2014 International Symposium on Nonlinear Theory and its Applications
NOLTA2014, Luzern, Switzerland, September 14-18, 2014

Generating surrogate time series from complex networks

Michael Small

Department of Mathematics and Statistics, University of Western Australia
Crawley, Western Australia, Australia.
Email: michael.small@uwa.edu.au

Abstract—The method of surrogate data is now well established and provides a framework for generating random signals that may then be used to test specific statistical hypotheses. We propose to extend this rationale to the various complex network reconstruction methods. Starting with an experimental time series we suggest building a complex network that represents the underlying dynamics of that signal. Then, using that network we generate random equivalent time series. This paper provides a brief review and road-map.

1. Surrogate Data

Surrogate data [26] arise when performing hypothesis testing with time series data [16]. The basic idea is that one generates, from a measured time series, an ensemble of time series that are both "like" the original and also consistent with a specific null hypothesis. By computing the value of some statistic on the data and the surrogates, and comparing the statistical distribution for the surrogates to the value for the true data one can say whether the surrogates are typical — and therefore whether the true data is consistent with the underlying hypothesis.

For example, to test whether an observed time series is consistent with independent and identically distributed noise, one may randomly shuffle the observed time series data. This freshly re-ordered time series has no temporal correlation (since the data occur in a random order) and yet has the same probability distribution as the data (since the data have been resampled without replacement, the surrogate consists of exactly the same numbers — just in a different sequence). Measuring statistical properties for an ensemble of surrogates, and then comparing these to the original data (for example, by measuring autocorrelation, complexity [34] or entropy [2]) one can make a statistical statement about whether the data is typical of the null hypothesis.

More complicated surrogate algorithms exist — including: the original tests for linear noise and a monotonic nonlinear transformation of linear noise [26], various extensions thereof [7, 3, 15, 14, 13], cyclic determinism [28], periodic orbits [22, 9], noisy trends [12], and many others.

Notable among these techniques are methods which deal with oscillatory signals. Methods such as the cycle shuffled surrogate approach [28] break the original time series into individual cycles and then re-order and reconnect these cycles. This preserves the dynamics within cycles, but breaks the dynamics across cycles. The so-call pseudo-periodic surrogate methods [22] provides a continuous extension of this idea. Rather than discrete breaks between cycles, the perturbations occur continuously and mirror the local dynamics [19]. The method actually adapts the idea of local modelling first proposed by Sugihara and May [23]. The same technique has also been adapted to test for state dependent noise processes [20] and shown to provide varying granularity[1] by judicious adjustment to the algorithm parameters [21].

The various algorithms described so far construct surrogate data which, by definition, are consistent with a specific hypothesis. There is an alternative interpretation. One can think of these algorithms as *models* of the observed data — consistent with some specific hypothesis [18]. The surrogates that these models generate are the simulations from the model. The hypothesis being test is characterised by the behaviour of the model. While there are technical issues [27, 18] concerning when the model is adequate to address a general hypothesis rather than a specific one (i.e. that the test is testing, for example, against all models that would exhibit periodic orbits, or just that particular one that was built) the basic principal is sound. In section 3 we think of a complex network as a new type of model of a dynamical system, and then seek to generate alternative time series that are independent realisations of the appropriate hypothesis. First, we need to introduce the idea of generating complex networks from time series. We do this in section 2.

2. Complex Networks

Based on the search for methods to distinguish deterministic aperiodic dynamics (loosely, "chaos") from noisy periodic orbits[2], Zhang, Luo and I [31] developed an algorithmic method to measure the periodic component in an oscillatory time series. The method did not require embedding [25] and essentially sought a similarity index between

[1] Pun intended.

[2] Of course, the distinction is formerly moot: noise can be considered as nothing more than very high dimensional dynamics, and, conversely, low dimensional chaos measured in a time series can be decomposed into a periodic signal plus small perturbations. Nonetheless, what we are really asking is are the perturbations from a purely periodic regime best modelled with deterministic or stochastic techniques.

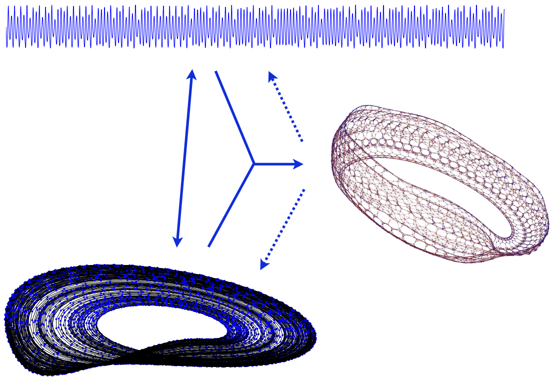

Figure 1: **Time series, attractors and networks.** The celebrated Taken's embedding theorem provides a methods to reconstruct an object equivalent to the underlying attractor from a scalar time series. The process is, of course, reversible. Methods exist to construct complex network representations of the dynamics from either the attractor or directly from the original time series. Procedures to do the reverse — creating equivalent time series or attractors from the network — are currently lacking.

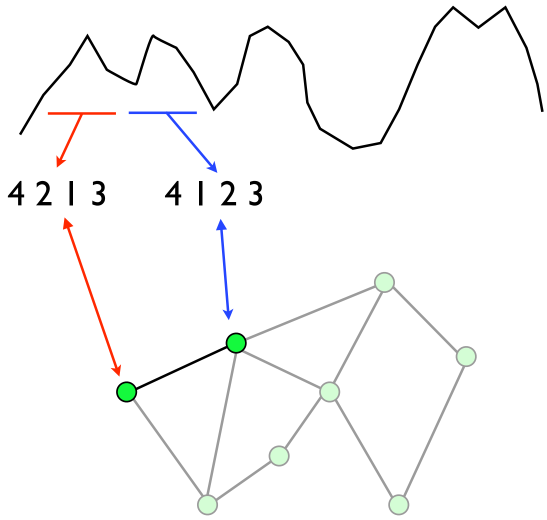

Figure 2: **Encoding time and space in a network.** Data within a window w is encoded as an ordinal sequence (describing the relative ordering of the observations) and then mapped to nodes on the network. Each ordinal sequence is a single node. Nodes are linked if the corresponding ordinal sequences occur in succession in the time series.

successive cycles. It was effectively an intermediate step from the surrogate methods developed previously and the idea of building a network from a time series [33], which Zhang and I first proposed in [32]. Fig. 1 provides a basic caricature of the correspondence between time series, attractor, and network.

The idea of [32] is actually deceptively simple. Take a time series and divide it into cycles — this is exactly the same process as [28], and the best way to approach it is still not clear[3]. Identify each cycle with a node of a network. Connect nodes if the corresponding cycles are "sufficiently close". Again, how to measure closeness is not clear, but there are perfectly sensible approaches outlined in both [31] and [32]. By doing this, one obtains a complex network representation of the topological relationship between cycles in the time series. As one would expect the resultant network for periodic or chaotic signals differ, and there is evidence of the complex network hubs corresponding to orbits of the system unstable periodic orbits [32].

Many alternative methods to construct networks from time series followed [8, 29, 10, 11, 30, 5] and these have been summarised in [4]. With the exception of visibility graph methods [8, 10], all other methods assign proximity based on similarity of state, while the visibility methods assign links to neighbours in time. There is a division between state-space *proximity* based methods and *temporal* closeness. In every case, these networks have found a wide range of applications and have helped to unveil features in experimental data not apparent to standard nonlinear time series techniques [16]. The discussion in [4] and references therein provide many examples, our own recent work is presented in [21].

Proximity based networks have turned out to be best at capturing topological properties of underlying *deterministic* dynamical systems, whereas temporal networks have been successful at characterising correlation structure in *stochastic* processes. While it is nice that the proximity methods provide a kind of dual space to the original attractor[4], generating simulations (or surrogates) from a model really requires the model to encode the deterministic temporal information directly.

We first proposed a method to do this in [17] and our first complete implementation is described in [24]. The method deviates from other approaches in that sequences of scalar points, of length w, are first encoded as an ordinal sequence [2, 1]. That is, the scalar points are replace by one of $w!$ sequences of the integers $1, 2, 3, \ldots, w$ where that sequence encodes their relative size. Those $w!$ possible sequence are then mapped to nodes of a network and the nodes are connected if the corresponding ordinal sequences occur in succession. Figure 2 summarises the process. There are several possible variants to this scheme, and [24] provides

[3]While apparently straightforward, the problem of robustly and computationally identifying cycles in a noisy (especially with correlated noise) time series is both ill-defined and deceptively complicated.

[4]For the proximity network, nodes are linked if they are close in space, whereas the trajectory of an embedded system links states that occur successively in time.

details of just one of these — we are working on alternatives.

Intuitively, this scheme combines both topological information of the attractor (via the ordinal sequences [1]) and also temporal information. This allows us to produce a random walk on the network which mimics the true dynamics of the underlying system.

3. Surrogate Models and Networks

The pseudo-periodic surrogate technique [22] generates random walks on the attractor, following the dynamics and moving at random between neighbours. What we propose here is a similar idea for networks created from time series — some sort of random walk on the network that can be mapped back into the original time series space to create a random trajectory through the same dynamical space (defined by the network).

There are a couple of restrictions to this plan. First, the proximity networks described above do not encode temporal information and, moreover, one needs a way to take the abstract nodes of a network and map them back to time series data. The approach we take to achieve this is to simply preserve this information in the network. That is, we keep the temporal information and a representative magnitude for each node. Hence we will probably need to associate with each node v_i a time t_i (this only makes sense for the original temporal networks and not the ordinal partition construction) and a representative time value $x_i \in \mathbf{R}$.

For any obvious implementation of such a random walk scheme on the proximity networks, one arrives at exactly the pseudo-periodic surrogate technique [22] — and doing so is only possible by preserving the information (t_i, x_i) for each node v_i of the network. This is because nodes in the proximity network exist in a one-to-one correspondence with embedded time series points, and moving between them is exactly equivalent to moving between the embedded points — the t_i are required to progress to the successor of each state (from node v_i, apply a random perturbation to arrive at one of its neighbours v_j and then move to its temporal successor, v_k, where $t_k = t_j+1$). As a side remark we note that computational simulations indicate that the *k-neighbour* approach [29] to proximity networks better match the dynamics of a deterministic dynamical system than the *ϵ-ball* approach [11]. The reason for this is related to the relative sparsity of the k-neighbour network and the fact that it constructs networks invariant under homeomorphic deformation of the phase space embedding [6].

Nonetheless, the ordinal approach will possibly be more interesting, because the nodes now encode an abstract dynamical state, rather than an individual point. Moreover, by constructing a directed and weighted network, one has a perfectly natural way of evolving between nodes in the network (from node v_i select one of the outgoing links with probability proportional to its weight and then move to that node v_j). This creates a random walk as a sequence of nodes on the network (essentially, a Markov chain over the ordinal partitions). However, because we cannot associate network nodes v_i with unique dynamical states, getting back to the time series is complicated.

Each node v_i has, associated with it, a set of time series values $\mathcal{X}_i = \{x_1, x_2, \ldots, x_k\}$. These are all the time series values that occupy (for example) the first component of occurrences of the ordinal sequence corresponding to v_i. Unfortunately, it is not sufficient to choose a member of that set at random: this will lead to an unnaturally discontinuous surrogate signal. Instead, our random walk must generate a sequence of N sets $\{\mathcal{X}_i\}_i$. Employing a greedy optimisation algorithm (which we are still working on) one selected members of these sets $z_i \in \mathcal{X}_i$ (one member from each set) in such a way as to maximise continuity:

$$\min \sum_{i=2}^{N} |x_{i-1} - x_i|.$$

Acknowledgement

MS is support by an Australian Research Council Future Fellowship (FT110100896) and Discovery Project (DP140100203). This report is the result of ongoing work which will be reported in full at a later date. Much of that work was done by Lvlin Hou, Konstantinos Sakellariou, and Michael McCullough.

References

[1] J. Amigó, S. Zambrano, and M. Sanjúan. True and false forbidden patterns in deterministic and random dynamics. *Europhys Lett*, 79:50001–p1–p5, 2007.

[2] C. Bandt and B. Pompe. Permutation entropy: A natural complexity measure for time series. *Physical Review Letters*, 88:174102, 2002.

[3] K. T. Dolan and M. L. Spano. Surrogate for nonlinear time series analysis. *Physical Review E*, 64:046128, 2000.

[4] R. V. Donner, M. Small, J. F. Donges, N. Marwan, Y. Zou, R. Xiang, and J. Kurths. Recurrence-based time series analysis by means of complex network methods. *International Journal of Bifurcation and Chaos*, 21:1019–1046, 2011.

[5] Z. Gao and N. Jin. Complex network from time series based on phase space reconstruction. *Chaos*, 19:033137, 2009.

[6] L. Hou. Private communication, 2014.

[7] D. Kugiumtzis. Surrogate data test for nonlinearity including nonmonotic transforms. *Physical Review E*, 62:R25–R28, 2000.

[8] L. Lacasa, B. Luque, F. Ballesteros, J. Luque, and J. Nuño. From time series to complex networks: The visibility graph. *Proc Natl Acad Sci USA*, 105(13):4972–4975, 2008.

[9] 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.

[10] B. Luque, L. Lacasa, F. Ballesteros, and J. Luque. Horizontal visibility graphs: Exact results for random time series. *Phys. Rev. E*, 80:046103, 2009.

[11] N. Marwan, J. F. Donges, Y. Zou, R. V. Donner, and J. Kurths. Complex network approach for recurrence analysis of time series. *Physics Letters A*, 373(4246–4254), 2009.

[12] T. Nakamura and M. Small. Small-shuffle surrogate data: Testing for dynamics in fluctuating data with trends. *Physical Review E*, 72:056216, 2005.

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

[14] T. Schreiber and A. Schmitz. Improved surrogate data for nonlinearity tests. *Physical Review Letters*, 77:635–638, 1996.

[15] T. Schreiber and A. Schmitz. Surrogate time series. *Physica D*, 142:346–382, 2000.

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

[17] M. Small. Complex networks from time series: Capturing dynamics. In *IEEE International Symposium on Circuits and Systems Proceedings*, pages 2509–2512, 2013.

[18] M. Small and K. Judd. Correlation dimension: A pivotal statistic for non-constrained realizations of composite hypotheses in surrogate data analysis. *Physica D*, 120:386–400, 1998.

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

[20] M. Small and C. Tse. Determinism in financial time series. *Studies in Nonlinear Dynamics and Econometrics*, 7(3), 2003. Article 5.

[21] M. Small, D. M. Walker, A. Tordesillas, and C. K. Tse. Characterising chaotic dynamics from simulations of large strain behaviour of a granular material under biaxial compression. *Chaos*, 2013. In press.

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

[23] G. Sugihara and R. May. Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series. *Nature*, 344:734–741, 1990.

[24] X. Sun, M. Small, Y. Zhao, and X. Xue. Characterizing system dynamics with a weighted and directed network constructed from time series data. *Chaos*, 24:024402, 2014.

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

[26] J. Theiler, S. Eubank, A. Longtin, B. Galdrikian, and J. D. Farmer. Testing for nonlinearity in time series: The method of surrogate data. *Physica D*, 58:77–94, 1992.

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

[28] 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.

[29] X.-K. Xu, J. Zhang, and M. Small. Superfamily phenomena and motifs of networks induced from time series. *Proc Natl Acad Sci USA*, 105:19601–19605, 2008.

[30] Y. Yang and H. Yang. Complex network-based time series analysis. *Physica A*, 387:1381–1386, 2008.

[31] J. Zhang, X. Luo, and M. Small. Detecting chaos in pseudoperiodic time series without embedding. *Physical Review E*, 73:016216, 2006.

[32] J. Zhang and M. Small. Complex network from pseudoperiodic time series: Topology versus dynamics. *Physical Review Letters*, 96:238701, 2006.

[33] J. Zhang, J. Sun, X. Luo, K. Zhang, T. Nakamura, and M. Small. Characterizing pseudoperiodic time series through the complex network approach. *Physica D*, 237:2856–2865, 2008.

[34] X.-S. Zhang, Y.-S. Zhu, N. V. Thakor, and Z.-Z. Wang. Detecting ventricular techycardia and fibrillation by complexity measure. *IEEE Transactions on Biomedical Engineering*, 46:548–555, 1999.