PHYSICAL REVIEW E **72**, 055201(R) (2005)

Testing for nonlinearity in time series without the Fourier transform

Tomomichi Nakamura,* Xiaodong Luo, and Michael Small
Department of Electronic and Information Engineering, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong

(Received 24 May 2005; revised manuscript received 19 August 2005; published 18 November 2005)

A method to test for nonlinearity in time series, without the need to apply the Fourier transform, is proposed. This method therefore avoids the drawbacks of previously proposed surrogate techniques associated with the estimation of the signal's power spectrum. The test addressed by this algorithm is that the data are generated by a stationary linear system. To achieve this, the algorithm takes advantage of the fundamentally different structure of linear and nonlinear systems.

DOI: 10.1103/PhysRevE.72.055201 PACS number(s): 05.45.Tp, 02.50.−r, 05.10.−a

The method of surrogate data has been proposed by Theiler *et al.* [1] to investigate nonlinear determinism in a time series. The most successful algorithms are as follows: (i) the Fourier transform (FT) algorithm and (ii) the amplitude adjusted Fourier transform (AAFT) algorithm. In an attempt to address some of the problems of these techniques, an improved surrogate method, iterative AAFT (IAAFT), has also been proposed by Schreiber *et al.* [2]. All of these techniques are *linear surrogate methods* because they are based on a linear process and address a linear null hypothesis.

It should be noted that a time series exhibiting strong periodicities is clearly not consistent with the hypothesis of linear noise [3]. Hence, in this Rapid Communication, our focus is only on a time series exhibiting irregular fluctuations: a time series which may be generated by either a stochastic linear process or a deterministic nonlinear dynamical system. The purpose of our method is to distinguish between linear noise and deterministic dynamics. The test addressed by our algorithm is that the data are generated by a stationary linear system.

Recently, there has been considerable concern raised about the efficacy of the standard linear surrogate techniques. Schreiber demonstrated that the power spectra of AAFT surrogates may not be identical to the original data [2,4], and Theiler observed that wraparound effects of the Fourier transform may lead to spurious high frequency content in the surrogates [5]. Kugimitzis identified further problems with these techniques and proposed palliative measures [6]. In this Rapid Communication, we introduce a method to test for nonlinearity without applying the Fourier transform unlike linear surrogate methods, thereby avoiding many of these problems entirely.

After describing our technique, we will present some numerical examples of the application of this algorithm to simulated time series data: a linear AR model and nonlinear transformations of a linear AR model, which are both linear systems; the Henon map and a chaotic neural network (CNN), which are both nonlinear systems. In each case, the data we will use are both noise free and contaminated by 20 dB Gaussian observational noise. Based on the results, we apply the proposed method to two actual data, NMR laser and cobalt data.

The basic premise of this technique is that the sum of independent realizations of a linear system is a new realization of the same linear system. But, this superposition principle is valid only for linear systems. The sum of independent realizations of a nonlinear deterministic system will exhibit a different dynamical behavior. In the following paragraphs we present our choice of discriminating statistics and then describe the proposed algorithm in more detail.

Dynamical measures are often used as discriminating statistics in hypothesis testing. We use the correlation dimension (CD) [2]. We note that when a system is linear stochastic, although the true CD should be infinite, the estimated CD can still provide useful information to understand some features of dynamics. Hence, we employ it as our discriminating statistics. To estimate the CD we apply the Gaussian kernel (GK) algorithm [7]. The GK method is more robust against noise and can give a unique value of dimension by extending the hard kernel function (or the Heaviside function) in the calculation of correlation integral to the general kernel functions. More details concerning CD estimates and the relevant problems are presented elsewhere [7,8].

To estimate the CD, it is necessary to first reconstruct the underlying deterministic attractor. For this purpose, a time delay embedding reconstruction is usually applied [9]. If our observed data comprises a set of N scalar measurements $\{s_t\}_{t=1}^{N}$, we can construct vectors $x_t=(s_t, s_{t-\ell}, \ldots, s_{t-(d-1)\ell}) \in \mathbf{R}^d$, where ℓ is a time-lag and d is an embedding dimension, and we call these *embedding parameters*. Assuming that the original attractor lies on a k-dimensional compact manifold, Takens proved that the transformation is an embedding if $d>2k$ [9]. It is well known that the proper estimation of invariant measures of the underlying system are obtained only for a suitable range of the embedding parameters. Many authors have considered the details of the very important problem of finding appropriate embedding parameters elsewhere; however, the issue is still unresolved [10]. To avoid possible confusion with problems arising from inappropriate estimation of embedding parameters, we use the "correct" time-lag and a large enough embedding dimension in all examples [11].

After the calculation of suitable test statistics, we need to determine whether the hypothesis addressed by this method shall be rejected or not. We employ Monte Carlo hypothesis testing [12]. The basic idea is the following. Let a discriminating statistic for the original data be Q_0 and let the statistic for test data sets be Q_i $(i=1,2,\ldots,N_s)$, where N_s is the number of test data sets. We inspect whether the estimated statistic of the original data falls within or outside the statistic distribution of the test data. When the statistic falls within

*Electronic address: entomo@eie.polyu.edu.hk

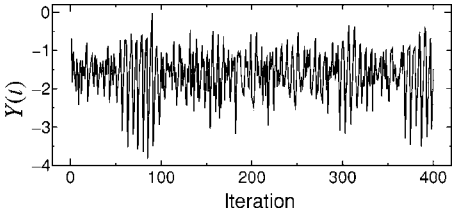

FIG. 1. The macroscopic mean activity of the CNN.

the distribution of the test data, we consider that the hypothesis may not be rejected.

The primary contribution of this Rapid Communication is an alternative method of producing data sets which do not require estimation of the power spectrum of the given data. To achieve our premise, we generate "test data" (see below) by adding two intervals. The algorithm is superficially similar to a moving average filter. Let the original data be $s(t)$ and the test data be $z(t)$. A data set $\hat{s}(t)$ is first constructed by

$$\hat{s}(t) = \alpha s(t) + (1 - \alpha)s(t + \tau), \tag{1}$$

where $0.4 < \alpha < 0.6$ and τ is called "shift." If τ is large enough so that $s(t)$ and $s(t+\tau)$ are (linearly) uncorrelated, then $\hat{s}(t)$ will have the same linear characteristics as $s(t)$. The data set $\hat{s}(t)$ is therefore equivalent to a FT surrogate of $s(t)$. The data $\hat{s}(t)$ do not have the same empirical distribution as the original data. To ensure that the test data have the same distribution as the original (and avoid any statistical bias) we must reorder it. Hence, to achieve the same distribution like an AAFT or IAAFT surrogate, we reorder $s(t)$ to have the same order as $\hat{s}(t)$ and the reordered data are the test data $z(t)$.

We note here that the proposed method is not strictly a surrogate test, although the purpose of the proposed method

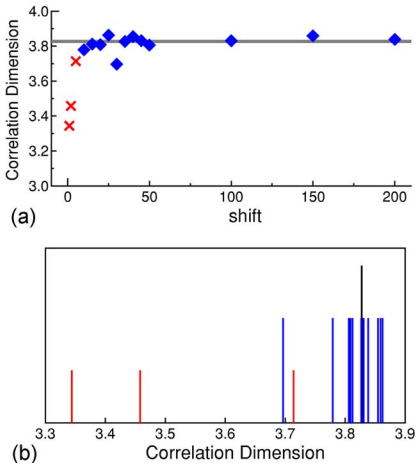

FIG. 2. (Color online) A plot of the correlation dimension (CD), where the model is the NLT of a linear AR model, ℓ is 1 and d is 5. (a) relation between estimated CD and shift, and (b) CD of the original data and the distribution of the test data. The horizontal line in panel (a) and longer line in panel (b) correspond to the CD of the original data, the $\times$ and the short lines when τ is not greater than the embedding window (that is, 1, 2, and 5), and the $\diamond$ and medium lines when τ is larger than the embedding window (that is, 10, 15, 20, 25, 30, 35, 40, 45, 50, 100, 150, and 200).

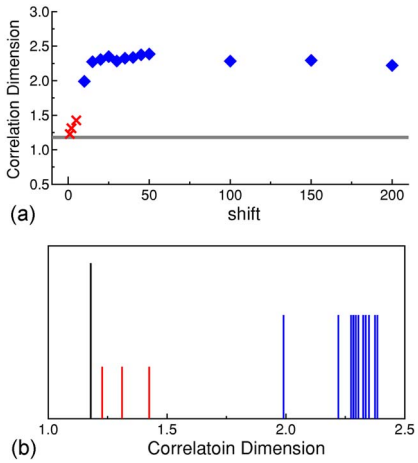

FIG. 3. (Color online) A plot of the CD, where the model is the Henon map, ℓ is 1, and d is 5. The notation is the same as in Fig. 2.

is the same as that of linear surrogate methods. The essential feature of the methodology of surrogate data is that generated surrogates preserve certain properties of the data and destroy others and are consistent with a specified null hypothesis. However, our proposed algorithm simply takes advantage of the fundamentally different structure of linear and nonlinear systems to investigate nonlinearity in the time series. Hence, we refer to data generated by the proposed algorithm not as "surrogate data" but as "test data."

We may demonstrate the rationale for this algorithm as follows. Suppose that the underlying system is linear, then both $s(t)$ and $s(t+\tau)$ are realizations of that linear process. If $s(t)$ and $s(t+\tau)$ are linearly uncorrelated (that is, τ is large enough), $\alpha s(t)+(1-\alpha)s(t+\tau)$ is also a realization of the underlying linear process [13]. Hence, the CD for $s(t)$ and $\alpha s(t)+(1-\alpha)s(t+\tau)$ will be the same. Conversely, suppose that the underlying system is deterministic (additional additive or dynamic noise is also admissible) and inherently nonlinear. For small values of τ, the substitution $s(t) \to \alpha s(t) + (1-\alpha)s(t+\tau)$ is nothing more than a change of coordinates:

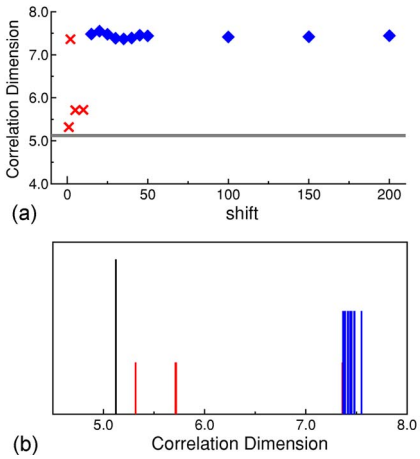

FIG. 4. (Color online) A plot of the CD, where the model is the CNN, ℓ is 1, and d is 10. The horizontal line in (a) and the longer line in (b) correspond to the CD of the original data, the $\times$ and the short lines when τ is 1, 2, 5, and 10, and the $\diamond$ and medium lines when τ is larger than 10.

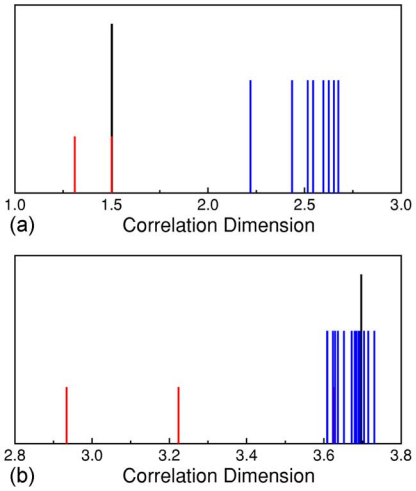

FIG. 5. (Color online) A plot of the CD, (a) NMR laser data and (b) cobalt data, where ℓ is 1 and d is 5. The longer line corresponds to the CD of the original data, the short lines when $\tau \leqslant 5$, and the medium lines when $\tau > 5$.

equivalent to smoothing out observational noise and identical to the distinction between a time-lag and a differential embedding. Hence, for small τ, we expect that the CD of the test data is preserved [14]. As τ increases, $s(t)$ and $s(t+\tau)$ become uncorrelated. In particular, for chaotic systems this will happen fairly rapidly. When $s(t)$ and $s(t+\tau)$ are uncorrelated, the dynamics of test data $\alpha s(t)+(1-\alpha)s(t+\tau)$ is equivalent to observing a different, more complex, dynamical system.

The critical point for this algorithm is therefore choosing τ such that $s(t)$ and $s(t+\tau)$ become linearly uncorrelated. In the most precise sense, this may be difficult to achieve, even if the autocorrelation function is almost zero. To avoid this difficulty, we change the number of the shifts step by step, for example, 1, 2, 5, 10, 20, and so on [15]. By doing this, independent test data are generated at each shift. Also, multiple random test data sets for a fixed shift can be generated by randomizing α. In either case, by gradually increasing the shift, for a linear system we expect the test and original data to converge. For nonlinear systems we expect them to become clearly distinct.

In practice, when applying this idea, we must also be careful that the shift should be larger than the maximum time-lag in the embedding space (the embedding window). If this is not the case, certain more complicated correlations appear in the embedding space and the preceding argument does not hold. For example, when the time-lag ℓ is 1 and the embedding dimension d is 5, the vectors constructed are $(t, t-1, \ldots, t-4)$. When the shift τ is 10, there would be no strong correlation between each component of successive embedded vectors. However, when ℓ is 1 and d is 15, the constructed vectors are $(t, t-1, \ldots, t-14)$. Hence, when τ is 10 (that is, the shift is smaller than the embedding window), some kind of correlation will appear between original and shifted embedded vectors. That is, the use of the shift to form the test data increases the embedding window and forms a different embedding. However, when a system is linear, the test data will have the same linear system when the embedding window is smaller than the shift used. On the other hand, when a system is nonlinear, the test data will no longer have the same nonlinear dynamics, even if the embedding window is smaller than the shift used.

We now demonstrate the application of our algorithm, and confirm our theoretical arguments with several models (two linear and two nonlinear): a linear AR model, nonlinear transformations (NLT) of a linear AR model data, the Henon map; and CNN. In each case the number of data points is 2048; the data used are both noise free and contaminated by 20 dB Gaussian observational noise. We use $\alpha=0.5$ [16] and τ is 1, 2, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 100, 150, and 200. The linear AR model is given by $x_t=A_0+A_1x_{t-1}+A_6x_{t-6}+\eta_t$, where $A_0=2.945\,206$, $A_1=0.300\,739$, and $A_6=0.202\,056$ [17], η_t is the Gaussian dynamical noise with standard deviation 1.0. We use x_t as the observational data. For the NLT of a linear AR model, the same linear AR model described above is used and a static monotonic nonlinear function $h(x)$ is used [18]. The function is

$$h(x) = \frac{\left[\frac{x-x_{min}-0.0001}{x_{max}-x+0.001}\right]^{\rho}}{1+\left[\frac{x-x_{min}-0.0001}{x_{max}-x+0.001}\right]^{\rho}}, \tag{2}$$

where x_{min} and x_{max} are the minimum and maximum value of x in the original time series, and $\rho=3$. The Henon map [19], as a second order difference equation, is given by $x_t=1.0+0.3x_{t-2}-1.4x_{t-1}^2$. We use x_t as the observational data.

Many measured physical quantities can be seen as an average derived from subsystems or microsystems. However, in the scalar time series analysis it is assumed to be a single probe of the system being investigated. Also, the averaging operation is one of the filters by which the mean data would become less chaotic. Hence, for investigating such a more complex and practical case, we use a CNN [20], where we use ten neurons to compose the network [21]. Let the ith chaotic neuron be $y_i(t)$. Then, the ensemble mean value $Y(t)$ of ten neurons is defined as $Y(t)=\frac{1}{10}\Sigma_{i=1}^{10}y_i(t)$, which can be regarded as a simple model of electroencephalography (EEG) data. We use $Y(t)$ as the observational data. Figure 1 shows typical behavior of the time series. It should be noted that we have confirmed that linear surrogate data methods can discriminate the above systems correctly using the CD but only provided that the embedding parameters are estimated appropriately.

Figure 2 shows the results when the model is the NLT of a linear AR model and the data is noise free, where the time lag ℓ is 1 and embedding dimension d is 5. Figure 2(a) shows that as τ increases, the CD values of the test data converge and these are almost identical to that of the original data. We apply Monte Carlo hypothesis testing [12] to the larger shift ($\tau > 5$) because these shifts are larger than the embedding window and we expect that the two intervals are linearly uncorrelated. Figure 2(b) shows that the CD of the original data falls within the distribution of the test data. According to the criterion mentioned previously, we cannot reject our hypothesis. This result indicates that if there is no correlation between $s(t)$ and $s(t+\tau)$ when using a τ, the test data have almost the same structure even if a larger shift is used. When the linear AR model is used, the data are con-

taminated by 20 dB observational noise, and other embedding dimensions are used; the results show essentially the same behavior as Fig. 2.

Figure 3 shows the results when the model is the Henon map and the data are noise free, where ℓ is 1 and d is 5. This result shows the opposite behavior as a previously presented result. Figure 3(a) shows that the CD of the test data is almost the same as that of the original data when the shift τ is 1, 2, and 5, which is the same behavior reported in [14]. However, Figs. 3(a) and 3(b) show that these values are no longer the same and the CD of the original data falls outside the distribution when $\tau > 5$. When the data are contaminated by 20 dB observational noise and other embedding dimensions are used, essentially the same behavior occurs. In all cases, from the results of these two examples, we find that if the system is linear, the CD of the original data falls within the distribution of the test data, and if the system is nonlinear, the CD of the original data falls outside the distribution of the test data.

Figure 4 shows the results when the model is the CNN described previously, and the data are noise free, where ℓ is 1 and d is 10. When $\tau > 10$, Fig. 4(a) shows that the CD of the original data is different from that of the test data and Fig. 4(b) shows that it falls outside the distribution. Although the Lyapunov dimension of this CNN is 1.043 032 [21], the estimated CD of the original data is about 5.1. As is obvious, it is difficult to estimate CD correctly in every situation. The possible reason why the estimated CD is much different is that the data used are the ensemble mean data of ten neurons. That is, the averaging operation is one of the filters by which information in the data would be lost. However, Figs. 4(a) and 4(b) clearly show that the CD of the original data is different from that of the test data when $\tau > 10$. Hence, we conclude that our method works well. When the data are contaminated by 20 dB observational noise and other embedding dimensions are used, the behavior is essentially the same as the results shown here.

The above results indicate the effect of different features of the reconstructed attractor for linear and nonlinear dynamics. From the results, we expect that if the system is linear, when τ is large enough and there is no correlation between two intervals in Eq. (1), the distribution of the CD of test data includes that of the original data. However, if the system is not linear, even when τ is large, the distribution does not include that of the original data. In these examples, the proposed algorithm works well for testing linear and nonlinear systems and also distinguishing linear from chaotic systems.

Finally, we have applied the proposed method to a NMR laser data, which is known to be nonlinear time series [2], and time intervals of γ-ray emissions of cobalt, which has been recognized as random. The number of data points used is 2048 in both the cases. Consistent with the presented results, Fig. 5 shows that the NMR system is nonlinear and the cobalt system is linear.

By taking advantage of different structures for linear and nonlinear systems, we described an algorithm to provide data sets for testing nonlinearity without applying the Fourier transform. Not only is this algorithm exceedingly simple, it also avoids some of the numerical issues concerned with estimating Fourier transforms from data. Our theoretical arguments and computational examples show that this algorithm succeeds in testing nonlinearity and discriminating well between linear and chaotic systems: even in cases (such as the CNN) where visual inspection of the data are inconclusive.

This research is supported by a Hong Kong University Grants Council Competitive Earmarked Research Grant (CERG) No. PolyU 5235/03E.

[1] J. Theiler *et al.*, Physica D **58**, 77 (1992).
[2] H. Kantz and T. Schreiber, *Nonlinear Time-Series Analysis* (Cambridge University Press, Cambridge, 1997).
[3] M. Small *et al.*, Phys. Rev. Lett. **87**, 188101 (2001).
[4] T. Schreiber, Phys. Rev. Lett. **80**, 2105 (1998).
[5] J. Theiler and P. E. Rapp, Electroencephalogr. Clin. Neurophysiol. **98**, 213 (1996).
[6] D. Kugiumtzis, Int. J. Bifurcation Chaos Appl. Sci. Eng. **11**, 1881 (2001).
[7] D. Yu *et al.*, Phys. Rev. E **61**, 3750 (2000).
[8] A. Galka *et al.*, Physica D **121**, 237 (1998).
[9] F. Takens, Lect. Notes Math. **898**, 366 (1981).
[10] C. J. Cellucci *et al.*, Phys. Rev. E **67**, 066210 (2003).
[11] Strictly speaking, there is no such thing as "correct" embedding parameters Ref. [10]. By "correct" we mean that the embedding dimension is the dimension in which the attractor may be embedded, we set the embedding lag for discrete or noisy systems to be 1.
[12] J. Theiler and D. Prichard, Physica D **94**, 221 (1996).
[13] By substituting the precise form of the linear process $s(t)=\Sigma_i a_i s(t-i)-\Sigma_j b_j \epsilon(t-j)$, where $\epsilon(t)\sim N(0,\sigma^2)$ and a_i and b_j are arbitrary autoregressive moving average (ARMA) parameters, this argument becomes rigorous.
[14] F. Mitschke *et al.*, Phys. Rev. A **37**, 4518 (1988).
[15] One of the useful methods to find two uncorrelated intervals is to estimate the autocorrelation. To generate test data, it is preferable to use τ where the autocorrelation function indicates zero or negative correlations.
[16] This number is not strict. We can use other numbers between 0.4 and 0.6 with consistent results.
[17] M. Small and K. Judd, Phys. Rev. E **59**, 1379 (1998).
[18] P. E. Rapp *et al.*, Int. J. Bifurcation Chaos Appl. Sci. Eng. **11**, 983 (2001).
[19] M. Henon, Commun. Math. Phys. **50**, 69 (1976).
[20] K. Aihara *et al.*, Phys. Lett. A **144**, 333 (1990).
[21] The largest Lyapunov exponent estimated directly from the model (not an embedding of the data) is 0.007 873 8 and the dimension is 1.043 032. This system is therefore consistent with chaotic dynamics and a deterministic fractal attractor. However, we do not wish to overinterpret these numbers as sure evidence of chaos.