Surrogate Test for Pseudoperiodic Time Series Data

Michael Small,[1,2,*] Dejin Yu,[2] and Robert G. Harrison[2]

[1]*Department of Electronic and Information Engineering, Hong Kong Polytechnic University,
Hung Hom, Kowloon, Hong Kong, China*
[2]*Department of Physics, Heriot-Watt University, Riccarton, Edinburgh EH14 4AS, United Kingdom*
(Received 18 January 2001; revised manuscript received 24 May 2001; published 16 October 2001)

For time series exhibiting strong periodicities, standard (linear) surrogate methods are not useful. We describe a new algorithm that can test against the null hypothesis of a periodic orbit with uncorrelated noise. We demonstrate the application of this method to artificial data and experimental time series, including human electrocardiogram recordings during sinus rhythm and ventricular tachycardia.

The method of surrogate data [1] is widely applied to test the null hypothesis that an observed time series is a typical realization of the output of a specific class of dynamical systems. This method is widely used in the analysis of experimental time series and provides a powerful tool in the search for determinism in apparently stochastic data. However, the current surrogate techniques have very limited utility when applied to a time series with a strong pseudoperiodic behavior.

The surrogate algorithm we describe in this Letter generates *pseudoperiodic surrogates* (PPS). This method is based on the well-known local-linear modeling methods described by Mees [2] and Sugihara and May [3]. Previously, Small and Judd advocated [4] and implemented [5] nonlinear radial basis modeling routines [6,7] as a form of surrogate hypothesis testing. The method we describe here is simpler and tests a more specific null hypothesis. This method may be applied to test against the null hypothesis of a periodic orbit with uncorrelated noise in the very large number of experimental systems that exhibit pseudoperiodic behavior.

By contrast, the three most successful, and widely applied, algorithms test for membership of the class of (i) independent and identical distributed (IID) noise processes, (ii) linearly filtered noise processes, and (iii) static monotonic nonlinear transformation of linearly filtered noise processes [1,8]. For time series data exhibiting strong pseudoperiodic behavior, the null hypotheses of IID or colored noise are obviously false. Therefore, apart from serving as a "sanity check," these existing algorithms are of limited use for such data.

For a pseudoperiodic time series, Theiler and Rapp suggested an alternative algorithm [9]: cycle shuffled surrogates. Analogously to IID noise surrogates, cycle shuffled surrogates are produced by shuffling the individual cycles within a time series. Hence, intracycle dynamics are preserved but intercycle dynamics are not. However, even with this new approach Theiler noted spurious long term correlations in the autocorrelation plot [9] for cycle shuffled surrogates. Furthermore, if the peak or troughs do not occur at precisely the same values, surrogates generated by this method are not able to preserve

both stationarity and continuity. By shifting individual cycles vertically, individual cycles can be matched and continuity will be preserved. However, such a transformation introduces nonstationarity in the surrogates that is absent in the original.

Each of these techniques is commonly applied to corroborate (or dismiss) observed nonlinearity in an experimental time series. However, for cyclic (i.e., pseudoperiodic) data the current methods perform poorly—or are simply not applicable. The PPS algorithm we describe in this Letter offers an entirely new surrogate generation algorithm that tests the null hypothesis that an observed time series is consistent with an (uncorrelated) noise-driven periodic orbit. Alternative hypotheses include deterministic nonperiodic intercycle dynamics or a periodic orbit with correlated noise.

By using correlation dimensions [10], we test this new algorithm with data from the Rössler attractor in both periodic and chaotic regimes. Our results show that this algorithm is able to distinguish between a noisy periodic orbit (the Rössler system exhibiting stable period 6 dynamics) and the chaotic Rössler contaminated with dynamic noise. We then apply our method to data from a periodic system driven by white and colored noises. The surrogate test is able to distinguish between these two cases. This demonstrates that this algorithm is a test against the null hypothesis of only uncorrelated noise in the intercycle dynamics. Finally, we apply this algorithm to human electrocardiogram (ECG) data during sinus rhythm and ventricular tachycardia (VT). When applied to these data, our algorithm demonstrates that sinus rhythm and VT are not consistent with an uncorrelated noisy periodic orbit.

Let $\{x_t\}_{t=1}^{N}$ be a scalar time series of N observations. For *embedding dimension d_e* and *embedding lag τ* we reconstruct the underlying dynamics according to the Takens' embedding theorem [11]:

$$z_t = (x_t, x_{t-\tau}, x_{t-2\tau}, \ldots, x_{t-(d_e-1)\tau})$$

for $t = (d_e - 1)\tau + 1, \ldots, N$. For notational convenience let us reindex the embedded time series to be $\{z_t\}_{t=1}^{\tilde{N}}$, where $\tilde{N} = N - (d_e - 1)\tau$. To construct a

surrogate we employ the following scheme: (1) Randomly choose an initial condition s_1, where $s_1 \in \{z_t : t = 1, \ldots, \tilde{N}\}$. Set $i = 1$. (2) For s_i choose a neighbor $s_j \in \{z_t : t = 1, \ldots, \tilde{N}\}$ with probability

$$\text{Prob}(s_j = z_t) \propto \exp \frac{-\|z_t - s_i\|}{\rho}. \qquad (1)$$

The parameter ρ is the *noise radius* and will be discussed latter. (3) Set $s_{i+1} = s_{j+1}$ and then increment i. (4) Repeat this procedure from step (2) until $i > N$. The vector time series $\{s_t\}_{t=1}^N$ represents a stochastic trajectory on the attractor approximated by $\{z_t\}_{t=1}^{\tilde{N}}$ with the same underlying dynamics (contaminated by noise) as the original system. The scalar surrogate time series $\{x_t^{(s)}\}_{t=1}^N$ is the first component of successive values of s_t.

It is important to note that $\{s_t : t = 1, \ldots, N\}$ and $\{z_t : t = 1, \ldots, \tilde{N}\}$ approximate the same attractor, but the reconstruction

$$\{(x_t^{(s)}, x_{t-\tau}^{(s)}, \ldots, x_{t-(d_e-1)\tau}^{(s)}) : t = (d_e - 1)\tau + 1, \ldots, N\}$$

does not. The attractor represented by this reconstruction is only an approximation to the underlying attractor and the sets $\{s_t : t = 1, \ldots, N\}$ and $\{z_t : t = 1, \ldots, \tilde{N}\}$. An example of this process for chaotic Rössler data is illustrated in Fig. 1.

Surrogates constructed in this way follow approximately the same vector field as the original data, but are contaminated with dynamic noise in such a way that any existing fine dynamics are obliterated. This includes any deterministic nonperiodic intercycle dynamic behavior, pseudoperiodic chaos, or periodic dynamics with colored noise. The
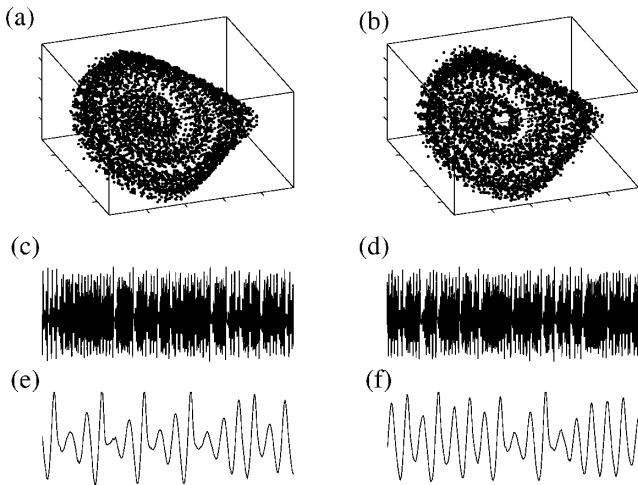

FIG. 1. Generation of PPS data from a chaotic Rössler time series. The top panels show the reconstructed attractors ($d_e = 3$, $\tau = 8$) for (a) the original data ($N = 5000$) and (b) the PPS data. Also shown is (c) the full original time series, (d) the full PPS time series, (e) a short section of the original, and (f) a short section of the PPS data. The surrogate was constructed with $d_e = 3$, $\tau = 8$, and $\rho = 0.005$. Note that the time series appear virtually indistinguishable. Correlation dimension estimates for the data and 30 PPS data sets are shown in Fig. 4.

important point is the suitable selection of the parameter ρ in (1). For large values of ρ the dynamics are poorly approximated. If ρ is extremely large ($\rho \rightarrow \infty$), then the points s_t are simply temporally uncorrelated random points in $\{z_t : t = 1, \ldots, \tilde{N}\}$ and the surrogate $\{x_t^{(s)}\}_{t=1}^N$ is equivalent to an ordinary IID surrogate (sampled with replacement rather than without). If ρ is too small ($\rho \rightarrow 0$) then the surrogate and original are identical. One must choose ρ so that the fine intercycle dynamics are obliterated but the intracycle dynamics are preserved.

We select ρ to be the value which maximizes the expected number of short segments (of length 2) that are the same for the original time series and a surrogate. This provides precisely the required balance between too much and too little noise. Figure 2 demonstrates the selection of ρ for the Rössler time series data.

The first of our applications of this algorithm is to the Rössler system. The Rössler equations are given by

$$\dot{x} = -(y + z),$$
$$\dot{y} = x + ay,$$
$$\dot{z} = 2 + z(x - 4).$$

For $a = 0.398$ this system exhibits broadband chaos, and for $a = 0.3909$ it exhibits period 6 behavior [12]. For these values of a we integrated this system for 5000 time steps of 0.2 time units. In each case we added Gaussian noise with a standard deviation of 0.05 to the x, y, and z components at each step. These time series, along with PPS data and reconstructed attractors, are illustrated in Figs. 1 and 3.

For each time series and 30 PPS data sets we estimated the correlation dimension according to the algorithm prescribed in [10] (Fig. 4). For the chaotic system, the data and surrogates are clearly distinct and in this case the null hypothesis of a periodic orbit with uncorrelated noise should be rejected. For the noisy period 6 system the data
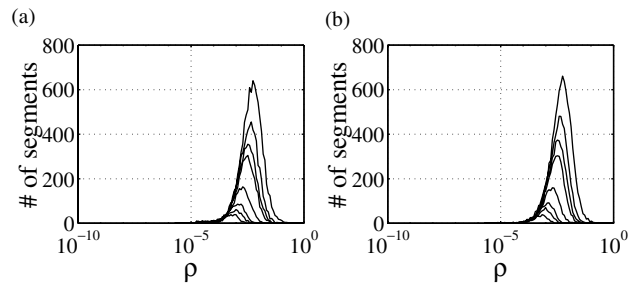

FIG. 2. Optimal selection of the parameter ρ. Panels (a) and (b) are the same calculation for chaotic and periodic Rössler systems, respectively. The time series are described in the text and shown in Figs. 1 and 3. In each case the plot shows a count of the number of times for which the data and a representative surrogate (of length 5000, $d_e = 3$, $\tau = 8$) are identical for more than n successive data points. The distinct curves in each plot are (from top to bottom) for $n = 2, 3, 4, 5, 10, 20, 30$, and 50. One clearly sees a distinct peak in each curve, suggesting an optimal value of ρ.
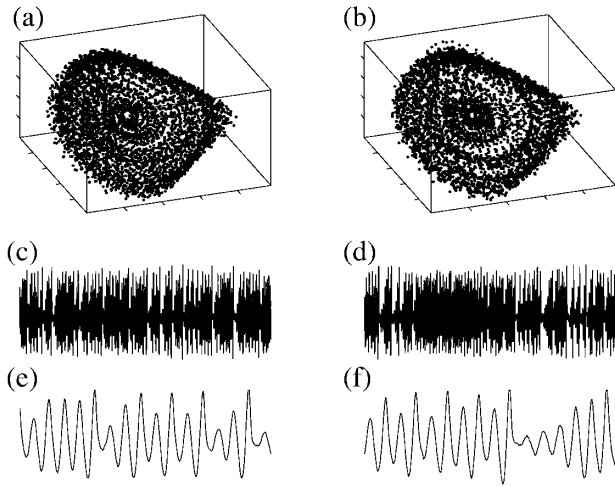
FIG. 3. Generation of PPS data from a noisy period 6 Rössler time series. The top panels show the reconstructed attractors ($d_e = 3$, $\tau = 8$) for (a) the original data and (b) the PPS data. Also shown is (c) the full original time series ($N = 5000$), (d) the full PPS time series, (e) a short section of the original, and (f) a short section of the PPS data. This surrogate was constructed with $d_e = 3$, $\tau = 8$, and $\rho = 0.0025$. A large range of values of ρ produced comparable results. Note that the time series and reconstructed attractors appear virtually indistinguishable. Correlation dimension estimates for the data and 30 PPS data sets confirm this and are shown in Fig. 4.

and surrogates are indistinguishable, and we are therefore unable to reject the null hypothesis. These results are consistent with what was expected.

We also tested this algorithm on periodic signals contaminated with various noise sources. Colored noise with a decorrelation time greater than p was added to a periodic signal of period p to produce a pseudoperiodic time series with a linear stochastic dependence of the intercycle intervals. The test signal was determined by the surrogate algorithm to be inconsistent with the null hypothesis. Indeed this is the simplest system that can lead to rejection of the null. Conversely, white noise added to a periodic signal was found to be consistent with the null.
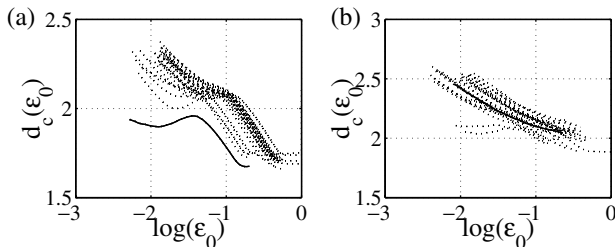

FIG. 4. Comparison of the correlation dimension for original and PPS data for the Rössler system. Panel (a) shows correlation dimension estimates ($d_e = 3$, $\tau = 8$) for the data shown in Fig 1 and 30 PPS data sets. Panel (b) is the same calculation for the data in Fig 3. For the chaotic time series the data and surrogates are clearly distinguishable; however, for the noisy periodic system there is no distinction.

Finally, we consider the application of the PPS algorithm to an experimental system. Figure 5 shows sections of two recordings of human ECG during sinus rhythm and VT. Each recording consists of 10 000 data points; the sampling rates are 50 Hz for sinus rhythm and 100 Hz for VT. Both time series were recorded with the experimental protocol described in [13] and are digitized at 10 bits. PPS data are also shown.

We repeated the calculation applied to the test data with these experimental recordings. For sinus rhythm the 30 PPS data sets were clearly distinguished from the ECG recordings. For ECG data recorded during VT the correlation dimension estimated for the data is distinguishable from the surrogates at the smallest length scales [$\log(\epsilon_0)$]. Hence, in both cases the null hypothesis of a periodic orbit with uncorrelated noise can be rejected [Figs. 5(a) and 5(b)].

The algorithm we have described in this Letter provides a robust method to test pseudoperiodic time series data against the null hypothesis of a periodic orbit with uncorrelated noise. Our results for data from the Rössler dynamical system, during periodic and chaotic regimes, and from a periodic system with either white or colored noise, demonstrate that this algorithm can differentiate between the underlying dynamics in these cases. In the case of the Rössler system data, the original time series are qualitatively similar in appearance, and uncovering the underlying dynamics directly would be difficult.

The issue of selection of ρ and for experimental data, d_e and τ has also been considered. The embedding
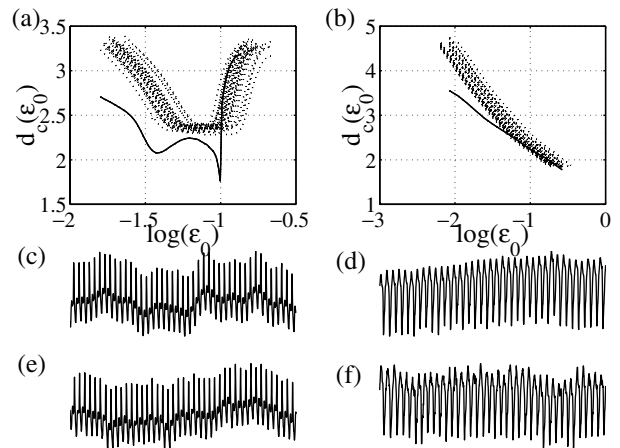

FIG. 5. PPS data calculation for human ECG data. Panel (a) is the correlation dimension estimates ($d_e = 6$, $\tau = 6$) for human ECG during sinus rhythm ($N = 10 000$) and 30 PPS data sets ($d_e = 6$, $\tau = 6$, $\rho = 0.005$). Panel (b) is the correlation dimension estimates ($d_e = 6$, $\tau = 6$) for human ECG during VT and 30 PPS data sets ($d_e = 6$, $\tau = 6$, $\rho = 0.0002$). Panel (c) shows a short section of the ECG during sinus rhythm (50 Hz) and panel (d) is a short section of the ECG during VT (100 Hz). Panels (e) and (f) are representative surrogate data sets for each. These surrogate tests indicate that human ECG during sinus rhythm and VT is not consistent with an uncorrelated noisy periodic orbit.

parameters d_e and τ must be selected so that the underlying dynamics are reconstructed by time delay embedding. This has been discussed in great detail elsewhere (see, for example, [14]). In this Letter, we provide a parameter-free method to select an appropriate value of ρ. Excessively large values destroyed the underlying dynamics, and led to temporally uncorrelated surrogates. Small values meant that the surrogates and data were virtually identical for large portions of their trajectories. By computing the effect of the variation of ρ on the expected frequency with which a surrogate is identical to the original data we were able to choose appropriate, intermediate values.

The dimension estimates shown in Figs. 4 and 5 provide additional corroboration that this algorithm is performing as proposed. For the data that led to the rejection of the null hypothesis, the dimension estimates for data and PPS time series differed most at small scale [small $\log(\epsilon_0)$]. This indicates that whereas at large scales the dynamics of data and PPS are identical for small scale they are significantly different. Therefore, the noise term (1) in the PPS algorithm has obliterated the fine structure present in the data.

Application of this algorithm to experimental data provided encouraging results. The experimental data used in this paper are highly noisy. The PPS time series is qualitatively very similar to the original data. The result that human ECG during both sinus rhythm and VT is consistent with deterministic nonperiodic intercycle dynamics is intriguing. This supports our observation of deterministic nonlinear dynamics during ventricular fibrillation [15,16]. The relatively subtle distinction for VT also supports our observation of comparatively low correlation dimension and entropy during VT [16]. However, it must be noted that the simplest explanation of this result is that the human ECG data are generated by a periodic system with correlated stochastic forcing.

*Corresponding author.
Email address: ensmall@polyu.edu.hk

[1] J. Theiler *et al.,* Physica (Amsterdam) **58D**, 77 (1992).
[2] A. I. Mees, Int. J. Bifurcation Chaos Appl. Sci. Eng. **1**, 777 (1991).
[3] G. Sugihara and R. M. May, Nature (London) **344**, 737 (1990).
[4] M. Small and K. Judd, Physica (Amsterdam) **120D**, 386 (1998).
[5] M. Small and K. Judd, Int. J. Bifurcation Chaos Appl. Sci. Eng. **8**, 1231 (1998).
[6] K. Judd and A. Mees, Physica (Amsterdam) **82D**, 426 (1995).
[7] M. Small and K. Judd, Physica (Amsterdam) **117D**, 283 (1998).
[8] J. Theiler and D. Prichard, Physica (Amsterdam) **94D**, 221 (1996).
[9] J. Theiler and P. Rapp, Electroencephalogr. Clin. Neurophysiol. **98**, 213 (1996).
[10] K. Judd, Physica (Amsterdam) **56D**, 216 (1992); **71D**, 421 (1994).
[11] F. Takens, Lect. Notes Math. **898**, 366 (1981).
[12] J. M. T. Thompson and H. B. Stewart, *Nonlinear Dynamics and Chaos* (Wiley, Chichester, 1986).
[13] M. Small *et al.,* Comput. Cardiol. **27**, 355 (2000).
[14] H. D. I. Abarbanel, *Analysis of Observed Chaotic Data* (Springer-Verlag, New York, 1996).
[15] M. Small *et al.,* Chaos **10**, 268 (2000).
[16] M. Small *et al.,* Chaos, Solitons and Fractals (to be published).