Constrained surrogate time series with preservation of the mean and variance structure

C. J. Keylock
Earth and Biosphere Institute, University of Leeds, Woodhouse Lane, Leeds, LS2 9JT, United Kingdom
(Received 6 August 2005; published 21 March 2006)

A method is presented for generating surrogates that are constrained realizations of a time series but which preserve the local mean and variance of the original signal. The method is based on the popular iterated amplitude adjusted Fourier transform method but makes use of a wavelet transform to constrain behavior in the time domain. Using this method it is possible to test for local changes in the nonlinear properties of the signal. We present an example for a change in Hurst exponent in a time series produced by fractional Brownian motion.

DOI: 10.1103/PhysRevE.73.036707 PACS number(s): 02.50.Ey, 05.45.Tp, 05.40.Ca, 89.75.−k

Distinguishing between linear and nonlinear processes from a time series of system outputs plays an integral role in the detection of deterministic chaos in physical systems. One approach is to formulate a null hypothesis for a specific class of processes and compare the system output to this hypothesis. The surrogate data method is a popular way to establish such a null hypothesis [1] and can be undertaken in two distinct ways: *Typical realizations* are Monte Carlo generated surrogates from a model that provides a good fit to the data; *constrained realizations* are surrogates generated from the data values to conform to certain properties of the data (such as its autocorrelative structure). This latter approach is more suitable for hypothesis testing as it does not require the definition of a pivotal test statistic [2]. In order to test the null hypothesis at a level of significance a with a two-tailed test, one can generate $2/a-1$ surrogates which, together with the original data leads to an appropriate number of realizations of the data. If all the surrogates are either greater or less than the original data, the null hypothesis may be rejected at the level a.

An early method for generating constrained realizations conforming to a linear Gaussian, stochastic process [1] that has subsequently seen several applications [3] has been termed the amplitude adjusted Fourier transform (AAFT) method [4]. An AAFT surrogate for a time series $\{x_i\}$, $i=1,\ldots,N$ is found by generating white-noise data $\{w_i\}$ and manipulating it so that the sequential order of the ranked values matches that for $\{x_j\}$. The phases of this ordered white noise sequence are then randomized in the Fourier domain and again rank ordered with respect to the original data. Some limitations with this method due to finite sample sizes and preservation of the Gaussian distribution [5] led [6] to propose an enhanced method known as the iterated AAFT (IAAFT), which has proven popular [7]. The squared amplitudes of $\{x_i\}$ are stored and then a random sort of the values in $\{x_i\}$ takes place. The Fourier transform of the random sort is taken and the squared amplitudes are replaced by those for $\{x_i\}$, with the complex phases retained. The Fourier transform is then inverted and rank ordering is used to map values in this surrogate series to those in $\{x_i\}$. The modification to the spectral behavior that results from this rank-ordering adjustment is dealt with by iterating this procedure until no further reordering occurs. The IAAFT method's popularity is due to its elegance, computational efficiency and ability to discriminate effectively between linear and nonlinear phenomena [4]. In a subsequent paper, a more general approach for constructing constrained surrogate data was proposed [8]. However, the computational time associated with the simulated annealing procedure for this algorithm has meant that it has not been as popular as the original IAAFT method.

More recently, a method for producing typical realizations has been proposed for cyclic data based on time-delay embedding [9]. This is known as the pseudoperiodic surrogates (PPS) method and for this case, the appropriate null hypothesis is that the data have been produced by periodic behavior driven by white noise. Although the surrogates are only typical realizations, the authors argue that their correlation dimension test statistic [10] is pivotal for the hypothesis of periodic behavior and the method seems to work well for this hypothesis. However, as shown in Fig. 1, this approach cannot be used successfully for testing the more common null hypothesis that the data are from a stationary, linear process, where nonlinearity is assessed using the asymmetry [2] $A(\lambda)=\langle(x_i-x_{i-\lambda})^3\rangle/\langle(x_i-x_{i-\lambda})^2\rangle^{3/2}$ where the angled braces indicate ensemble averaging. The (IAAFT) and our proposed algorithms both accept the null hypothesis for the stationary case and reject it at the 5% significance level for the nonlinear process. These results are not surprising because the PPS method is designed to test a different null hypothesis. Because it generates typical realizations, the PPS algorithm is also not constrained to the values in the original dataset and gives a weaker match to the original spectrum than the other methods as is clear in Fig. 2. In particular, note that at low frequencies there has been a significant enhancement of the energy for the two examples based on the Rössler attractor using the PPS method. This problem has been previously noted for AAFT surrogates [6].

Our approach to surrogate generation uses a wavelet transform to preserve local values in the time domain. It has been previously proposed to use wavelets to generate surrogate data [11] but that method was based on a randomization of the wavelet detail coefficients at a particular dyadic scale. The potential advantage of the wavelet approach is that behavior in the time-frequency plane can be preserved. Hence randomization of the coefficients on the time axis offers no clear advantage over the IAAFT, at least for univariate time

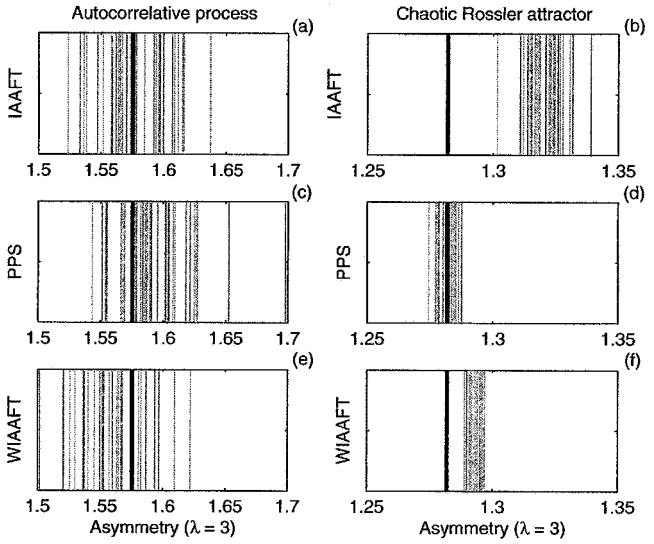

FIG. 1. The asymmetry measure $A(\lambda=3)$ for a second-order autocorrelative process (a), (c), and (e) and a chaotic path from the Rössler attractor (b), (d), and (f). In each case the long black line records the asymmetry for the original data and the short gray lines the values for 39 surrogates. The IAAFT surrogates are displayed in (a) and (b), (PPS) surrogates ($d_e=3$, $r=8$, $a=0.005$) in (c) and (d), and surrogates generated with our wavelet-based variant of IAAFT (WIAAFT) in (e) and (f). The autocorrelative process is given by $x_i=0.8x_{i-1}-0.25x_{i-2}+0.2\varepsilon$ where ε is a zero mean, unit variance Gaussian distribution. The Rössler equations are $\dot{q}=-(r+s), \dot{r}=q+\alpha r, \dot{s}=2+t(q-4)$ where, following Ref. [9], we choose $\alpha=0.398$ to obtain a chaotic response. These equations were integrated for 20 480 time steps of 0.1 units and values for r extracted. The first 10 240 values were then discarded and one in every ten of the remainder were regularly sampled.

series. Our method for generating surrogates is wavelet based, but overcomes this deficiency and preserves the patterns in the mean and variance while randomizing nonlinear properties of the signal such as the Hurst exponent.

Our algorithm for generating a single surrogate time series proceeds as follows:

(i) Take the stationary or maximal overlap discrete wavelet transform MODWT of the signal over dyadic scales $2^{j-1}j=1,\ldots,J$. This will produce J sets of detail coefficients $\{D_i^j\}$ each containing N values (because the MODWT is an undecimated transform) and a set of approximation coefficients $\{A_i^J\}$ representing the unresolved scales. In this study we have analyzed signals with a length 2^J so that all information (beyond a constant) is contained in the detail coefficients. The MODWT has the additional useful properties that it is well defined for any N (not restricted to a multiple of 2^J), and produces detail coefficients and spectra unaffected by circularly shifting the data (i.e., the coefficients are robust to the time at which one breaks into the signal). The variance of the detail coefficients at each level is equivalent to the Fourier spectrum of the signal [12]. The choice of wavelet basis function affects the time-frequency properties of the wavelet decomposition. In order to deal with nonstationary signals it is advantageous to use a wavelet with a high number of vanishing moments [12]. The results in Figs. 1–4 all use a Daubechies wavelet with 16 vanishing moments [13].

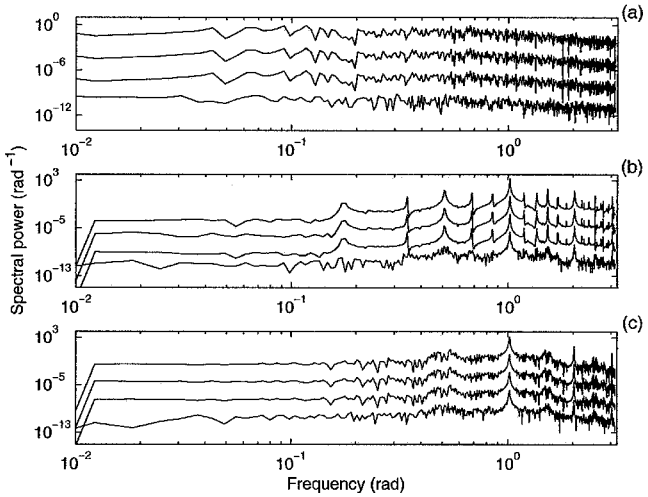

FIG. 2. Power spectra for a second order autocorrelative process (a), a path on a periodic Rössler attractor (b), and a chaotic Rössler attractor (c). Each plot shows four spectra displaced from each other by 10^3 for clarity. The upper is the spectrum for the original data, the second is an IAAFT surrogate, the third is for a surrogate produced using our algorithm and the bottom for a PPS surrogate. The data in the middle figure were obtained from the Rössler equations in the same way as described in Fig. 1 except for α, which was altered to 0.3909 to give period 6 behavior [9].

(ii) Obtain a constrained realization of the detail coefficients at each level by applying the IAAFT algorithm (i.e., apply the IAAFT as if each set of detail coefficients was a time series in its own right) to yield surrogate detail coefficients $\{D_i^j(\text{surr})\}$, $i=1,\ldots,N; j=1,\ldots,J$. Because this is a constrained realization method the values of the detail coefficients are preserved. Because the frequency behavior of these coefficients is retained, this method eliminates the ran-

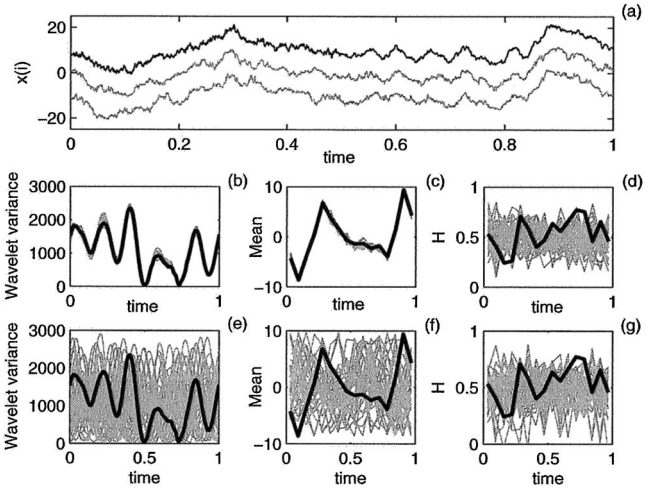

FIG. 3. Time series and surrogate analysis for fractional Brownian motion with a change in H from 0.4 to 0.65 at the mid point. The dimensionless time series (black) and two surrogates using the new algorithm (gray) are shown in the top panel, with vertical displacements to aid visibility. The second row shows properties of the original signal (black) and surrogates (gray) for our method and the third row shows similar information for IAAFT surrogates.

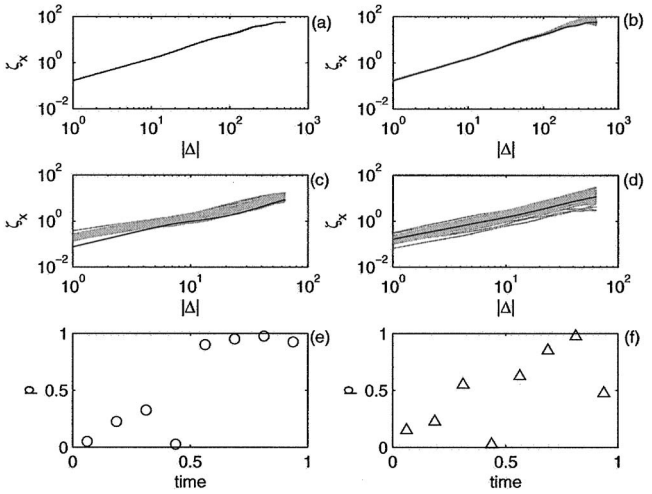

FIG. 4. Hurst exponent properties for the time series in Fig. 3. The plot on the left is for our surrogates and that on the right is for the IAAFT method. (a) and (b) a log-log plot of $\zeta_x=\langle|x(t)-x(t-\Delta)|^2\rangle$ vs $|\Delta|$ over the whole time series for $\Delta=1,\ldots.512$. Hence from Eq. (1) the gradient is $2H$. (c) and (d) have the same axes but the calculation is undertaken over the last 256 values in the time series for $\Delta=1,\ldots,64$ (e) and (f) give the probability of obtaining a particular value for H for the data (estimated from the slope of the line from plots similar to the above figure) as a function of the values for H of the 39 surrogates based on eight blocks of 256 values.

domization problem associated with Ref. [11]. We then have a surrogate for the detail coefficients at a particular level. In addition, we produce the mirror image of this surrogate $\{D_i^j(\text{mirror})\}=\{D_k^j(\text{surr})\}$, $i=1,\ldots,N;k=N,N-1,\ldots,1$.

(iii) Match the surrogate and its mirror image to the original detail coefficients at a particular level by circularly rotating the values until an appropriate error function is minimized. Retain either the surrogate or the mirror image according to the lowest value for the error function. This locates the energy peaks at the appropriate locations in the signal, preserving the temporal structure. We have found that there is little advantage to using more than a simple least-squares function at this stage.

(iv) Perform the inverse MODWT on the chosen surrogate detail coefficients for each level and the original approximation coefficients to produce a surrogate of the original time series and use the rank-ordering method from the standard IAAFT algorithm to recover the values of the original time series.

(v) As with IAAFT algorithm, the rank ordering has degraded the accuracy of our spectral representation of the signal. Hence we require an iterative procedure to obtain convergence. We have found it best to implement these subsequent iterations in exactly the same manner as the IAAFT due to the relative speeds of the wavelet and Fourier transforms and the degree to which our wavelet-based first stage to the algorithm approximates a local minimum.

Figures 1 and 2 show that our algorithm is able to generate surrogates that discriminate between linear and nonlinear processes and provide as good a match to the power spectrum of the data as other algorithms. Figure 3 shows properties of the surrogates for a time series consisting of 2048 values for fractional Brownian motion with a Hurst exponent (H) of 0.4 for the first half of the record and a change to $H=0.65$ for the second half. Figure 3(a) illustrates the localization properties of our algorithm The surrogates (in gray) retain the positions of the major maxima and minima of the original series (black). Hence they have the nice property that they "look like" the original data. The next two rows of Fig. 3 show various properties of both the signal and 39 surrogates generated by our method (middle) and the IAAFT (bottom). Our algorithm preserves the local mean and variance structure while randomizing the nonlinear aspects of the signal. The latter is assessed by estimating H using a second derivative method [14] for 16 blocks of 128 consecutive values. The mean values were determined for the same blocks, while the wavelet variance was calculated using Ref. [15].

The advantage of producing surrogates that "look like" the data is not merely aesthetic. This property allows us to formulate further hypotheses that cannot be tackled successfully with standard IAAFT surrogates. For example, consider the case of fractional Brownian motion B defined by

$$B(0)=0,$$

$$\langle|B(t)-B(t-\Delta)|^2\rangle=\sigma^2|\Delta|^{2H}. \quad (1)$$

If one considers the whole of the time series in Fig. 3 then the surrogates from IAAFT and our algorithm yield a similar average H (Fig. 4). However, specific segments of IAAFT surrogates do not preserve the local value for σ^2, meaning that variability in both the local variance and the time series increments will contribute to the local estimate of H. Figure 3 shows that this will not be the case for our surrogates, meaning that, from Eq. (1), the variability in H for the surrogates is merely due to differences in the increments. This provides a means for detecting changes in H as the surrogates will retain an average measure of H (defined over the whole time series) while the data will reflect the local value. Figures 4(c) and 4(d) show example estimates of H for the final 256 values of the time series in Fig. 3 determined for $\Delta=1,\ldots,64$, and Figs. 4(e) and 4(f) give the probability of H based on the data plus surrogates for all eight windows of 256 values across the data set. The transition between the values for the first and last halves of the data is clear in the case of our surrogates, but much more confused for IAAFT surrogates. Hence our surrogates have correctly found the transition from $H=0.4$ to $H=0.65$. The higher value for H for the data compared to the surrogates for the final increment is clear from Fig. 4(c).

Given the range of phenomena that may be described by fractional Brownian motion or related fractal noises that are stationary in the increments, the surrogate algorithm proposed in this paper may have application in a number of areas where one is concerned with changes in the properties of a signal through time. Our conceptually simple alteration to the highly effective IAAFT approach permits this surrogate generation method to be extended to consider a wider class of null hypotheses.

[1] J. Theiler, S. Eubank, A. Longtin, B. Galdrikian, and J. D. Farmer, Physica D **58D**, 77 (1992).
[2] J. Theiler and D. Prichard, Physica D **94D**, 221 (1996).
[3] D. Prichard and C. P. Price, Geophys. Res. Lett. **20**, 2817 (1993); J. Theiler and P. E. Rapp, Electroencephalogr. Clin. Neurophysiol. **98**, 213 (1996).
[4] D. Kugiumtzis, Phys. Rev. E **60**, 2808 (1999).
[5] D. K. Ivanov, H. A. Pösch, and C. Stumpf, Chaos **6**, 243 (1996).
[6] T. Schreiber, A. Schmitz, Phys. Rev. Lett. **77**, 635 (1996).
[7] T. Schreiber, A. Schmitz, Physica D **142D**, 346 (2000); K. Lehnertz, *et al.*, J. Clin. Neurophysiol. **18**, 209 (2001); C. Rieke, *et al.*, IEEE Trans. Biomed. Eng. **50**, 634 (2003); D. Poggi, *et al.*, Geophys. Res. Lett. **31**, L05102 (2004).
[8] T. Schreiber, Phys. Rev. Lett. **80**, 2105 (1998).
[9] M. Small, D. Yu, and R. G. Harrison, Phys. Rev. Lett. **87**, 188101 (2001); M. Small and C. K. Tse, Physica D **164D**, 187 (2002).
[10] K. Judd, Physica D **56D**, 216 (l992); **71D**, 421 (1994); M. Small and K. Judd, *ibid.* **120D**, 386 (1998).
[11] M. Breakspear, M. Brammer, and P. A. Robinson, Physica D **182D**, 1 (2003).
[12] D. B. Percival and A. T. Walden, *Wavelet Methods for Times Series Analysis* (Cambridge University Press, Cambridge England, 2000).
[13] I. Daubechies, *Ten Lectures on Wavelets* (SlAM, 1992).
[14] J. Istas and G. Lang, Ann. I.H.P. Phys. Theor. **33**, 407 (1994)
[15] B. Whitcher, S. D. Byers, P. Guttorp, and D. B. Percival, Water Resour. Res. **42**, 1054 (2002). Algorithms for implementing this method can be found at http://lib.stat.cmu.edu