A SURROGATE METHOD FOR DISCRETE MOVEMENT DATA

Paul Taylor[1], Michael Small[2], Kwee-Yum Lee[1], Raul Landeo[1], Damien O'Meara[1, 3] and Emma Millett[1, 3]

School of Exercise Science, Australian Catholic University, Sydney, NSW, Australia[1]
School of Mathematics and Statistics, The University of Western Australia, Perth, WA, Australia[2]
New South Wales Institute of Sport, Sydney, NSW, Australia[3]

Sample entropy can be an effective tool for the investigation of human movement variability. However, before applying the method, it can be beneficial to employ an analysis to confirm that observed data is not solely the result of stochastic processes. This can be achieved using surrogate methods. Previous investigations have used surrogate methods within human gait data, yet no appropriate method has been applied to discrete human movement. This article proposes a surrogate method for discrete movement data. The technique reliably generated surrogates for discrete joint angle time series, effectively destroying fine-scale dynamics of the observed signal and maintaining macro structural characteristics (e.g., Mean, SD). Comparison of entropy estimates indicated that observed signals contained deterministic dynamics.

KEY WORDS: deterministic dynamics, entropy, pseudo periodic surrogate, movement variability

INTRODUCTION: A tool which has been shown to be effective in investigating movement variability is sample entropy (Preatoni, Ferrario, Dona, Hamill, & Rodano, 2010). Sample entropy allows quantification of the regularity of a signal allowing inference to the complexity of the organism or system producing the signal (Lake, Richman, Griffin, & Moorman, 2002; Preatoni et al., 2010; Richman & Moorman, 2000). However, as entropy quantifies the regularity of signals that are stochastic, deterministic or a combination of both, a method which can demonstrate that a biological signal is not solely stochastic in nature is beneficial. This outcome can be achieved by contrasting observed data with data generated from surrogate methods (Small, Nakamura, & Luo, 2007; Theiler, Eubank, Longtin, Galdrikian, & Doyne Farmer, 1992). Surrogate methods can produce time series which resemble observed data yet present properties consistent with a non-deterministic signal.
Various surrogate techniques exist for different applications (Small et al., 2007). Due to its cyclical nature, human gait has previously been investigated using a pseudo periodic surrogate method (Miller, Stergiou, & Kurz, 2006; Preatoni et al., 2010). However, this method is inappropriate for discrete movements as the discontinuities that exist from one trial to the next prohibit time delay embedding. The purpose of this article is to propose a surrogate method which can be applied to discrete movement data and to determine its reliability. Sample entropy will then be used to test for deterministic dynamics within discrete human movement using the generated surrogates.

METHODS: Sixty four shoulder (I/E rotation) and elbow (flex/ext) joint angular displacement time series were obtained from 10 informed, consenting, male participants [24.1 (3.3) years; 176.6 (5.9) cm; 76.4 (7.8) kg] during the performance of an overarm throwing task across four sessions. Equipment and laboratory configuration were the same as reported previously (Taylor, Lee, Landeo, O'Meara, & Millett, 2015).
A generalisation of the pseudo periodic surrogate method (Small, Yu, & Harrison, 2001) is proposed for discrete movement data. This method will result in outcomes similar to those achieved using a Small surrogate method (Nakamura & Small, 2005). The following outlines the proposed method:

1. Let x_{ij} and y_{ij} be the j^{th} scalar time point from the i^{th} trial of observed joint angle time series (e.g. where x_{ij} is elbow angular displacement and y_{ij} is the same for the shoulder). Let the concatenated time series *X* and *Y* be;

$$X = (x_{ij})_{\substack{i=1,\dots,N \\ j=1,\dots,T_i}}$$
$$Y = (y_{ij})_{\substack{i=1,\dots,N \\ j=1,\dots,T_i}}$$

where *N* is the total number of trials collected, T_i is the total number of data points in the i^{th} trial and *X* and *Y* are matrices with dimensions $\sum_{i=1}^{N} T_i$.

2. Then the concatenated time series *X* and *Y* are combined to form a phase space, *P,* where *P* is a matrix with dimensions $2 * \sum_{i=1}^{N} T_i$;

$$P = (X_{ij}, Y_{ij})_{\substack{i=1,\dots,N \\ j=1,\dots,T_i}}$$

3. Initial (*A*) and final (*B*) conditions of individual trials within *P* are extracted where *A* and *B* are both *2 x N* matrices;

$$A = (x_{i1}, y_{i1})_{i=1,\dots,N}$$
$$B = (x_{iT_i}, y_{iT_i})_{i=1,\dots,N}$$

4. Elements of *P* are then shuffled (randomly resampled with replacement) to form the surrogate P_s. First an initial current state $P_{s(i,t)}$ is selected at random from *A*. Set *t* = 1.

5. To select the next state of P_s first noise is added to the current state creating *C*;

$$C = P_{s(i,t)} + \rho g P_{s(i,t)}$$

where *ρ* is a constant and *g* is Gaussian noise;

$$g \sim N(0,1)$$

6. The state in *P* which is closest to the noisy current state *C* created above is identified as $k_{m,n}$ using the least root mean square difference between *C* and each column of the matrix *P*. Then the next state of P_s is defined as the successor;

$$P_{s(i,1+t)} = k_{m+1,n+1}$$

7. The state $P_{s(i,1+t)}$ is now the current state of P_s. Increment *t*. The next state of P_s is selected by repeating steps 5–6. The process of incrementing *t* and selecting the next state continues until the current state of P_s is equal to one of the sets in *B*.

8. The value *i* can then be incremented and steps 4–7 repeated to obtain the next surrogate.

An optimal value for *ρ* (Step 5) elicits the greatest number of small segments within the surrogate time series (Small et al., 2001), providing an optimal balance between effectively destroying the fine-scale dynamics of the signal and maintaining its macro structure. As *ρ* increases, so too will the number of small segments, towards a maximum, before returning toward zero (as $\rho \rightarrow \infty$). Ranges of *ρ* were tested in order to determine the optimal value for each block of throws analysed.

Elbow and Shoulder time series were concatenated and combined to form the two dimensional phase space from which the respective surrogates were drawn. The number of surrogates generated matched the number of throws in the observed data for each block. Surrogates with similar length (± 1SD) as the mean length in the original data were accepted to maintain comparability. If this criterion was not met, the surrogate was rejected and the process repeated.

To demonstrate the ability of the technique to produce surrogates which approximate the macro structure of the original data, surrogate Mean, SD and data length were compared to that of observed signals using Mann-Whitney U tests. Furthermore, these values were assessed for reliability using intraclass correlation and standardised typical error tests (Hopkins, 2000, 2011).

Sample entropy (Lake et al., 2002; Richman & Moorman, 2000) estimates of the observed and surrogate data were used for statistical inference. It was hypothesised that the observed time series would return lower sample entropy than surrogates as they are not solely the result of noisy, random processes, but contain some element of deterministic dynamics. The lower entropy score of the observed data would reflect the increased regularity of a signal under the control of the neuromuscular system as opposed to the random, stochastic process producing the surrogate. Sample entropy (*m* = 2 and *r* = 0.1) was estimated for the

concatenated real and surrogate time series of the three joint angles for all blocks of throws. These scores were compared using the Mann-Whitney U test. Non parametric statistics were employed as data did not display normality (Peat & Barton, 2005).

RESULTS: Surrogates were effectively generated for all observed throws. An example of elbow and shoulder time series and their respective surrogates can be seen in Figure 1.

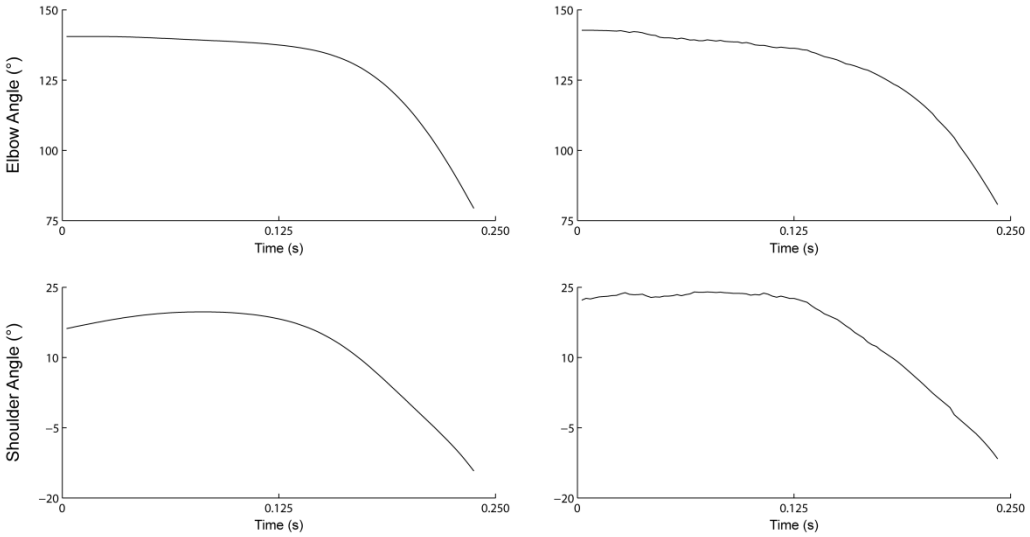

Figure 1: Example of elbow and shoulder data (left panels) and their respective surrogates (right panels)

Results of the Mann Whitney U tests indicated that observed joint angular displacement time series had significantly lower sample entropy ($p \leq 0.01$) than their respective surrogate (Figure 2).

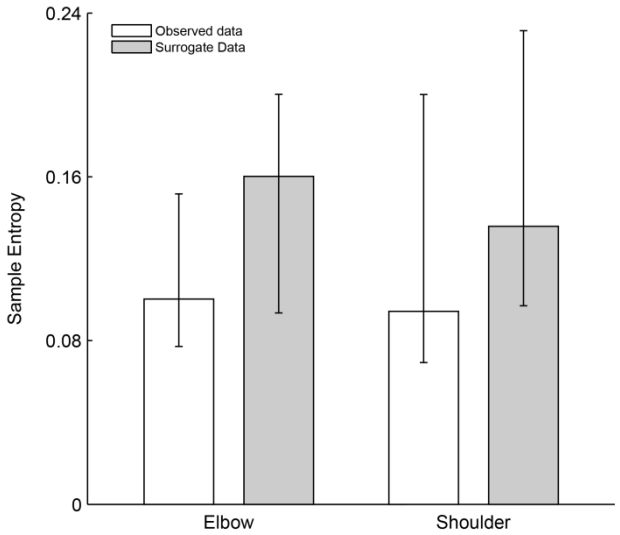

Figure 2: Difference between observed and surrogate entropy estimate (± IQR) for elbow and shoulder time series, significant at $p \leq 0.01$

The comparison of macro characteristics (mean, SD and length) showed no significant differences between the real and surrogate throws ($p \geq 0.68$). Reliability analysis indicated that the surrogate generation algorithm was able to consistently produce this output as indicated by an ICC ≥ 0.99 and a small standardised typical error (≤ 0.1) (Hopkins, 2000, 2011).

DISCUSSION: Comparison of the sample entropy score for both real and surrogate data (Figure 2) indicated that the observed discrete human movement is not solely the product of non-deterministic 'noisy' processes. Hence, the surrogate method effectively disrupted the micro structure of the signal while the macro characteristics of the observed data were maintained. Similar results have been shown in previous work using a pseudo periodic surrogate with normal gait and race walking (Miller, et al., 2006; Preatoni, et al., 2010). Coupled with the ability of the algorithm to consistently produce the expected outcome, the proposed surrogate method is both a valid and reliable technique to investigate the presence of deterministic dynamics in other discrete human movement time series.

CONCLUSION: Sample entropy can provide useful information about the complexity and organisation of the neuromuscular system. Being able to prove the existence of determinism in signals to be analysed using entropy measures lends validity to any observed outcomes which can be attributed to purposeful changes within the organism and not solely to stochastic/random processes. Together, sample entropy and the proposed surrogate method may provide a useful tool for the further exploration of movement variability within discrete sporting movements.

REFERENCES:

Hopkins, W. (2000). Measures of reliability in sports medicine and science. *Sports Medicine, 30*(1), 1-15. doi: 10.2165/00007256-200030010-00001

Hopkins, W. (2011). Precision of measurement. In W. Hopkins (Ed.), A New View of Statistics. Retrieved from newstats.org/precision.html.

Lake, D. E., Richman, J. S., Griffin, M. P., & Moorman, J. R. (2002). Sample entropy analysis of neonatal heart rate variability. *American Journal of Physiology - Regulatory, Integrative and Comparative Physiology, 283*(3), R789-R797. doi: 10.1152/ajpregu.00069.2002

Miller, D. J., Stergiou, N., & Kurz, M. J. (2006). An improved surrogate method for detecting the presence of chaos in gait. *Journal of Biomechanics, 39*(15), 2873-2876. doi: http://dx.doi.org/10.1016/j.jbiomech.2005.10.019

Nakamura, T., & Small, M. (2005). Small-shuffle surrogate data: Testing for dynamics in fluctuating data with trends. *Physical Review E, 72*(5), 056216.

Peat, J., & Barton, B. (2005). *A Guide to Data Analysis and Critical Appraisal* (1st ed.). Malden, MA: Blackwell Publishing.

Preatoni, E., Ferrario, M., Dona, G., Hamill, J., & Rodano, R. (2010). Motor variability in sports: A non-linear analysis of race walking. *Journal of Sports Sciences, 28*(12), 1327-1336. doi: 10.1080/02640414.2010.507250

Richman, J. S., & Moorman, J. R. (2000). Physiological time-series analysis using approximate entropy and sample entropy. *American Journal of Physiology-Heart and Circulatory Physiology, 278*(6), H2039-H2049.

Small, M., Nakamura, T., & Luo, X. (2007). Surrogate data methods for data that isn't linear noise. In C. W. Wang (Ed.), *Nonlinear Phenomena Research Perspectives* (pp. 55-81). New York, NY: Nova Science Publishers.

Small, M., Yu, D., & Harrison, R. G. (2001). Surrogate test for pseudoperiodic time series data. *Physical Review Letters, 87*(18), 188101.

Theiler, J., Eubank, S., Longtin, A., Galdrikian, B., & Doyne Farmer, J. (1992). Testing for nonlinearity in time series: the method of surrogate data. *Physica D: Nonlinear Phenomena, 58*(1–4), 77-94. doi: http://dx.doi.org/10.1016/0167-2789(92)90102-S