USING A BREAKPOINT TO DETERMINE THE OPTIMAL CUT-OFF FREQUENCY

David R. Mullineaux

School of Sport and Exercise Science, University of Lincoln, Lincoln, United Kingdom

The aims of this study were to compare methods of determining the optimal cut-off frequency (CFopt) for a Butterworth filter. CFopt were determined for leg displacement data for treadmill running through residual analysis using regression (RA0reg), integral of the power spectral density (PSD), and both these methods analysed through a new ‘breakpoint’ method. RA0reg did not correlate with other methods suggesting poor concurrent validity. The ‘breakpoint’ method correlated significantly between several methods. CFopt was least for anterio-posterior and highest for vertical directions for all methods ($p<0.05$). Settings for RA0reg and PSD can have substantial effects on CFopt, but the ‘breakpoint’ is not affected as much by the settings. Future research should attempt to standardise settings and explore the criterion validity of the methods to determine CFopt.

KEY WORDS: Butterworth, filtering, residual, smoothing.

INTRODUCTION: The Butterworth filter (Butterworth 1930) is one of the most commonly used filters for smoothing data in biomechanics. A principal setting is to define a cut-off frequency (CF) that is ‘optimal’ (CFopt) for retaining the majority of the movement signal and removing the majority of the noise. There have been several analytical approaches to identifying the CFopt including residual analysis of first derivative data (Jackson, 1979), residual analysis of displacement data using regression (RA0reg; Wells & Winter 1980) and a fixed integral of the power spectral density (PSD; Kram et al., 1998).

The approaches to determining the CFopt are all influenced by choices with the data capture and reduction (e.g. sampling rate; analysis of whole time-series or of separate phases). In addition, choices with the analytical settings influence the CF. For example, for RA0reg the CFopt is determined by the portion of residuals selected for the regression, and for PSD the CFopt corresponds to the size of the fixed integral of the PSD (e.g. 99%; Kram et al., 1998).

The choices for the settings for the analytical approaches all have substantial effects on the determination of the CFopt. Further research is required to confirm the appropriateness of these settings, and to attempt to standardise these settings. However, an approach that is more robust that is independent of settings is required to provide a more stable result. These methods all produce a curve similar to a parabola, where the ‘breakpoint’ is the vertex. Hence, it is proposed CFopt corresponds to the ‘breakpoint’ defined as the minimum of the curve once rotated so that its start and end points are zero. Consequently, the aim is to compare the CFopt determined using previous methods (99% PSA; RA0reg) and adapting these methods using the ‘breakpoint’.

METHODS: The test data were from a previously published study (Mullineaux, 2017) for running on a non-motorized curved-treadmill (Curve, Woodway). Healthy males (n=10; 21.4±4.1years; 75.6±7.0kg; 1.81±0.08m) ran for 30s at 2.15±0.11m/s, calculated from the left-leg posterior toe marker’s displacement and stance duration. Twenty retroreflective markers (greater-trochanter; lateral and medial femoral-epicondyle and malleolus; mid thigh and shank; heel; dorsal first metatarsal-phalange and inter-phalange joints) were captured at 200Hz (9 Raptor cameras; Cortex5.2 software; Motion Analysis Corporation). Data were analysed in Matlab (2015a; Mathworks) to determine the CFopt for a fourth-order critically damped recursive Butterworth filter. Five methods of determining CFopt were calculated as described below.

Residual analysis of zero derivative by regression (RA0reg; Wells & Winter 1980; Figure 1). This first involved calculating multiple sets of filtered coordinate data using filter CF ranging from 0.5 Hz to 25 Hz (e) in 0.1 Hz increments. The RMSD (residual) between each set of filtered and raw coordinate data were then calculated and the residual v CF plotted. From the approximate maximum feasible frequency in the signal (d, which was set to 12.5 Hz) to the maximum CF used (e), a linear regression was fitted to determine the intercept (a). The CF corresponding to where the residual equalled or exceeded 'a' (b) was taken as CFopt.

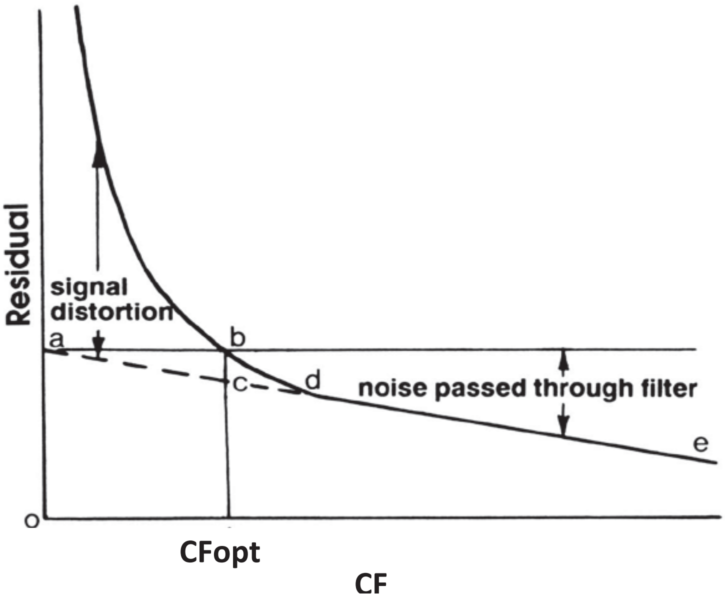

Figure 1. Determining the optimal cut-off frequency (CFopt) using the residual analysis of zero derivative by regression method (adapted from Winter, 1990). See text for explanation.

Residual analysis of zero derivative by breakpoint (RA0bkpt). This involved calculating the same residual v CF data as for RA0reg. The data were then offset normalised by subtracting the maximum value. The data then underwent a 'curve-rotation' so that the residual at the maximum CF equalled zero. The minimum value of the rotated curve was defined as the breakpoint and the corresponding CF was taken as CFopt.

PSD integral of 99% (PSD99). This first involved using a DC-offset (i.e. subtract mean), fourth order polynomial detrending to remove very-low frequency errors, and a Hanning window with correction to taper the end of the signals to zero. Next a Fast Fourier Transformation was calculated, and the CF corresponding to the 99% integral of the PSD was taken as CFopt.

PSD breakpoint (PSDbkpt). By inverting the PSD, and applying the 'curve-rotation', the breakpoint was determined as the CFopt.

Amplitude spectral density breakpoint (ASDbkpt). By calculating the amplitude of the PSD, inverting, and applying the 'curve-rotation', the breakpoint was determined as CFopt.

Across the 30 second data recording, the CFopt for the 5 methods were calculated for each Cartesian component for each marker of each subject separately (i.e. 5×3×20×10). Any CFopt greater than 15 Hz was treated as an outlier and deleted. Between each pair of the same CFopt data, Pearson product moment correlations (r) were calculated, and the average number of significant correlations determined ($r>0.707$ for df=10-2; maximum of 10 as n=10). For PSDbkpt and ASDbkpt the corresponding integrals of the PSD were calculated as PSDbkpt% and ASDbkpt%, respectively. For each Cartesian component of all markers, the minimum and maximum values for the 5 CFopt, PSDbkpt% and ASDbkpt% values were calculated. In addition, for each Cartesian component of all markers combined, the mean and SD were calculated, and the pairwise comparisons of xy, xz and yz were calculated using paired t-tests (where x is anterio-posterior; y is medio-lateral; z is vertical). The statistical significance level was 0.05.

RESULTS: The descriptive statistics of the CFopt for the various methods are described in Table 1. The highest CFopt was found for RA0reg, with a narrow range across all Cartesian coordinates and markers of 8.8 to 11.4 Hz (mean ± SD; 10.4 ± 0.4 Hz for x). The lowest CFopt were found for PSDbkpt, ranging from 1.3 to 7.0 Hz (2.5 ± 0.5 Hz for x). For each method, the majority of pairwise comparisons of CFopt were statistically significantly different between Cartesian coordinates (p<0.05).

Table 1
Descriptive statistics of 5 optimal cut-off frequency methods and integrals under the power spectral density for 2 of the 5 methods presented for each Cartesian coordinate

		RA0reg	RA0bkpt	PSD99	PSDbkpt	ASDbkpt	PSDbkpt%	ASDbkpt%
Mean	x	10.4	3.5	2.8	2.5	4.7	99.3	99.9
	y	10.2	4.0	4.5	3.3	6.0	98.6	99.7
	z	10.6	4.8	3.8	3.7	5.9	99.4	99.8
SD	x	0.4	0.4	0.5	0.5	0.7	0.3	0.0
	y	0.3	0.5	0.9	0.4	0.8	0.4	0.1
	z	0.2	0.8	0.8	0.7	0.8	0.3	0.1
Min	x	9.0	2.9	1.3	1.3	2.5	98.0	99.5
	y	8.8	2.7	2.2	1.4	3.7	96.2	98.5
	z	9.4	3.3	2.5	2.5	3.8	96.7	97.8
Max	x	11.4	5.1	4.5	4.2	6.8	99.9	100.0
	y	11.0	6.5	8.0	7.0	9.5	99.9	100.0
	z	11.3	6.0	6.6	5.5	9.1	99.9	100.0
Sig	xy		Y	Y	Y	Y	Y	Y
	xz	Y	Y	Y	Y	Y		Y
	yz	Y	Y	Y	Y		Y	Y

Note: units of Hz; Y for Sig indicates significance differences between the pairwise comparisons of Cartesian coordinates (p<0.05), and; for PSD99, 3 outliers >15Hz were deleted.

The average number of significant correlations between methods was greatest between each of the breakpoint methods (Table 2). The most commonly used RA0reg method had few significant correlations with any other method. There were few significant correlations with the areas under the PSD (≤1.7 out of 10), although the suitability of a Pearson product moment correlation is limited as the range was small at only 96.2 to 100% (see Table 1).

Table 2
Average number of significant correlations between 5 optimal cut-off frequency methods and for 2 of these methods the integrals under the power spectral density

	RA0reg	RA0bkpt	PSD99	PSDbkpt	ASDbkpt	PSDbkpt %	ASDbkpt%
RA0reg		0.3	0.0	0.0	1.3	0.0	0.3
RA0bkpt	0.3		**5.0**	**6.0**	0.0	0.3	1.7
PSD99	0.0	**5.0**		**4.3**	**3.3**	1.3	0.7
PSDbkpt	0.0	**6.0**	**4.3**		0.3	0.3	0.7
ASDbkpt	1.3	0.0	**3.3**	0.3		1.7	0.7
PSDbkpt%	0.0	0.3	1.3	0.3	1.7		0.3
ASDbkpt%	0.3	1.7	0.7	0.7	0.7	0.3	

Note: highest counts in bold (maximum of 10).

DISCUSSION: For RA0reg, which is the most commonly used method to determine CFopt, the settings used resulted in CFopt being higher than many researchers have used in past. Although the data is not presented, changing these settings substantially altered CFopt. RA0reg was first used with manually digitised data (Wells & Winter, 1980), and with the greater sampling frequencies and accuracy of automated motion capture system being more frequently used today the settings required needs further exploration and standardisation. RA0reg had poor correlations with the other methods, indicating low concurrent validity.

In using a fixed 99% area under the PSD resulted in a lower CFopt (2.8 ± 0.5 Hz for x) compared to RA0reg. Proposed areas range from 99% (e.g. Kram et al., 1998) to 95% (e.g. Sinclair et al., 2013). In this study the area ranged from 96.2 to 100%, hence a fixed area of 95% would have resulted in much lower CFopt. This suggests a fixed area is not appropriate to account for the varying frequency contents across coordinates, markers or different studies, and supports the recommendation that different CFopt should be used depending on the variables of interest (Giakas & Baltzopoulos, 1997).

A new method to determine CFopt based on the breakpoint of the curves was proposed. The main advantage of this approach is that no settings are required, although the inputted data (e.g. whole recording; 200 Hz sampling frequency) may affect the results that has not been explored. This breakpoint approach was adaptable, and was applied to both RA0 and PSD. It was found that there were several significant correlations particularly between RA0bkpt, ASDbkpt and PSDbkpt, supporting concurrent validity. ASDbkpt produced the highest CFopt with means of 4.7, 6.0 and 5.9 Hz for x, y and z, respectively (Table 1), which are closest to CFopt often used in the literature. It is probable that a different CFopt should be used for each marker, but this has not been presented.

In this study, participant data were used, which has the benefit of retaining the complexity of the signal. However, these make it is more difficult to determine the validity of the methods, hence mathematically-generated signals with known noise may also assist in the future for exploring the criterion validity of the methods.

CONCLUSION: All methods used for calculating CFopt indicated each Cartesian coordinate requires a differing CFopt, which for this non-motorised treadmill running data would be lowest for the anterio-posterior and highest for the vertical directions. The varying settings of the methods can have substantial effects on CFopt, hence standardisation of these is required. To achieve this, further research is required to determine the criterion validity of these or other methods for determining CFopt using both participant and mathematically generated signals.

REFERENCES: Butterworth, S. (1930). On the theory of filter amplifiers. *Experimental Wireless and the Radio Engineer*, 7, 536-541.

Giakas, G. & Baltzopoulos, V. (1997). Optimal digital filtering requires a different cut-off frequency strategy for the determination of the higher derivatives. *Journal of Biomechanics*, 30(8), 851-855.

Jackson, K.M. (1979). Fitting of mathematical functions to biomechanical data. *IEEE Transactions on Biomedical Engineering*, 26(2), 122-124.

Kram, R., Griffin, T.M., Donelan, J.M. & Chang, Y.H. (1998). Force treadmill for measuring vertical and horizontal ground reaction forces. *Journal of Applied Physiology*, 85(2), 764-769.

Mullineaux, D.R. (2017). CI2 for creating and comparing confidence-intervals for time-series bivariate plots. *Gait and Posture*, 52, 367-373.

Sinclair, J., Richards, J., Taylor, P.J., Edmundson, C.J., Brooks, D. & Hobbs, S.J. (2013). Three-dimensional kinematic comparison of treadmill and overground running. *Sports Biomechanics*, 12, 272-82.

Wells, R.P. & Winter, D.A. (1980). Assessment of signal and noise in the kinematics of normal, pathological, and sporting gaits. In *Proceedings of the Special Conference of the Canadian Society for Biomechanics* (pp 92-93). London, Ontario: CSB.

Winter, D.A. (1990). *Biomechanics and Motor Control of Human Movement*, Hoboken, NJ: Wiley.