A Surrogate Technique for Investigating Deterministic Dynamics in Discrete Human Movement

Paul G Taylor[a], Michael Small[b], Kwee-Yum Lee[a], Raul Landeo[a], Damien M O'Meara[a, c] and Emma L Millett[a, c]

[a]School of Exercise Science, Australian Catholic University, Sydney, NSW, Australia.
c/- The School of Exercise Science
Australian Catholic University
Locked Bag 2002, Strathfield, NSW 2135, Australia
Paul.taylor@acu.edu.au
kweeyum@hotmail.com
Raul.landeo@acu.edu.au

[b]School of Mathematics and Statistics, The University of Western Australia, Perth, WA Australia
c/- School of Mathematics and Statistics
The University of Western Australia
35 Stirling Highway
Crawley, Perth, WA 6009, Australia
Michael.small@uwa.edu.au

[c] The New South Wales Institute of Sport, Sydney, NSW, Australia
Building B Level 1
6 Figtree Drive
Sydney Olympic Park, NSW 2127, Australia
Damien.omeara@nswis.com.au
Emma.millett@nswis.com.au

Date of submission: **6th July 2015**

Corresponding Author: Paul Taylor

c/- The School of Exercise Science

Australian Catholic University

Locked Bag 2002, Strathfield, NSW 2135, Australia

paul.taylor@acu.edu.au

Abstract

Entropy is an effective tool for investigation of human movement variability. However, before applying entropy, it can be beneficial to employ analyses to confirm that observed data is not solely the result of stochastic processes. This can be achieved by contrasting observed data with that produced using surrogate methods. Unlike continuous movement, no appropriate method has been applied to discrete human movement. This article proposes a novel surrogate method for discrete movement data, outlining the processes for determining its critical values. The proposed technique reliably generated surrogates for discrete joint angle time series, destroying fine-scale dynamics of the observed signal, while maintaining macro structural characteristics. Comparison of entropy estimates indicated observed signals had greater regularity than surrogates and were not only the result of stochastic but also deterministic processes. The proposed surrogate method is both a valid and reliable technique to investigate determinism in other discrete human movement time series.

Introduction

Human movement variability has received increasing attention over the last 30 years and has historically been attributed to noisiness within the neuromuscular system (Newell, Deutsch, Sosnoff, & Mayer-Kress, 2006). Contemporary investigations hypothesise that variability is not representative of purely stochastic processes but rather manifestation of intrinsic, deterministic, dynamical systems (Newell & Corcos, 1993), which can facilitate motor learning, improve performance and prevent injury (Bartlett, 2007; Davids, Glazier, Araujo, & Bartlett, 2003; Preatoni et al., 2013). Sample entropy is an effective tool for investigating movement variability (Preatoni, Ferrario, Dona, Hamill, & Rodano, 2010). Sample entropy can quantify 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. If a signal can be shown to contain deterministic dynamics then it may provide evidence against the null, variability as noise, hypothesis. Furthermore, it provides confidence that inferences made about observed changes or differences in regularity are the result of purposeful rather than random processes. 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). Many of these techniques deal with intrinsically stochastic signals. These methods may be applied to deterministic data by pre-filtering the observed signal to remove the deterministic component. However, segmentation of data into noise and deterministic components can result in spurious effects (Theiler & Eubank, 1993). When dealing with human movement data, surrogate methods designed for use with deterministic signals need to be considered. 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). This method derives a noise contaminated signal from a reconstruction of the underlying deterministic dynamic (a phase space created via time delay embedding consistent with Takens (1981) theorem). However, this method is inappropriate for discrete movements. This is due to the data consisting of *N* short time series rather than the type of continuous and repetitive time series which facilitate time delay embedding. That is, despite resembling a continuous and periodic variable when concatenated together, the final value of one trial/cycle does is not a neighbour to the initial value of the next, excluding the pseudo periodic surrogate and other surrogate methods which employ time delay embedding. Therefore, the purpose of this article is to propose a generalisation of the pseudo periodic surrogate method, without time delay embedding, which can be applied to discrete

movement data. It is expected that this technique will produce outcomes similar to those of the Small shuffled surrogate method (Nakamura & Small, 2005, 2006), whereby the sequence of data is shuffled on a fine scale, destroying the micro structure of the original data (relationship between each datum and those immediately surrounding it), while the macro structural elements of the data (mean, variance, length) are maintained. The use of the proposed technique, quantification of critical values and the implementation of sample entropy to test for deterministic dynamics within discrete human movement will then be outlined.

Method

Participants

This project was approved by the Australian Catholic University Human Research Ethics Committee. Ten male participants [24.1 (3.3) years; 176.6 (5.9) cm; 76.4 (7.8) kg] provided informed consent and had their data included in this study. The task chosen to demonstrate surrogate generation was an overarm throw toward a target. Participants were seated on an adjustable piano stool with knee and ankle angles approximating 90° and anatomical orientation respectively. The piano stool was placed 7 m from a projection screen (5 m x 3 m) upon which a 70 cm round target consisting of 5 concentric circles was projected with the target centre being at a height of 2 m. Participants were seated such that their frontal plane was oriented perpendicular to a line projected from the centre of the target to the piano stool. Participants attended two sessions where they performed two blocks of trials with 16 throws per block. The choice of 16 throws per block was based on previous work (Taylor, Lee, Landeo, O'Meara, & Millett, 2015). Kinematic data were collected using a 10 camera (6 MX and 4 T-series) Vicon (Oxford Metrics, Oxford, UK) motion capture system, operating at 400 Hz. A Basler A602fc camera (Basler AG, Germany) recording at 100 Hz was used to capture ball release for later data cropping. Following data collection, three-dimensional joint angles – shoulder internal/external rotation and flexion/extension at the elbow and wrist – were

calculated. All angle data were cropped from the first target-directed motion of the finger marker through to ball release. Following investigation of the residuals (Winter, 2005) and frequency content of the data, all time series were filtered at 12 Hz using a 4th order Butterworth filter.

Surrogate Technique

The following details the surrogate generation 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, with no new entries (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.

This method is documented here using two concatenated input variables (*X* and *Y* in step 1). Researchers should use the knowledge of their own data to ensure there is a suitable level of appropriateness when selecting these input variables, avoiding the use of unrelated or irrelevant combinations. However, as long as this level of appropriateness is maintained there is no theoretical limit to the number of input variables that are used to form the phase space *P* at step 2. As such the matrix *P* could be defined such that its dimensions are $V *$

$\sum_{i=1}^{N} T_i$ where *V* is equal to the number of input variables. Matrices *A* and *B* would then be *V* x *N* in dimension.

Determining *ρ*

An optimal value for *ρ* elicits the greatest number of small segments within the surrogated time series (Small, Yu, & Harrison, 2001), providing an optimal balance between effectively destroying the fine-scale dynamics of the signal and maintaining its macro structure. A small segment is defined as any run of surrogate data of length between 2 and the total length of the surrogate, identical to one existing at any point within the original data set. The segment is created when a switch in the sequence of data in *P*, currently being sampled to provide the next state of P_s, occurs. When *ρ* is very small (at or approaching zero) the number of small segments will be zero as original data and surrogate will be identical. As *ρ* increases, so too will the number of small segments, towards a maximum, before returning toward zero (as $\rho \rightarrow \infty$). A large range of values for *ρ* (0–5; increments of 0.1) were tested 100 times using a block of data of one participant (Figure 1a). This identified the probable range (0.1 – 0.9 and 0.1 – 2.0 for two and three dimensional phase spaces respectively) over which to test for individual peaks in small segments (Figure 1b). Each participant's data were then tested over this range five times, and the *ρ* value associated with the highest mean number of small segments was selected (e.g., Figure 1c). This resulted in an individualised value for *ρ* to be used for surrogate generation for each block of 16 throws for each participant.

Figure 1 near here

Discrete Data Surrogate Generation

To demonstrate the use of the technique with different multiples of input variables, two different surrogate generations were conducted. First, Elbow and Shoulder time series were concatenated and combined to form a two dimensional phase space from which the respective surrogates were drawn. Next, wrist time series were included to form a three dimensional phase space and the process was repeated. 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. This process resulted in two elbow and two shoulder surrogate time series, from the two and three dimensional phase space generation, being produced for each observed throw included in the study. In addition, one wrist surrogate was produced via the three dimensional phase space for each observed throw.

Validity and Reliability

The biomechanical data used in this investigation was filtered, as is convention, to remove any systematic noise introduced by the data collection equipment. However, since surrogate data can appear similar to unfiltered/raw data, the surrogate generation process was also carried out on the raw movement data in addition to the filtered data. This analysis ensured that any observed differences in regularity between the data and its surrogate was the result of the methodology and not due to increased regularity introduced to the signal via the post collection smoothing. That is, if the raw data and its surrogate, as well as the filtered data and its surrogate, are both significantly different in regularity, this can be attributed to the surrogate method and not to any other conditioning of the observed data.

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, the ability for these values to be produced reliably was tested by repeating the surrogate generation process 6 times for each included block of throws. The mean, SD and length of the resultant data were assessed for reliability using intraclass correlation and standardised typical error tests (Hopkins, 2000, 2011). This was performed for surrogates produced both via two and three dimensional phase space.

Comparing Real and Surrogate Data

Sample entropy values quantify the regularity of a signal by assessing the probability that two sequences of points extracted from a time series of length *N,* which are similar for a period of *m* points within a tolerance *r*, will remain similar for a period of *m + 1* points excluding self matches (Lake et al., 2002; Richman & Moorman, 2000). The sample entropy 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 estimates than surrogates as they are not solely the result of noisy, random processes, but contain some element of deterministic dynamics. The lower entropy estimate 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.

The choice of values for the parameters *m* and *r* will affect the outcome of the entropy estimate, and consistency between parameters used for real and surrogate data comparison is the key concern. Still, values of $m = 2$ and $m = 3$ as well as a range of *r* values (0.1 – 0.3) were tested as recommended (Yentes et al., 2013) to determine these values. As a result, the parameters of $m = 2$ and $r = 0.1$ were employed. Sample entropy estimated for the concatenated real and surrogate time series of the three joint angles for all blocks of throws. These estimates were compared using the Mann-Whitney U test. Non parametric statistics were employed as data did not display normality (Peat & Barton, 2005).

Results

Surrogate generation was successfully conducted via the documented algorithm using both two and three dimensional phase spaces. An example of concatenated real and surrogate data as well as a single real and surrogate throw can be seen in Figure 2 (two dimensional phase space).

Figure 2 near here

The comparison of macro characteristics (mean, SD and length) showed no significant differences between the real and surrogate throws ($p \geq 0.68$). There were also no significant differences between the mean, length and SD of elbow and shoulder surrogates produced via two and three dimensional phase space ($p \geq 0.61$). The group mean value of ρ was significantly higher for the three dimensional phase space surrogate generation ($p < 0.01$). However, the number of short segments produced by this increased ρ value was no different ($p = 0.55$) between two and three dimensional applications. 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 of ≤ 0.1 (Hopkins, 2000, 2011).

Comparison between the sample entropy estimate of real and surrogated data for Elbow, Shoulder and Wrist angles can be seen in Figure 3. Results of the Mann Whitney U tests indicated that observed time series had significantly lower sample entropy ($p \leq 0.05$) than their respective surrogate for all joint angles across both two and three dimensional phase space generation. This was observed for both the filtered and unfiltered/raw data. There was no significant difference between the entropy estimates of the elbow and shoulder surrogates produced via the two and three dimensional phase space ($p \geq 0.08$).

Figure 3 near here

Discussion

The purpose of this paper was to propose a surrogate generation method for discrete movement data and to illustrate its use - i.e., to demonstrate that these data were not solely the result of stochastic processes. Shoulder, elbow and wrist joint angle time series were taken from an overarm throwing task and appropriate surrogates generated. Reliability analyses suggest that this method can be depended upon to consistently produce the expected outcomes. All surrogate time series effectively maintained the overall trends in the observed data (Figure 2), as confirmed by the Mann Whitney U results showing no significant difference in the mean, SD and length between real and surrogate data. While the

macro characteristics of the observed data were maintained, comparison of the sample entropy estimate for both real and surrogate data (Figure 3) showed that the observed discrete human movement is not solely the product of non-deterministic 'noisy' processes. Furthermore, repeating the process using unfiltered/raw data produced the same results indicating that the differences between observed and surrogate data is the result of the surrogate method and not from any post-processing (increased regularity due to filtering) of the data.

The documented method is theoretically capable of producing surrogates using any number of input variables, greater than or equal to two, by creating an equally dimensioned phase space. To demonstrate this, two and three variables were used to form two and three dimensional phase spaces respectively. Results showed that surrogates were effectively created using both approaches. However, despite no significant differences in the macro characteristics, the entropy estimates or in the number of short segments created, the selected values for ρ were significantly higher for each participant in the three dimensional phase space approach. This can be attributed to the requirement of a greater noise radius to effectively select the nearest noisy neighbour due to the increased distance between trajectories that exist in a higher dimensional phase space. Qualitatively, it did appear that the increased ρ resulted in 'noisier' surrogates being produced via the three dimensional phase space, supported by the p values of the compared surrogates appearing to approach significance ($p \sim 0.08$). In addition to determining whether the variables being combined to form the phase space in this method are appropriate for the task, researchers should also ensure that the dimensions employed have the desired effect on surrogate outcomes.

This study is not the first to investigate the use of surrogate techniques with human movement data. Previous work using a pseudo periodic surrogate with normal walking and race walking (Miller et al., 2006; Preatoni et al., 2010) successfully displayed the presence of deterministic dynamics within the time series taken from these tasks. While the discrete data used in this study can appear cyclical when concatenated (Figure 2a), discontinuities are

present which do not exist in cyclical data. Hence, the discrete, separate trajectories of the current data required a new method capable of producing multiple surrogates with multiple random walks from a single phase space formed by embedding multiple observed time series as opposed to one created by time delay embedding (Takens, 1981) such as with pseudo periodic surrogates. Hence, the concatenated data of two or more joint rotations (step 1 in Surrogate Technique) are brought together to form the phase space (step 2 in Surrogate Technique) which maintained the biomechanical relationship between variables.

In conclusion, the proposed method effectively produced surrogates for comparison with collected discrete movement data. This comparison identified that the observed signal is not solely the result of stochastic processes suggesting the presence of deterministic dynamics. Coupled with the ability of the algorithm to consistently produce the expected outcome, the modified small shuffle surrogate method is both a valid and reliable technique to investigate the stated hypotheses in other discrete human movement time series.

Acknowledgements

Michael Small is funded by an Australian Research Council Future Fellowship (FT110100896) and Discovery Project (DP140100203). These organisations had no input in the conduct of this research.

References

Bartlett, R. (2007). Movement coordination and movement variability. *Sports Biomechanics, 6*(2), 119-120. doi: 10.1080/14763140701400931

Davids, K., Glazier, P., Araujo, D., & Bartlett, R. (2003). Movement systems as dynamical systems: The functional role of variability and its implications for sports medicine. *Sports Medicine, 33*(4), 245-260. doi: http://dx.doi.org/10.2165/00007256-200333040-00001

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. doi: http://dx.doi.org/10.1103/physreve.72.056216

Nakamura, T., & Small, M. (2006). Applying the method of small–shuffle surrogate data: testing for dynamics in fluctuating data with trends. *International Journal of Bifurcation and Chaos, 16*(12), 3581-3603. doi: http://dx.doi.org/10.1142/s0218127406016999

Newell, K. M., & Corcos, D. M. (1993). Issues in variability and motor control. In Newell & Corcos (Eds.), *Variability and Motor Control* (pp. 1-12). Champaign, IL: Human Kinetics.

Newell, K. M., Deutsch, K., Sosnoff, J., & Mayer-Kress, G. (2006). Variability in motor output as noise: A default and erroneous proposition. In K. Davids, S. Bennett & K. Newell (Eds.), *Movement system variability* (pp. 3-23). Champaign, IL: Human Kinetics.

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

Preatoni, E., Hamill, J., Harrison, A. J., Hayes, K., Van Emmerik, R. E. A., Wilson, C., & Rodano, R. (2013). Movement variability and skills monitoring in sports. *Sports Biomechanics, 12*(2), 69-92. doi: 10.1080/14763141.2012.738700

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. doi: http://dx.doi.org/10.1103/physrevlett.87.188101

Takens, F. (1981). Detecting strange attractors in turbulence. In D. Rand & L.-S. Young (Eds.), *Dynamical Systems and Turbulence, Warwick 1980* (Vol. 898, pp. 366-381). Berlin: Springer.

Taylor, P. G., Lee, K.-Y., Landeo, R., O'Meara, D. M., & Millett, E. (2015). Determining optimal trial size using sequential analysis. *Journal of Sports Sciences, 33*(3), 300-308. doi: 10.1080/02640414.2014.942679

Theiler, J., & Eubank, S. (1993). Don't bleach chaotic data. *Chaos, 3*(4), 771 - 782. doi: http://dx.doi.org/10.1063/1.165936

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

Winter, D. A. (2005). *Biomechanics and motor control of human movement* (3rd ed.). Hoboken, NJ: Wiley.

Yentes, J., Hunt, N., Schmid, K., Kaipust, J., McGrath, D., & Stergiou, N. (2013). The Appropriate Use of Approximate Entropy and Sample Entropy with Short Data Sets. *Annals of Biomedical Engineering, 41*(2), 349-365. doi: 10.1007/s10439-012-0668-3

Figure Captions

Figure 1: Results of testing over a large range of ρ values (a), probable range for individual values of ρ (b) and results of testing over this range for a single participant (c)

Figure 2: All throws concatenated and a single throw for observed (a & b) and surrogate (c & d) data

Figure 3: Median (± inter-quartile range) sample entropy estimate for observed and surrogate data across the three included joint rotations. All surrogate data sample entropy estimates were significantly greater than their respective observed data estimate ($p < 0.05$).

Figure 1

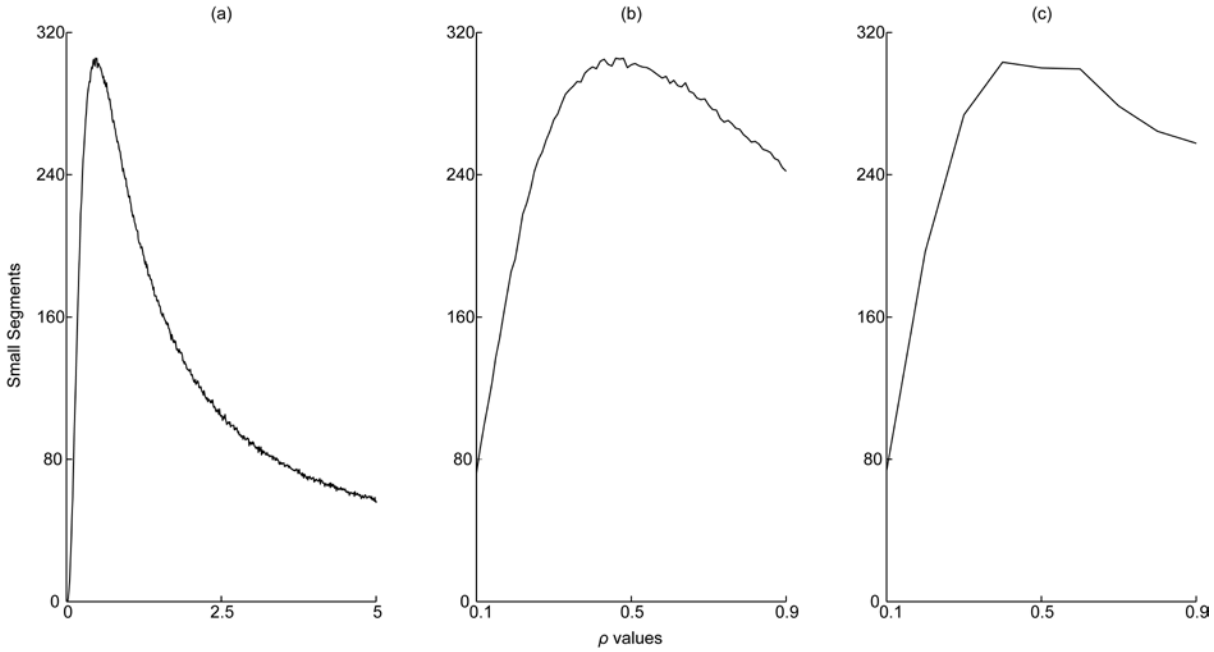


Figure 2

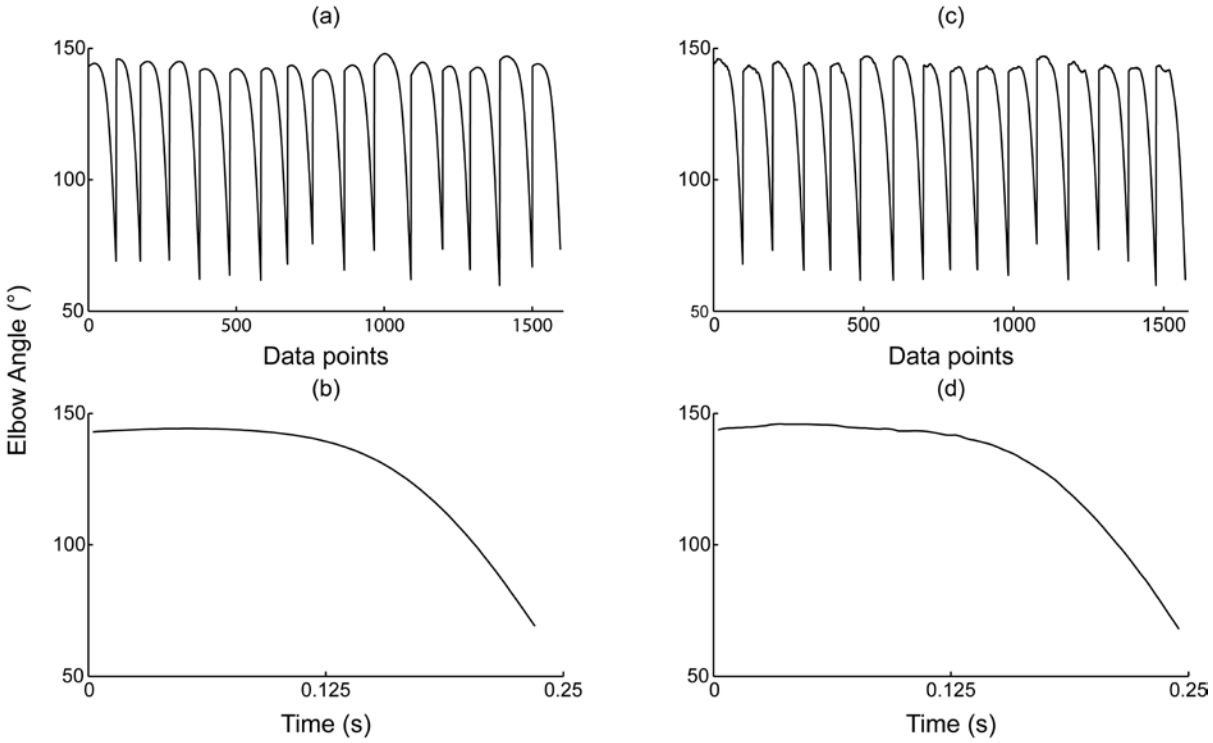
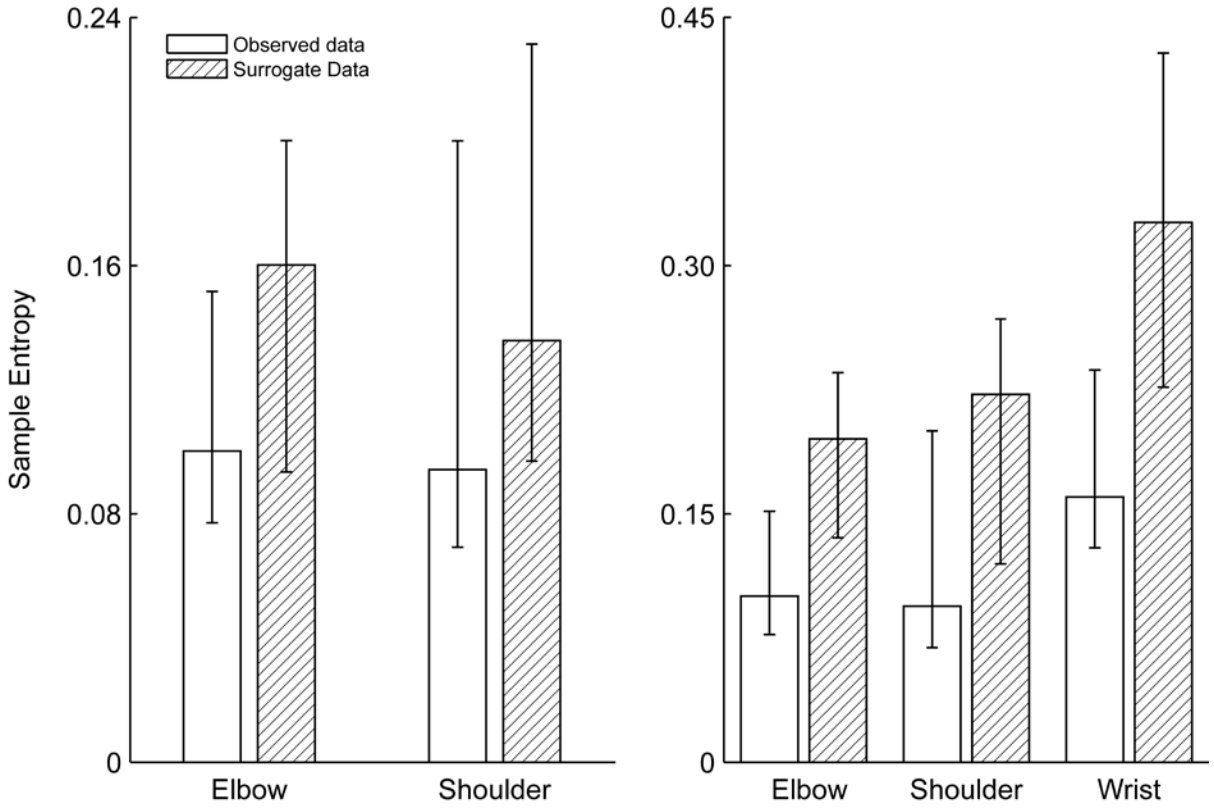

(a)
(c)
(b)
(d)
Elbow Angle (°)
Data points
Time (s)
150
100
50
0
500
1000
1500
0.125
0.25

Figure 3


Observed data
Surrogate Data
Sample Entropy
0.24
0.16
0.08
0
Elbow
Shoulder
0.45
0.30
0.15
0
Elbow
Shoulder
Wrist