2015 International Symposium on Nonlinear Theory and its Applications
NOLTA2015, Kowloon, Hong Kong, China, December 1-4, 2015

Revisiting surrogate generation for cyclic time series

Michael Small[†] Jiahao Su[†,‡] and Ming Wai Chan[†]

†School of Mathematics and Statistics, University of Western Australia
Crawley, WA, Australia.
‡Electrical and Computer Engineering, University of Maryland
College Park, Maryland, United States
Email: michael.small@uwa.edu.au

Abstract—The cycle shuffled surrogate algorithm provides a straightforward method to randomise cyclic time series. This randomisation can then be employed as a form of Monte-Carlo hypothesis testing — do the randomised realisations differ, statistically, from the original? If they do, then one may conclude (with some additional caveats) that the original data included deterministic inter-cycle dynamics: deterministic chaos, for example. In this communication we will re-examine this algorithm, point to several technical issues that may arise and discuss suitable palliatives.

1. Introduction

When chaotic dynamics and nonlinear time series analysis first became *en vogue* [12], evidence of chaos arose in a wide range of settings (for example [13]). However, the primary indicators of chaos; positive Lyapunov exponents [19] and fractional correlation dimension [2, 3]; were problematic for several reasons. Some of those problems were technical and address with later, improved, algorithms [12]. However, another statistical problem remained: to say with confidence that the observed data represented something "'interesting" one needed a statistical model of the appropriate "boring" alternatives. For random looking signals, the method of surrogate data came to the rescue [17].

Surrogate data methods provide a framework with which to test whether observed data is consistent with a specific hypothesis by generating an ensemble of realisations —- the *surrogates* — that are both "like" the original data and also consistent with the hypothesis under consideration. For example, to test the hypothesis that observed data is independent and identically distributed noise, one can generate surrogates by shuffling the order of observations in the original time series. Other hypotheses posited by Theiler [17] tested for correlated noise by shuffling the phases of the Fourier transform. These algorithms worked well and spawned a cottage industry in developing increasingly esoteric surrogate generation algorithms (the contributions of the current first author include [8, 9, 10, 11, 7, 5]). However, problems remain in those instances where the surrogate simply looked wrong. In particular, if the data had cyclic oscillations then most of these methods would lead to rejection of the hypothesis under consideration simply because the surrogates look different to the data.

Theiler and Rapp [18] initially demonstrated a solution to this problem in 1996 — the *cycle surrogate algorithm*. The algorithm is very simple and intuitive: first break the time series signal into cycles and then shuffle the order of those cycles before reassembling the signal. Long term deterministic dynamics would be destroyed and one could test the hypothesis of a noisy periodic orbit. The method was originally demonstrated for, and shown to work well with, strongly cyclic time series of epileptic electroencephalogram recordings. The mechanistic shuffling posed some conceptual problems for the statement of the null hypothesis and did not always work equally smoothly for all time series. Some finesse was required and an embedding based [16] alternative [15, 14] was proposed to circumvent these issues.

However, the cycle shuffle algorithm remains, and remains intuitive and attractive. Unfortunately, it does still suffer from some issues, including problems of aliasing which have previously not been properly recognised. In the following sections we catalogue these problems, and propose solutions.

2. Corrected cycle shuffled surrogates

The cycle shuffle surrogate algorithm was briefly described in the introduction and is illustrated in Fig. 1. Figure 1 depicts a typical experimental time series — in this case a sound recording of a constant tone. The sound is produced by the vibration of an air column in a woodwind instrument with air intake periodical occluded by a vibrating reed. The initial and final phase of the sound have been removed and depicted in Fig. 1 is the presumably "stationary" sustained phase. As the sound is a single musical tone it is approximately periodic and even typically characterised by its Fourier spectrum. Just as with vibrating strings [6, 4] it is natural to ask whether this recording is the output of a chaotic dynamical system. To test this hypothesis we can estimate Lyapunov exponents, correlation dimension and build nonlinear models. However, to ensure that the results we obtain are not artefacts of some simpler system, surrogate data should be generated and also tested alongside the original. The challenge is to generate surrogate data that looks like this experimental recording

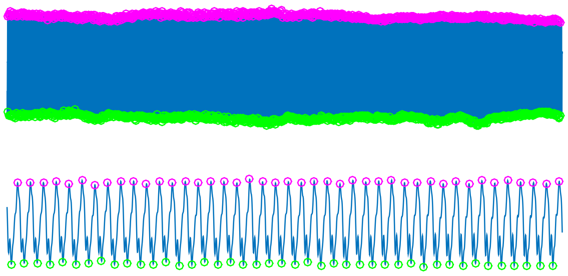

Figure 1: **The cycle shuffled surrogate method.** Two views of the same time series (a sound recording from a standard B♭ clarinet intoning the note A_5, toward the upper end of the clarino register, sustain only — attack and decay removed). Peaks and troughs are identified (using a sliding window method) to identify individual cycles of the recording. These are then separated and then pieced back together in a random order in Fig. 2.

but that contains no long-term (inter-cycle) determinism. Hence, cycle shuffled surrogates are applied to the cyclic decomposition of Fig. 1.

In Fig. 1 we have identified the location of each peak and trough of each cycle. While there are also issues associated with doing this automatically, we have manually checked that the estimates in this case are accurate. From these individual cycles, we can then break the time series into these cycles, re-arrange and re-order, and re-assemble — shifting cycles vertically to ensure continuity. Unfortunately, ensuring continuity introduces long-term non-stationarity through correlations across the peaks or troughs. In each sub-panel of Fig. 2 this non-stationarity is evident.

In would be trivial to detect statistical difference between the time series in Fig. 1 and a random ensemble like those evident in Fig 2. However, it would be erroneous to conclude that the original data contains long term deterministic dynamics — the non-stationarity introduced in these surrogates may merely be due to correlations across the breakpoints. The problem is that when the surrogates are re-assembled each individual cycle is moved vertically to preserve continuity. Unfortunately, this amounts to a shuffling of the difference between the trough locations — and even if this is a random process it is not necessary mean-zero and so the shuffled version is not necessarily stationary. The solution that we are exploring is to perform a linear dilation, affine transformation or rotation (one is free to choose from amongst the various alternatives — each only needing an additional free parameter to enforce stationarity) on each cycle in addition to the vertical translation — this ensures that not only are the break points continuous, but that the location of each trough is fixed between the surrogates.

Let $\{y_t\}_{t=1}^{N}$ be the time series and suppose that we identify breakpoints $\{t_i\}_{i=1}^{k}$ ($t_i < t_{i+1}$ and $t_i \in [1,N] \cap \mathbf{Z}$) splitting the time series into $k-1$ cycles. Without loss of generality,

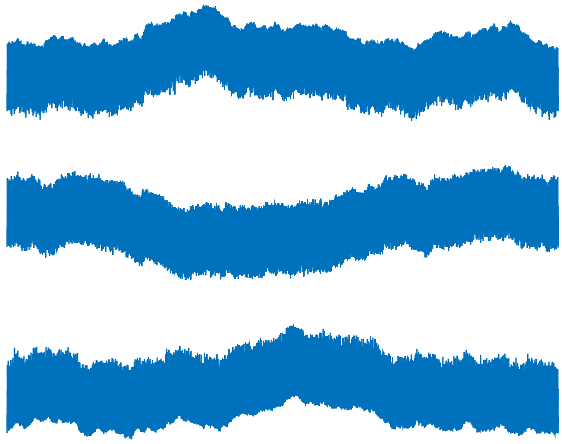

Figure 2: **Naïve cycle shuffled surrogates for experimental data.** Three standard cycle shuffled surrogates are depicted for the time series in Fig. 1. In the upper panel cycles are split at the peak; in the middle panel at the mid-point of each cycle; and, in the lower panel at the trough. In each case the act of piecing these back together has introduced non-stationarity not present in the original. One could trivially conclude that the null hypothesis is not true, however, is this evidence of nonlinear determinism?

suppose $t_1 = 1$ and $t_k = N$ (otherwise, we keep whatever occurs before t_1 and after t_k fixed). A cycle shuffle surrogate is generated by constructing the sequence of pairs

$$\{(t_1,t_2),(t_2,t_3),(t_3,t_4)\ldots,(t_{k-1},t_k)\}$$

and then reordering those pairs

$$\{(t_{\pi(1)},t_{\pi(1)+1}),(t_{\pi(2)},t_{\pi(2)+1}),(t_{\pi(3)},t_{\pi(3)+1})\ldots,\\ (t_{\pi(k-1)},t_{\pi(k-1)+1})\}$$

where π is a permutation of the integers $1,2,3,\ldots,k-1$. The surrogate z_t is then constructed iteratively from the segments

$$\{y_{t_{\pi(i)}},\ldots,y_{t_{\pi(i)+1}}\}.$$

Let

$$C_i = \{y_{t_{\pi(i)}},\ldots,y_{t_{\pi(i)+1}}\}$$

denote the i-th such segment. For each segment we perform a translation correction to obtain $\tilde{C}_i = C_i - y_{t_{\pi(i-1)+1}}$ to ensure continuity, and then z_t is the concatenation of these corrected segments $z_t := \{\tilde{C}_1|\tilde{C}_2|\cdots|\tilde{C}_k\}$. This is the process depicted in Fig. 2, and it is clear that preserving continuity is at the expense of stationarity.

To preserve both continuity and stationarity we must replace the correction operation $\tilde{C}$ with a more complex translation and possibly either rotation or dilation to ensure that both endpoints are preserved. Whereas $\tilde{C}_i$ translates C_i that the first point of $\tilde{C}_i$ is identical to the successor of the last point of $C_i \tilde{-} 1$ we must also ensure that the final

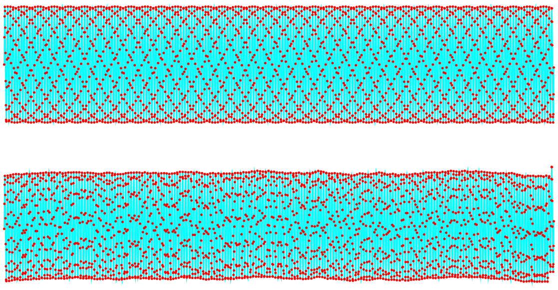

Figure 3: **Naïve cycle shuffled surrogates for an exactly periodic time series.** Depicted as a solid cyan line is the time series, sampled at the sampling times of the red dots. Clearly there is an aliasing effect due to sampling at a frequency not exactly divisible by the period. When the cycle surrogate is constructed (with any of the methods described in this communication) that long-term aliasing pattern is destroyed — the lower panel.

points line up as well. This introduces a single additional degree of freedom — we can either parameterise that with a dilation (scaling $\tilde{C}_i$ by a factor λ), affine transformation ($y = mx + b$) or a rotation (through an angle θ). Either approach will ensure that the sequence of breakpoints are preserved $y_{t_i} \equiv z_{t_i}$ for all i and hence the surrogates are stationary in the same sense as the original data. The randomisation achieved by shuffling the data is then a genuine randomisation of the shape of the individual cycles and a proper test of no inter-cycle determinism.

3. Aliasing

Implementing this solution works well for a wide range of experimental time series — including those depicted here. But it is worth noting that it does present new challenges. Consider strictly periodic orbits of a linear or nonlinear system and require only that the period is not exactly divisible by the sampling time interval. A typical such time series is depicted in the upper panel of Fig. 3. There is a clear visual pattern in the location of the sampled points, due to this aliasing. The cycle shuffled surrogate (constructed with any of the methods described in this paper) destroys that pattern and produces subtle variation that can be detected as a statistical discrepancy between data and surrogate (we have found that self-mutual information is sufficiently sensitive to detect this discrepancy). One would therefore (incorrectly) reject the hypothesis that this signal is periodic. The presence of small to moderate observation noise[1] does not resolve this problem.

At present we do not have a suitable correction to overcome the problem of aliasing. However, it is necessary only to be aware of it and interpret results of these surrogate algorithms accordingly. It is really only a problem with strictly periodic signals — in which case the nature of the signal should be obvious. For more natural signals the dilation and rotation corrections are sufficient. In the next, and final, section we apply these corrected cycle shuffled methods to the data of Fig. 1 with typical nonlinear measures as test statistics. Moreover, aliasing is only truly a "problem" as we are asking the wrong question — clearly the aliased times series *does* have non-trivial long term deterministic dynamics, they are just not properly of the underlying dynamical system.

[1] Exactly how much noise one can tolerate is a function of the sampling time as well.

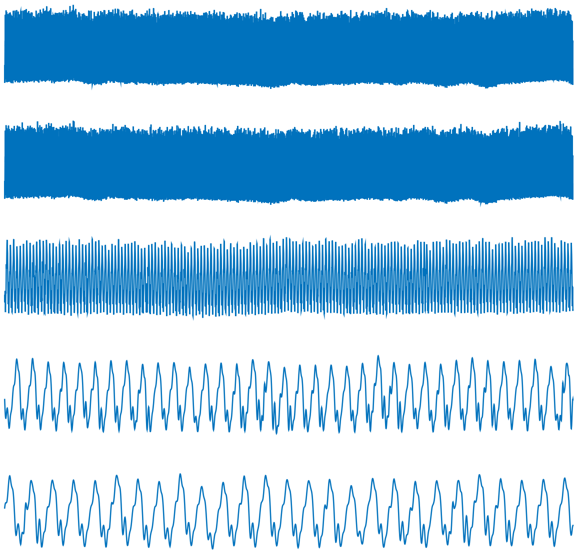

Figure 4: **Corrected cycle shuffled surrogates.** Here we depict five corrected (via an affine in time correction) cycle shuffled surrogates of the data depicted in Fig. 1. The lower three panels are enlarged so as to depict the detail and illustrate the cycle-to-cycle variability. Clearly, these surrogates look much more like the original data than the surrogates of Fig. 2.

4. The chaotic clarinet

We now conclude by returning to the experimental data of Fig. 1. In Fig. 4 we depict five surrogates generated in such a way as to preserve the trough values precisely using an affine transformation as described in this paper. Clearly, the surrogate now "look" like the data. Drift associated with the naïve application of the cycle shuffled method in Fig. 2 is now eliminated. Figure 5 illustrates a comparison of estimates of correlation dimension, entropy and noise level [20, 1] for the experimental data and an ensemble of 50 surrogates. As the data is clearly atypical of the surrogates we reject the null hypothesis and conclude that the data contains long-term deterministic dynamics —

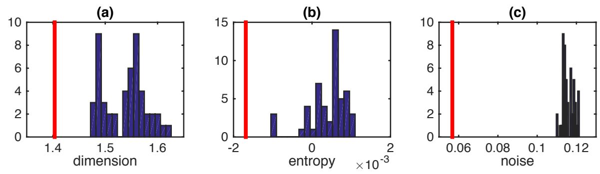

Figure 5: **Hypothesis test** In the three panels we present histograms of: (a) correlation dimension; (b) entropy; and, (c) noise using the Gaussian Kernel Algorithm (embedding lag of 28 and embedding dimension of 2 and 5). The value for the data is depicted by a red bar and the ensemble for the surrogates as the blue histogram. For all three statistics, the data is significantly different from the surrogates.

it is consistent with a chaotic dynamical system.

Finally, we must note that the corrected cycle shuffled algorithm implemented in this section (there are alternatives, which we have only alluded to) keeps the trough positions fixed. If one was to construct a Poincaré section from these trough values this would be a trivial and useless statistic with which to test the underlying dynamics. This, in part, is the reason for our choice of attractor based statistics in this section.

Acknowledgement

MS is support by an Australian Research Council Future Fellowship (FT110100896) and Discovery Project (DP140100203). JS was support by a UWA-USTC winter research training internship.

References

[1] C. Diks. Estimating invariants of noisy attractors. *Physical Review E*, 53:R4263–R4266, 1996.

[2] P. Grassberger and I. Procaccia. Characterization of strange attractors. *Physical Review Letters*, 50:346–349, 1983.

[3] P. Grassberger and I. Procaccia. Measuring the strangeness of strange attractors. *Physica D*, 9:189–208, 1983.

[4] K. Judd and A. Mees. Modeling chaotic motions of a string from experimental data. *Physica D*, 92:221–236, 1996.

[5] X. Luo, T. Nakamura, and M. Small. Surrogate test to distinguish between chaotic and pseudo periodic time series data. *Physical Review E*, 71:026230, 2005.

[6] T. Molteno and N. Tufillaro. Torus doubling and chaotic string vibrations — experimental results. *J Sound Vib*, 137:327–330, 1990.

[7] T. Nakamura, Y. Hirata, and M. Small. Testing for correlation structures in short term variabilities with long term trends of multivariate time series. *Physical Review E*, 74:41114, 2006.

[8] T. Nakamura, X. Luo, and M. Small. Testing for nonlinearity in time series without the Fourier transform. *Physical Review E*, 72:055201(R), 2005.

[9] T. Nakamura and M. Small. Testing for dynamics in the irregular fluctuations of financial data. *Physica A*, 366:377–386, 2006.

[10] T. Nakamura and M. Small. Testing for random walk. *Physics Letters A*, 362:189–197, 2006.

[11] T. Nakamura, M. Small, and Y. Hirata. Testing for nonlinearity in irregular fluctuations with long term trends. *Physical Review E*, 74:026205, 2006.

[12] M. Small. *Applied Nonlinear Time Series Analysis: Applications in Physics, Physiology and Finance*, volume 52 of *Nonlinear Science Series A*. World Scientific, Singapore, 2005.

[13] M. Small, K. Judd, M. Lowe, and S. Stick. Is breathing in infants chaotic? Dimension estimates for respiratory patterns during quiet sleep. *Journal of Applied Physiology*, 86:359–376, 1999.

[14] M. Small and C. Tse. Applying the method of surrogate data to cyclic time series. *Physica D*, 164:187–201, 2002.

[15] M. Small, D. Yu, and R. G. Harrison. Surrogate test for pseudo-periodic time series data. *Physical Review Letters*, 87:188101, 2001.

[16] F. Takens. Detecting strange attractors in turbulence. *Lecture Notes in Mathematics*, 898:366–381, 1981.

[17] J. Theiler and D. Prichard. Constrained-realization Monte-Carlo method for hypothesis testing. *Physica D*, 94:221–235, 1996.

[18] J. Theiler and P. Rapp. Re-examination of the evidence for low-dimensional, nonlinear structure in the human electroencephalogram. *Electroencephalogr Clin Neurophysiol*, 98:213–222, 1996.

[19] A. Wolf, J. B. Swift, H. L. Swinney, and J. A. Vastano. Determining Lyapunov exponents from a time series. *Physica D*, 16:285–317, 1985.

[20] D. Yu, M. Small, R. G. Harrison, and C. Diks. Efficient implementation of the Gaussian kernel algorithm in estimating invariants and noise level from noisy time series data. *Physical Review E*, 61:3750–3756, 2000.