW load increment are significant to the Malaysian subjects at 95% significant level at all maximum points of the curves where heel touches the platform and toe-off. P-values for tests are less than 0.0001.

CONCLUSION: FDA has an ability to convert data into functional forms. In this case, GRF data of different load carriage condition was analyzed. The process starts by transforming the data to functional data object. Then the data was smoothed to reduce noise in the system. Next curve registration was applied, which results in distinguished curve features for significant different detection. Finally, permutation test is done to evaluate the significance of increasing a 10% BW load while walking. Thus, FDA proves to be one of the best methods in analyzing biomechanics data and the study shows that a 10% BW load increment is significant to Malaysian military.

REFERENCES:

Birrell, S.A., Hooper, R.H. & Haslam, R.A. (2007). The effect of military load carriage on ground reaction force. *Gait & Posture,* 26, 611-614.

Crane, E.A, Cassidy, R.B., Rothman, E.D., & Gerstner, G.E. (2010). Effect of registration on cyclical kinematic data. *Journal of Biomechanics,* 43, 2444–2447.

Good, P. (1994). Permutation tests: a practical guide to resampling methods for testing hypotheses, *Springer Series in Statistics*. Springer-Verlag, New York.

Ikeda, T., Dowd, M. & Martin, J. L. (2008). Application of functional data analysis to investigate seasonal progression with interannual variability in plankton abundance in the Bay of Fundy, Canada. *Estuarine, Coastal and Shelf Science,* 78, 445-455

Lee, M.H., Roan, M. & Smith, B. (2009). An application of principal component analysis for lower body kinematics between loaded and unloaded walking. *Journal of Biomechanics* 42, 2226–2230.

Newella, J., McMillan, K., Grant, S., & McCabe, G. (2006). Using functional data analysis to summarize and interpret lactate curves. *Computers in Biology and Medicine*, 36, 262-275.

Pal, M.S., Majumdar, D., Bhattacharyya, M., Kumar, R., & Majumdar, D. 2009. Optimum load for carriage by soldiers at two walking speeds on level ground. *International Journal of Industrial Ergonomics,* 39, 68-72.

Parkinson, R.J. & Callaghan, J.P. (2009). The role of dynamic flexion in spine injury is altered by increasing dynamic load magnitude. *Clinical Biomechanics,* 24(2), 148-154.

Ramsay, J.O., Hooker, G., & Graves, S. (2009). *Functional Data Analysis with R and MATLAB*, New York, Springer. Chap. 8, pp. 117-124.

Ramsay, J.O. & Silverman, B.W. (2005). *Functional Data Analysis*, 2nd Edition. Springer, New York.

Sadeghi, H., Allard, P., Shafie, K., Mathieu, P.A., Sadeghi, S., Prince, F., & Ramsay, J. 2000. Reduction of gait data variability using curve registration. *Gait and Posture* 12, 257–264

Song, J.J., Deng, W., Lee, H.J., & Kwon, D. (2008). Optimal classification time-course gene expression data using functional data analysis. *Computational Biology and Chemistry,* 32(6), 426-432.

Acknowledgement: This work was supported by grants UKM-GUP-2011-154 and UKM-DLP-2011-049 from the Universiti Kebangsaan Malaysia. We also would like to thank PALAPES UKM for providing the military personnel as the subjects for the experiment.