THE DETERMINATION OF BREAK POINTS IN TIME SERIES DATA

Anthony J. Gorman, Alexander P. Willmott and David R. Mullineaux
School of Sport and Exercise Science, University of Lincoln, Lincoln, UK

The aim of this paper was to demonstrate three possible methods for the quantification of breakpoints. Data were the elbow angle during the pull phase of 10 strokes of ergometer rowing. First, a piecewise linear regression of two segments separated by a breakpoint was applied iteratively whereby the minimum pooled squared error of both regressions was the breakpoint. Second, the last local maxima, identified through zero-crossings of the first derivative, was the breakpoint. Thirdly, the last time point at which the rate of change of the joint angle data crossed a threshold was identified as the breakpoint. Each of these methods was checked against visual determination of the breakpoints. The threshold method was closer to the breakpoint as identified by visual inspection than were the linear regression and local maxima methods, and appeared suitable for this application.

KEY WORDS: breakpoint, linear, maxima, threshold, regression, statistics.

INTRODUCTION: The collection of time series data is common in biomechanics. These data are rarely rectilinear and often include changes along the y-axis. When these changes are abrupt, these are often termed breakpoints (BPs) (Stasinopoulos & Rigby, 1992).
The use of visual inspection of kinematic data for determining key movement events has previous been validated (Mickelborough et al., 2000), yet while it is easy to identify significant fluctuations and BPs in the data visually, there is a need for quantitative methods so as to circumvent possible subjectivity, and for the creation of algorithms to detect BPs in a given time series for large data sets.
Time series data are usually analysed to detect local minima or maxima to be quantified as change points. Other methods include the detection of points at which the sign of the curvature changes. These points of inflexion however, are less commonly explored. In other areas including physiology, techniques have been used for quantifying BPs in time series data through regression modelling (Muggeo, 2003; Cannon, Kolhorst, & Buono, 2009). Such techniques remain rarely applied in biomechanics, yet they could provide easy quantification of the fluctuations in time series data. This could provide a useful tool for numerous biomechanical applications, for example, in the detection of the onset of joint motion. The purpose of this paper was to evaluate the effectiveness in detecting the initiation of a joint angle change of three algorithms: (1) regression modelling; (2) detection of the local maxima from the derivatives of the signal values; and (3) through changes in the gradient of the first derivative of the signal values in relation to a threshold.

METHODS: Test data were obtained from 24 females (mean ± SD: age, 19.1 ± 1.6 years; height, 1.69 ± 0.09 m; mass, 67.5 ± 5.9 kg), who each provided informed consent. Participants rowed continuously for 10 minutes on a Dynamic ergometer (Concept2, Morrisville, VT). Kinematic data were obtained from five passive, spherical, retro-reflective markers, of 9.5mm diameter affixed to anatomical landmarks of the shoulder, elbow, and wrist joints, and to the lateral sides of the upper arm and forearm. A further seven markers were placed on the ergometer, on the handle and the foot stretcher. The positive X-axis pointed along the length of the ergometer, towards the feet of the subject, the Z-axis was vertically up and the Y-axis was the cross-product of Z and X (pointing left).
Three-dimensional kinematics of the markers were recorded at a rate of 150Hz using 11 Motion Analysis Corporation (MAC) Raptor Digital Cameras (MAC, Santa Rosa, CA). All marker identification was completed using Cortex v5.3.1 (MAC) and data analysed using custom written MATLAB code (R2014b; MathWorks, Natick, MA). Data were smoothed using a zero lag 4th order Butterworth low-pass filter with a cut-off frequency of 7Hz.
Ergometer and joint kinematics were analysed in three dimensions, and the elbow angle was defined as the angle between the longitudinal axes of the upper arm and forearm (where 180° was full extension). The turn-points of the rowing motion were defined as the instants at which

the velocity of the centre of the ergometer handle in the X-axis changed from positive to negative (catch), and from negative to positive (finish). Data from catch to subsequent finish constituted one ‘pull’ phase. Over each pull, joint position and velocity data were normalised as a function of time to 101 data points using cubic spline interpolation. The pull phases of the first 10 rowing strokes of each participant were taken for analysis.

The BP of the initiation of elbow motion from greater extension to greater flexion was firstly determined through visual inspection (VI) of the plotted joint position data by the researcher. This method used as a comparison for the subsequent three algorithms, which were each applied separately to the data. The first algorithm was a piecewise linear regression (LR), with two segments, separated by a BP (Figure 1). Using an iterative approach, each joint angle data point was in turn treated as a potential BP, and least-squared linear regressions were fitted to the data on either side. One regression was calculated from the first data point forwards to the BP, the second from the end data point back to the BP. The latter data were then reversed along the time axis so that the regression ran from the BP to the end of the data. The BP at which the pooled sum of the squared errors of both regressions was least was taken as representative of the time of the initiation of elbow flexion. For the second algorithm, termed the local maxima (LM) method, joint angle data were differentiated with respect to time before each positive to negative change in the sign of the first derivate was detected. Each of these changes were categorised as a local maxima. The last maxima was termed the BP, and here represented the time instant of the first uninterrupted move into flexion of the elbow. A third algorithm, termed the threshold method (TH) was also applied to the data. First, the time point at which the joint angle data crossed below a threshold of 100° was determined. Working from this point backwards towards the start of the data, the gradient of the first derivative of each joint angle data point was assessed against a second threshold. The first time instant at which this gradient became less negative than -0.25°/%pull, indicated the start of joint flexion and was taken as the BP of the joint angle data. Data were presented as means and standard deviations, and values were compared using paired t-tests at an alpha level of 0.05 using SPSS (v.22, IBM, Armonk, NY).

RESULTS: The mean estimated times of the elbow angle BPs (as a percentage of the pull phase) across all strokes of all participants were 39 ± 6%, 54 ± 7%, 27 ± 4%, and 37 ± 5% for VI, the LR, LM, and TH methods, respectively. For each participant, the LR method consistently estimated BPs of elbow flexion that were significantly later in time than the other methods (TH, $p < 0.05$; LM, $p < 0.05$; VI, $p < 0.05$). The LR method appeared to estimate BPs that were later than the actual onset of elbow flexion (Figure 1.b), identifying BPs that were during the transition of the elbow moving into flexion, rather than at its initiation. The estimates derived from the LM method were closer in time to the VI BPs, yet were significantly different ($p < 0.05$). The TH method however, appeared to determine BPs at the initiation of the change of the joint angle (Figure 1.). The TH method showed no significant difference when compared to VI ($p < 0.05$), and was therefore considered comparable. Figure 1 illustrates an example of the three calculated BPs of the elbow angle, displaying the time intersection of the two regressions and the points of the last local maxima and threshold crossing for the LR, LM, and TH methods, respectively.

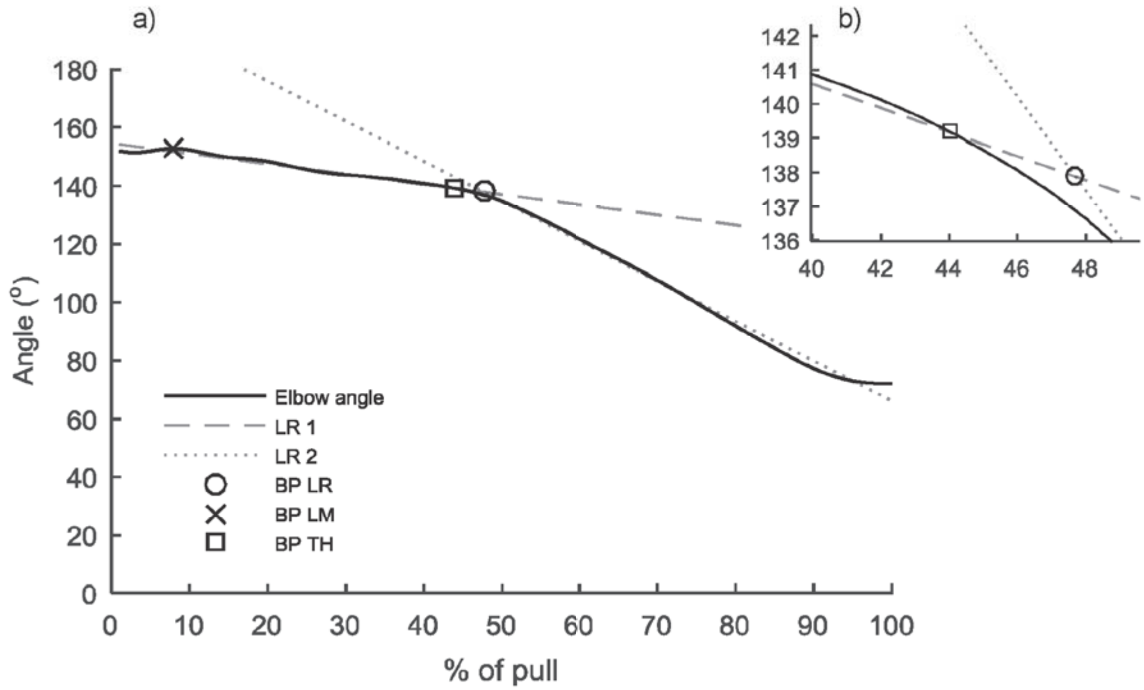

Figure 1. a) Mean (n = 10 strokes) elbow angle data for 1 participant. LR1 and LR2 represent both linear regressions fitted for the LR method; BP LR, BP LM, and BP TH, are the breakpoints as determined by the linear regression, local maxima, and threshold methods, respectively. b) Area of interest around the estimated BP LR and BP TH enlarged from Figure 1a).

DISCUSSION: A number of considerations became apparent when assessing the applicability of these methods. The sensitivity of the LR method was affected by the rate of change from extension to the onset of flexion. The slower the transition, the larger the radius of the curve in the data and the further in time the estimated BP was from the initiation of the curve away from the first regression (Figure 1). Furthermore, the gradient of the two regressions fitted to the data were dependent upon the extremes of the elbow angle data; for example, a return into extension of the elbow at the end of the pull altered the gradient of the second regression, consequently altering the intersection of the two regressions and the estimated BP. Running the LR method over a selected sample of the data set, around a prior estimation of the BP location, would have alleviated this, yet the usability of the LR method would reduce if the data being assessed displayed large variability in the predicted location of the BP and if, for each trial, individual data samples had to be identified before the LR analysis. This could be further complicated by the number of data points included in the calculation of each regression, so as to reduce the effects of prior or subsequent data fluctuations affecting the gradients of the regression equations.

If elbow extension was observed at the start of the pull phase and, after the initial move into flexion, the elbow continued to move into flexion at an increasing rate, the LM method appeared to identify the moment that the flexion began. However, if at the start of the pull elbow flexion were observable as a reaction to overcoming the inertial force of accelerating the ergometer handle, the likelihood of false positive identification of the BP would increase, especially if the elbow did not return into extension before the true BP (e.g. Figure 1a). The LM method is reliant on a change in the sign of the angular velocity to detect a BP, whereas in some cases observed here, the BP is actually a marked change from slow to rapid in the rate of flexion, rather than a move from extension to flexion. Such false positives, or even the non-detection of a BP, would need to be reduced. To decrease these, an extension of the LM method could comprise the inclusion of a threshold across which the data has to pass before the last LM is taken (Cannon et al., 2009).

The TH method produced the closest BPs to VI, however, the determination of the limits of the thresholds would need careful prior consideration. The thresholds chosen here were determined through pilot inspection of the data and were aided by the predictability of the movement being assessed. To provide a more quantitative approach to determining noise in the data for the setting of the thresholds, the TH method would benefit from a more objective, systematic method for estimating threshold values from a subset of the data prior to the expected BP location.

Neither without limitations, nor without the need for adaptation based on the time series data itself, the methods presented here demonstrate steps in quantifying the changes in times series data. The TH method appears a simple and suitable algorithm for detection of the initiation of joint motion of the type assessed here, where there are clear changes in directionality of the time series data and only one BP is present. However, this method would require refinement if it is to be applicable to data displaying greater temporal fluctuations.

CONCLUSION: This study describes the effectiveness of three methods for determining the BPs of time series data. On application to this test data and the determination of the onset of elbow flexion, the TH method was found to be effective for the detection of the BP of elbow joint motion, displaying greater accuracy than both the LR and LM methods when compared to VI. This study presents quantitative methods to assess changes in time series data and future research should focus on the refinement of these techniques.

REFERENCES:

Cannon, D.T., Kolhorst, F.W., & Buono, M.J. (2009). On the determination of ventilatory threshold and respiratory compensation point via respiratory frequency. *International Journal of Sports Medicine*, *30*(3), 157-162.

Mickelborough, J., van der Linden, M. L., Richards, J., & Ennos, A. R. (2000). Validity and reliability of a kinematic protocol for determining foot contact events. *Gait and Posture, 11,* 32-37.

Muggeo, V.M.R. (2003). Estimating regression models with unknown break-points. *Statistics in Medicine*, *22*(2), 3055-3071.

Stasinopoulos, D.M., & Rigby, R.A. (1992). Detecting break points in generalised linear models. *Computational Statistics and Data Analysis*, *13*, 461-471.