

Detecting Determinism in Time Series: The Method of Surrogate Data

Michael Small, *Member, IEEE*, and Chi K. Tse, *Senior Member, IEEE*

Abstract—We review a relatively new statistical test that may be applied to determine whether an observed time series is inconsistent with a specific class of dynamical systems. These *surrogate data* methods may test an observed time series against the hypotheses of: i) independent and identically distributed noise; ii) linearly filtered noise; and iii) a monotonic nonlinear transformation of linearly filtered noise. A recently suggested fourth algorithm for testing the hypothesis of a periodic orbit with uncorrelated noise is also described. We propose several novel applications of these methods for various engineering problems, including: identifying a deterministic (message) signal in a noisy time series; and separating deterministic and stochastic components. When employed to separate deterministic and noise components, we show that the application of surrogate methods to the residuals of nonlinear models is equivalent to fitting that model subject to an information theoretic model selection criteria.

Index Terms—Hypothesis testing, minimum description length, noise separation, nonlinear modeling, surrogate data.

I. INTRODUCTION

AN IMPORTANT problem in many areas of signal processing is to determine whether an observed time series is deterministic, contains a deterministic component, or is purely stochastic. Equivalently, one may consider the problem of separating an observed time series into deterministic (message) and stochastic (noise) components. In this paper we review the method of surrogate data and suggest two alternative techniques that may be employed for effective system identification and separation of signal and noise.

The method of surrogate data provides a rigorous way to apply statistical hypothesis testing to experimental time series. One may apply the method of surrogate data to determine whether an observed time series has statistically significant deterministic component. Three standard linear techniques are widely applied in the physical and biological sciences (see for example, [1]) to test the hypotheses of [2]: i) independent and identically distributed (i.i.d.) noise¹; ii) linearly filtered noise; and (iii) static monotonic nonlinear transformation of linearly filtered noise. Surrogate methods require one to generate an ensemble of artificial time series (the *surrogates*) that are both “like” the data being tested and consistent with

the hypothesis of interest. One then statistically compares the data and surrogates. Many extensions of this basic method have been suggested ([3] and [4]) but, in this current work, we are particularly interested in a method that may be applied to test for nonlinear nonperiodic determinism in apparently periodic time series [5], [6].

Surrogate methods identify whether an observed time series contains determinism. By themselves, they cannot separate the noise and deterministic components. We employ nonlinear modeling techniques [7] that utilize the information theoretic *minimum description length* (MDL) [8] to fit data without over fitting. These methods have been successfully employed to identify and model the deterministic component of time series of many experimental systems [9]–[11]. We show that these methods may be employed to separate the noise and signal in an observed time series. Alternatively, we show that surrogate data methods applied to the model residuals may also be used as a model fitting criterion. Surrogate data methods may then be used to separate certain classes of noise, such as i.i.d. or colored noise, from complex deterministic dynamics in a data set. A similar interpretation of surrogate data methods has been suggested by Takens [12]. Takens considered fitting some model to a time series and applying i.i.d. surrogate tests to the residuals. He showed that if the residuals are i.i.d., then the model offers a good fit to the data. We extend this result and use it to show that surrogates may be employed to determine whether a model captures all the deterministic structure in a time series.

Section II gives a review of existing surrogate data methods and describes the application of pseudoperiodic surrogates (PPSs) in more detail. In Section III, we present several examples of these methods, both to detect deterministic dynamics and filtered noise in an observed signal. Finally, in Section IV we conclude.

II. SURROGATE DATA AND NONLINEAR MODELING

In this section, we review the three main techniques employed in this paper. Section II-A concerns the application of surrogate data hypothesis testing: a standard method in the dynamical systems literature. In Section II-B, we describe the PPS generation scheme, introduced recently by Small and coworkers [5]. Section II-B also discusses some new work concerning the application of this algorithm to arbitrary time series data (not just those that exhibit periodic trends). Finally, in Section II-C, we review the nonlinear modeling and information theoretic techniques we employ to model time series data. This modeling method has been utilized in several previous publications. Here, we propose that also be employed to separate noise from deterministic dynamics.

Manuscript received January 15, 2002; revised October 29, 2002. This work was supported by the Hong Kong Polytechnic University under Grant G-YW55. This paper was recommended by Associate Editor N. Ling.

The authors are with the Department of Electronic and Information Engineering, Hong Kong Polytechnic University, Kowloon, Hong Kong (e-mail: Michael.Small@polyu.edu.hk).

Digital Object Identifier 10.1109/TCSI.2003.811020

¹By i.i.d. we mean that the time series observations are drawn *independently* from *identical* probability distributions.

A. Linear Surrogate Methods

Standard surrogate techniques were first suggested by Theiler and colleagues in 1992 [2]. Using surrogate data methods, one tests an experimental time series for membership of a specific class of dynamical systems. That is, we test whether the process that generated a time series is an instance of a specific form of system—for example, i.i.d. or linearly filtered noise. Let $\{y_t\}_{t=1}^N := \{y_1, y_2, y_3, \dots, y_N\}$ denote a time series of N measurements. Where the meaning is clear we will drop the indexing and abbreviate this to $\{y_t\}_t$. For each class of dynamical system \mathcal{H}_0 , one generates an ensemble of N surrogates $\{s_t^{(n)}\}_t$ ($n = 1, 2, \dots, N$), consistent with both the data $\{y_t\}_t$ and the class of dynamical systems being tested. These surrogates represent typical realizations of \mathcal{D} , the dynamical system that generated $\{y_t\}_t$, if $\mathcal{D} \in \mathcal{H}_0$. One then computes some statistic $d(\cdot)$ for the data and surrogates. If $d(\{y_t\}_t)$ is significantly different from the ensemble $d(\{s_t^{(n)}\}_t)$ ($n = 1, 2, 3, \dots, N$), then, the class of dynamical systems \mathcal{H}_0 may be rejected as the likely origin of $\{y_t\}_t$. If $d(\{y_t\}_t)$ is not atypical of $\{d(\{s_t^{(n)}\}_t) : \forall n\}$ then membership of \mathcal{H}_0 may not be rejected. It is important to note that failure to reject a particular class of dynamical systems is not the same thing as accepting that class as the likely origin of the data. Failure to reject only implies that the particular statistic $d(\cdot)$ was unable to distinguish between data and surrogates.

Two issues remain to be addressed: the exact nature of the algorithm used to generate surrogates consistent with meaningful hypotheses; and the selection of $d(\cdot)$.

Surrogate generation algorithms are best illustrated by summarizing those originally suggested by Theiler and colleagues [2]. The three algorithms are known as Algorithms 0, 1, and 2.

Algorithm 0 Surrogates generated by this algorithm are i.i.d. noise. To generate i.i.d. noise surrogates, one simply shuffles the data. The shuffling process will destroy any temporal correlation and surrogates thus generated are essentially random observations drawn (without replacement) from the same probability distribution as the data.

Algorithm 1 Surrogates generated by this algorithm are linearly filtered noise. To generate these surrogates one takes the discrete Fourier transform of the data and shuffles (or randomise) the phases of the complex conjugate pairs. Note that the phases of the complex numbers must be shuffled pairwise to preserve the realness of the inverse Fourier transformation. The surrogate is the inverse Fourier transform. By shuffling the phases but maintaining the amplitude of the complex conjugate pairs the surrogate will have the same power spectrum (and autocorrelation) as the data, but will have no nonlinear determinism.

Algorithm 2 Surrogates generated by this algorithm are (approximately, see [4] and [6]) static monotonic nonlinear transformations of linearly filtered

noise. Generating surrogates with this algorithm can be somewhat awkward. The various caveats are amply discussed in the literature [4], [6] and will not be discussed here. The algorithm aims to preserve both the power spectrum and probability distribution of the data (and is therefore well suited to non-Gaussian time series). Generate a Gaussian time series of the same length as the data, and reorder it to have the same rank distribution. Take the Fourier transform of this and randomise the phases (as for Algorithm 1). Finally, the surrogate is obtained by reordering the original data to have the same rank distribution as the inverse Fourier transform. By rank distribution we mean that the i th value in both the time series and surrogate data sets will be the j th largest, for all i and j .

We note that there is considerable discussion of the merits of these algorithms (particularly Algorithm 2) in the literature. In most cases the cautious application of these algorithms should provide adequate results, for more information on the technical details we refer the reader to the review [4] or the discussion of [6].

Finally, to apply these surrogate methods one must choose a suitable statistic $d(\cdot)$. As these methods were originally introduced as a “sanity test” for correlation dimension estimation [2], correlation dimension is a popular choice [1], [6]. A complete discussion of choice of statistic may be found in [13], and Small and Judd [3] demonstrate that correlation dimension is indeed a good choice. We will paraphrase these results and say that the chosen statistic should be independent of the surrogate generation method (linear autocorrelation or mean are therefore bad choices for Algorithm 2) and yet sensitive to deviation from the class of dynamical systems being tested (thus, nonlinear prediction is a good choice for the three linear hypotheses). The results in this paper use a correlation dimension estimation algorithm proposed by Judd [14], [15] that has been shown to be a good choice of test statistic for many classes of dynamical systems of interest [3]. For the sake of brevity, we do not define correlation dimension here, other than to say that it is a measure of structural complexity in the observed dynamics.

B. PPSs

The surrogate generation algorithms described in the previous section test for independent noise, linear noise or statically filtered linear noise. Very many time series of interest exhibit periodic fluctuations and are obviously inconsistent with linear noise. In Fig. 1 we show examples of linear surrogates generated from two typical time series. In the case of a system exhibiting a strong periodic component (Chua’s circuit) these linear surrogates are clearly inadequate. For such systems, an alternative approach has been suggested [5]. In [6], we described this approach and applied it to several artificial and experimental systems. The PPS algorithm generates surrogates that preserve coarse deterministic features (such as periodic trends) but destroy fine structures (such as deterministic chaos). In Fig. 2, we

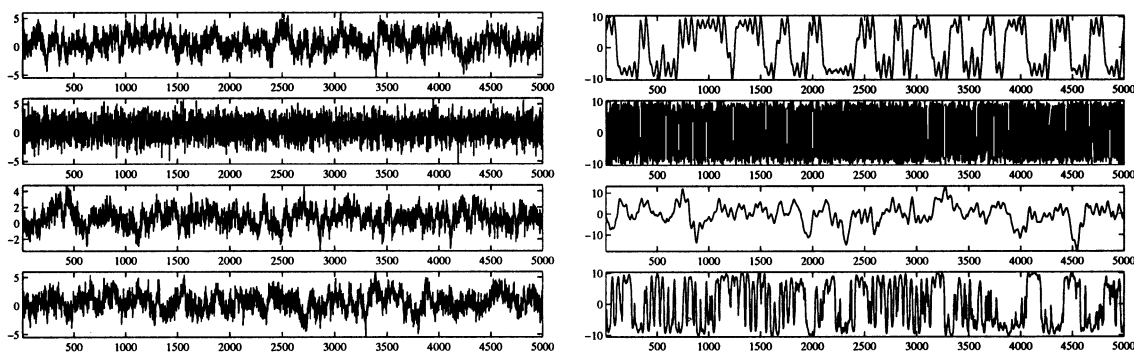


Fig. 1. Linear (Algorithms 0, 1, and 2) surrogates for two time series. The Ikeda map contaminated with linear noise (left side panels) and Chua’s circuit in its chaotic regime with both dynamic and observational noise (right). The top two panels are the original time series. Below these are (in order): Algorithm 0, Algorithm 1, and Algorithm 2 surrogates. Note that the surrogates appear qualitatively like the Ikeda data, but dissimilar to the Chua’s circuit simulation. Chua’s circuit contains nonlinear determinism which is not modeled adequately as a static monotonic nonlinear transformation of linear filtered noise. Quantitative analysis reveals that both these data sets are clearly distinct from these surrogates.

show that this method can be applied to differentiate between deterministic chaos with noise and a periodic orbit. The algorithm generates surrogates that are very like the original data, yet only have the large-scale deterministic features.

The basic algorithm for this method is the following.

1. Construct the vector delay embedding $\{z_t\}_{t=1}^{N-d_w}$ from the scalar time series $\{y_t\}_{t=1}^N$ according to

$$z_t = [y_t, y_{t+\tau}, y_{t+2\tau}, \dots, y_{t+d_e\tau}]$$

where the *embedding dimension* d_e and *embedding lag* τ remain to be selected. The *embedding window* d_w is defined by $d_w = d_e\tau - 1$.

2. Call $\mathcal{A} = \{z_t | t = 1, 2, \dots, N - d_w\}$ the *reconstructed attractor*.
3. Choose an initial condition $s_1 \in \mathcal{A}$ at random.
4. Let $i = 1$.
5. Choose a near neighbor $z_j \in \mathcal{A}$ of s_i according to the probability distribution

$$\text{Prob}(z_j = z_t) \propto \exp \frac{-\|z_t - s_i\|}{\rho}$$

where the parameter ρ is the *noise radius*.

6. Let $s_{i+1} = z_{j+1}$ be the successor to s_i .
7. Increment i .
8. If $i < N$ go to step 5.
9. The surrogate time series is $\{(s_t)_1\} := \{(s_1)_1, (s_2)_1, (s_3)_1, \dots, (s_N)_1\}$, the scalar first components of $\{s_t\}_t$.

The above algorithm has three parameters: ρ , d_e , and τ . Selection of embedding parameters d_e and τ is discussed at length in the literature (see, for example, [16] and the references therein) and we do not consider this problem here. The noise radius ρ is selected according to the suggestions of [5]. We choose ρ such that the expected number of sequences of length two or more that are identical for data and surrogates is

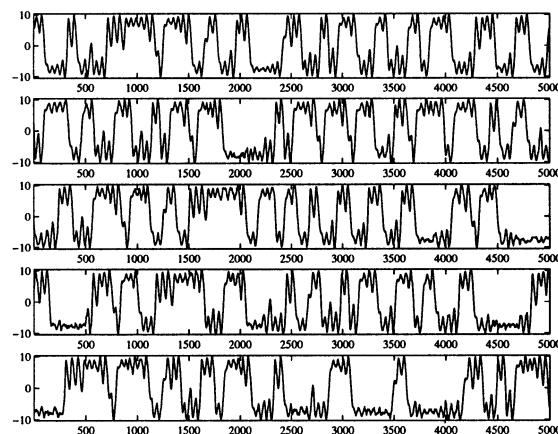


Fig. 2. Surrogates generated by the PPS algorithm for Chua’s circuit data. Note that the data and surrogates are qualitatively similar. However, quantitative analysis reveals that this data is clearly distinct from the system modeled by the surrogates—a period orbit with uncorrelated noise. Each of these surrogates lacks the deterministic “signature” of chaos. The original data are depicted in the top panel.

maximized. This selection criteria provides a balance between: 1) too much randomization (few identical sequences of length; and 2) too little (data and surrogate near identical).

According to [5], this algorithm can be applied to test for non-periodic deterministic dynamics in pseudoperiodic time series.² Conversely, dynamical systems consistent with these surrogates exhibit a periodic orbit with uncorrelated noise. For time series exhibiting pseudoperiodic dynamics, rejection of this class of systems implies the existence of nonperiodic deterministic dynamics. However, for time series that do not exhibit pseudoperiodic dynamics, we have shown [6] that surrogates generated by this method are actually consistent with short-term deterministic dynamics. Rejection of this test therefore is evidence for long-term deterministic dynamics. The distinction between “long term” and “short term” is perhaps ill defined, and certainly dependent on the selection of ρ . However, for the current study a more precise definition is beyond our requirements. It is sufficient to note that surrogates generated by this method appear

²A pseudoperiodic time series is one which exhibits a definite periodic trend: there exists $\tau_0 > 0$ such that the autocorrelation $\rho(\tau)$ has a nontrivial peak $\rho(\tau_0) > 0$ at τ_0 .

much more like the data than Algorithms 0, 1, and 2 surrogates. Rejection of the PPS algorithm indicates the presence of non-trivial deterministic dynamics in the data.

C. Nonlinear Modeling

In the last two sections, we have discussed various surrogate techniques. We now turn our attention to a class of nonlinear models [7] utilising an information theoretic stopping criterion [8]. The difference between these techniques is not as great as it may seem. In the previous sections, we have deliberately described surrogate data *hypothesis testing* in terms of dynamical systems and classes of models. Conversely, in [3] and [12], it has been shown that noisy iterated predictions of a model may be used as a form of hypothesis testing.

For a scalar time series $\{y_t\}_t$ we perform a time delay embedding [17] according to

$$z_t = [y_t, y_{t-1}, y_{t-2}, \dots, y_{t-d_w}]. \quad (1)$$

(There is a minor notational change between this embedding and that in Section II-B. This distinction is cosmetic and largely a matter of convenience.) We then model the dynamics of the map $z_t \mapsto z_{t+1}$ by the construction of a function f such that

$$y_{t+1} = f(z_t) + e_t$$

where the *residuals* e_t are expected to be i.i.d. random variates. The function $f(\cdot)$ is of the form

$$f(z_t) = \lambda_0 + \sum_{i=1}^m \lambda_i y_{t-\ell_i} + \sum_{j=1}^n \lambda_{m+j} \phi\left(\frac{\|z_t - c_j\|}{r_j}\right) \quad (2)$$

where $\lambda_i \in \mathbf{R}$ are the *weights*; $\ell_i \in \mathbf{R}$ such that $0 \leq \ell_i < \ell_{i+1} \leq d_w$ are the *lags*; $\phi(x) = \exp x^2/2$ are the *basis functions*; $c_j \in \mathbf{R}^{d_w}$ are the *centers*; and $r_j \in \mathbf{R}^+$ are the *radii*. In fact, an often more useful generalization of this scheme is described in [9]. In addition to each of these parameters, the model size $(m+n)$ must also be selected³ so that the variate e_t are indeed i.i.d. but yet model overfitting must be avoided. Note that the selection of the model form (2) is based solely on the computational algorithms at our disposal. Consideration of many other similar schemes may be found in the literature (see, for example, [11]).

To provide a suitable model fitting criterion we employ *minimum description length* (MDL) as described by Rissanen [8] and applied to radial basis modeling by Judd and Mees [7]. Roughly speaking, the description length of a data set is the length of shortest description that can be employed to reconstruct the entire data set. For random variates, that code will simply be the data itself and the description length will be the length of that data set. For data containing some determinism, the shortest description length of that data will be the description of a model of the deterministic component and the model prediction errors.

Rissanen [8] showed that the description length of a parameter λ_i (specified to some finite *precision* δ_i) is

$$\log \frac{\gamma}{\delta_i}$$

³The model selection scheme described in [7] implies that m and n are not independent.

where γ is a constant related to the number of bits in the mantissa of the binary representation of λ_i . Denote by Λ the model parameters $\lambda_0, \lambda_1, \dots, \lambda_{m+n}$. Then, if we assume⁴ that the model (2) may be completely described by the linear parameters λ_i ($i = 0, 1, \dots, m+n$) then the description length of the model is

$$L(\lambda_0, \lambda_1, \dots, \lambda_{m+n}) = L(\Lambda) = \sum_{i=0}^{m+n} \log \frac{\gamma}{\delta_i}. \quad (3)$$

The description length of the data *with this model* is then given by

$$L(x) = L(x|\Lambda) + L(\Lambda) \quad (4)$$

where $L(x|\Lambda)$ is the negative log likelihood of the model prediction errors under the assumed distribution (5). The likelihood of a data set is the probability of observing that particular set of values. Assuming that the model prediction errors are Gaussian distributed noise with standard deviation σ^2 , the negative log likelihood is given by

$$L(x|\Lambda) = -\ln \text{Prob}(x|\Lambda) = \frac{1}{(2\pi\sigma^2)^{N/2}} \exp \frac{e^T e}{2\sigma^2}. \quad (5)$$

Typically, one will substitute the estimate $\hat{\sigma} = \sqrt{e^T e}$ for the unknown variance of the noise.

A model may then be assessed based on its description length as given in (4). The best model of a particular data set is that for which this quantity is lowest.

One observes that the important quantity in the above computation is the set of precisions of the parameters δ_i . It can be shown [7] that these precisions may be calculated as the solution of

$$(Q \cdot [\delta_0 \delta_1 \delta_2 \dots \delta_{m+n}]^T)_i = \frac{1}{\delta_i} \quad (6)$$

where Q is the second partial derivative of $f(\cdot)$ with respect to the model parameters. In Fig. 3, we show the typical behavior of MDL and root-mean-square error for a model of the data depicted in (2).

III. DETECTING AND EXTRACTING DETERMINISM

We now apply the methods discussed in Section II to simulated and experimental data. We show that linear and nonlinear surrogates as described in Sections II-A and II-B can be successfully applied to realistic data to differentiate between determinism and stochastic behavior. We then consider the application of surrogate techniques to nonlinear modeling and show that they behave the same as minimum description length and may be used as a form of model selection criteria.

In Section III-A, we present the application of surrogate methods to various engineering systems. Section III-B compares minimum description length and surrogate methods as model selection criteria.

⁴This is an approximation intended to make the calculations that follow substantially easier. A full treatment is provided in [9].

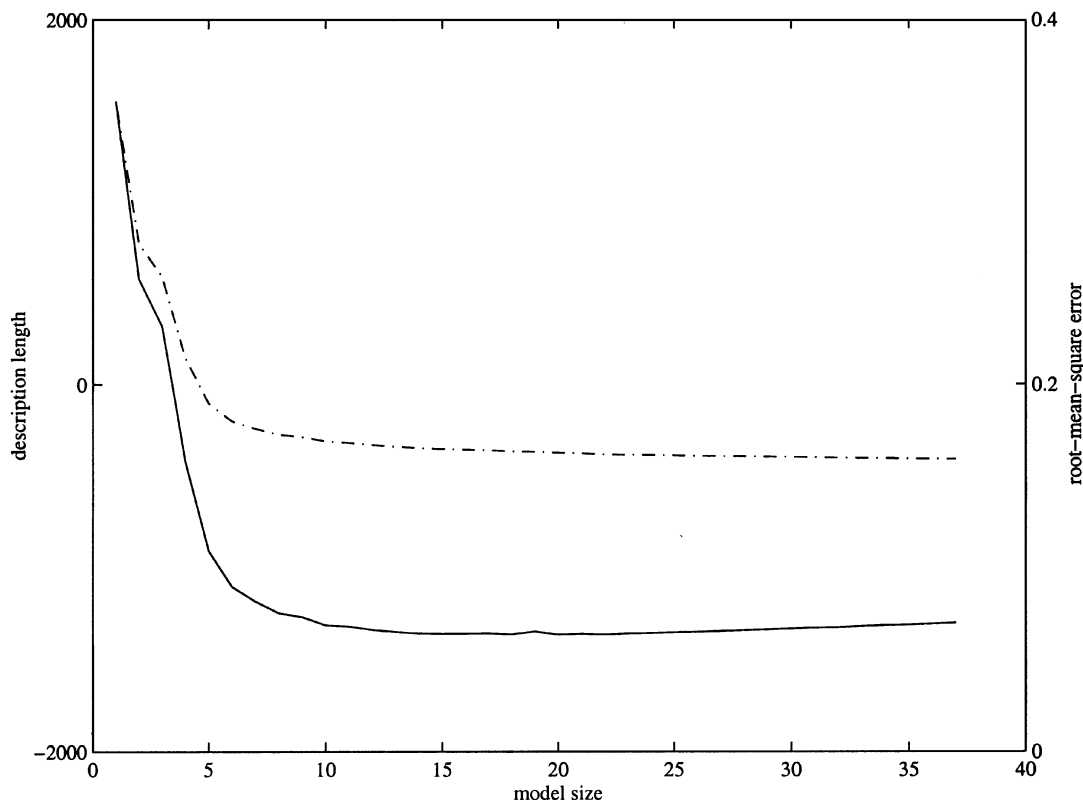


Fig. 3. Description length and root-mean-square of the residuals as a function of model size for the data depicted in the upper panel of Fig. 2. The dot-dashed line is root-mean-square prediction error, the solid line is description length. Note that the root-mean-square prediction error is a monotonically decreasing function of model size: as the model gets bigger the errors get smaller. However, the description length is not monotonic, a minimum is observed at model size 22. Increasing the model size beyond the minimum will cause description length to increase.

A. Detecting Determinism

We consider the application of the linear surrogate algorithms and the PPS algorithm to six test systems: i.i.d. random variates (generated from a radioactive decay process); a colored noise process; a monotonic nonlinear transformation of colored noise; Ikeda type chaos with observational and dynamic noise; and the Chua circuit during both periodic (with observational noise) and chaotic behavior (with observational and dynamic noise). For completeness, we describe the equations of motion and parameter settings employed in this paper.

i.i.d. Random Variates: 5000 uniform random variates were obtained as the product of two independent and uniformly distributed integers between 1 and 256.

Colored Noise: Colored noise was generated as realizations of the random walk process $x_t = 0.5x_{t-1} + 0.3x_{t-2} + 0.1x_{t-3} + e_t$ where e_t are Gaussian random variates with mean 0 and standard deviation 1.

Monotonic Transformation of Colored Noise: Generated by applying the monotonic nonlinear transformation $g(x) = x^3$ to colored noise (as above).

Ikeda Chaos: The Ikeda map is given by

$$\begin{pmatrix} \dot{x} \\ \dot{y} \end{pmatrix} = \begin{pmatrix} 1 + \mu(x \cos \theta - y \sin \theta) \\ \mu(x \sin \theta + y \cos \theta) \end{pmatrix} \quad (7)$$

and

$$\theta = 0.4 - \frac{6}{1 + x.^2 + y.^2} \quad (8)$$

where, for $\mu = 0.7$, this system is chaotic.

Chua's Circuit: Chua's circuit equations are given by

$$\begin{pmatrix} \dot{x} \\ \dot{y} \\ \dot{z} \end{pmatrix} = \begin{pmatrix} a(y - x - \phi(x)) \\ x - y + z \\ -by \end{pmatrix} \quad (9)$$

and

$$\phi(x) = m_1 x + \frac{(m_0 - m_1)}{2} (|x + 1| - |x - 1|) \quad (10)$$

where $m_0 = -8/7$ and $m_1 = -5/7$. We investigate the behavior of the system for: $a = 9$ and $b = 14(2/7)$ ("double scroll" chaos); $a = 8.86$ and $b = 16$ (stable period 2 dynamics); and, $a = 9.4$ and $b = 16$ ("single-band" chaos).

In each case, observational and dynamic noise are computer generated pseudorandom numbers conforming to a Gaussian distribution. The distinction between *observational* and *dynamic* noise is the following. Let $f(x)$ be the discrete time map under consideration (or for a flow the flow integrated over one time interval) and $g(x)$ be the scalar measurement function. Then

$$z_{t+1} = f(z_t) + \delta_t$$

and the time series we consider is

$$y_t = g(z_t) + \epsilon_t.$$

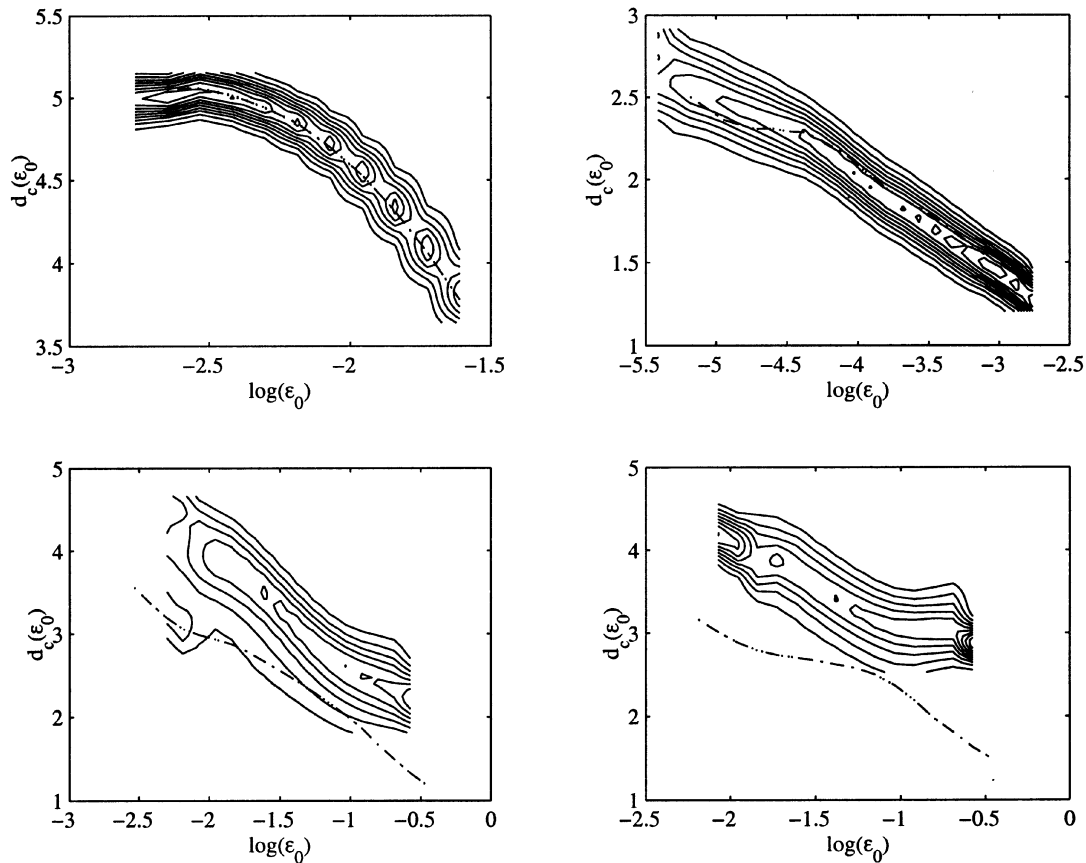


Fig. 4. Probability distribution (shown as a contour plot) of correlation dimension estimates for 50 Algorithm 2 surrogates and the original data for data from four different systems is shown. Any value beyond the lowest (outside) isocline would indicate an outlier of the surrogate distribution. The four systems considered are: a colored noise process (top left), a monotonic nonlinear transformation of colored noise (top right), and Chua's circuit in chaotic (bottom left) and periodic (bottom right) regime.

The random variates δ_t are dynamic noise (these affect the time evolution of z_t), and the variates ϵ_t are observational noise (these do not affect the evolution of z_t).

In each case, we employ correlation dimension [14], [15] as a test statistic. The algorithm described in [14] and [15] estimates correlation dimension d_c as a function of viewing scale ϵ_0 , hence $d_c(\epsilon_0)$. Therefore, the correlation dimension for each time series is not a single number, but a curve. To compare curves, we either plot the results graphically (as in Figs. 4 and 5) or compute a single statistic, the *deviation*. The deviation is computed to be the maximum difference between the correlation dimension estimated for the data and the mean of the correlation dimension for the surrogates, expressed in terms of the standard deviation of the surrogates. The results of these calculations are summarized in Table I. If these distributions were Gaussian, then a value of deviation greater than three would be statistically significant. Although it does not necessarily follow that these distributions are Gaussian, a value of deviation greater than about five is certainly significant and anything below that is questionable. Fig. 4 displays the results of the Algorithm 2 surrogate calculation for four of the data sets considered here, and Fig. 5 is the same result for the PPS Algorithm. A more detailed examination of the PPS algorithm is presented in [6].

Table I shows that the test systems are identified correctly: i.i.d. noise is not rejected by Algorithm 0 (although it is spuriously rejected by Algorithm 1); colored noise is rejected by Al-

gorithm 0, but not by the others; and, the monotonic nonlinear transformation of linearly filtered noise is clearly distinct from Algorithm 0 and 1 surrogates, but not Algorithm 2. Of the deterministic systems, all are inconsistent with the linear (Algorithms 0, 1, and 2) surrogates, and all are inconsistent with the PPS algorithm. This is most notable because the Ikeda data are visually indistinguishable from noise. Furthermore, it is worth noting that surrogates generated by the PPS algorithm appear qualitatively similar to the data (unlike the linear surrogates) but quantitative analysis demonstrates that they are significantly different.

That Table I shows i.i.d. noise being rejected by Algorithm 1 is an indication of why one needs to apply this hierarchy of surrogate algorithms. In contrast, we note that the PPS data is (according to Table I) indistinguishable from the Ikeda map data. This is a demonstration of why failure to reject a surrogate test is not evidence that the hypothesis is true; only that we have failed to find evidence that it is false. By varying the test statistic, one readily observes the converse. It is simply that in this particular case (a two-dimensional map), correlation dimension is not sensitive to the distinction between the data and surrogates.

Figs. 4 and 5 provide further illustration of these calculations for four data sets. The noise data (colored noise and a static monotonic transformation of colored noise) are indistinguishable from the surrogates (correlation dimension estimates are coincident). However, both Algorithm 2 and PPS data are

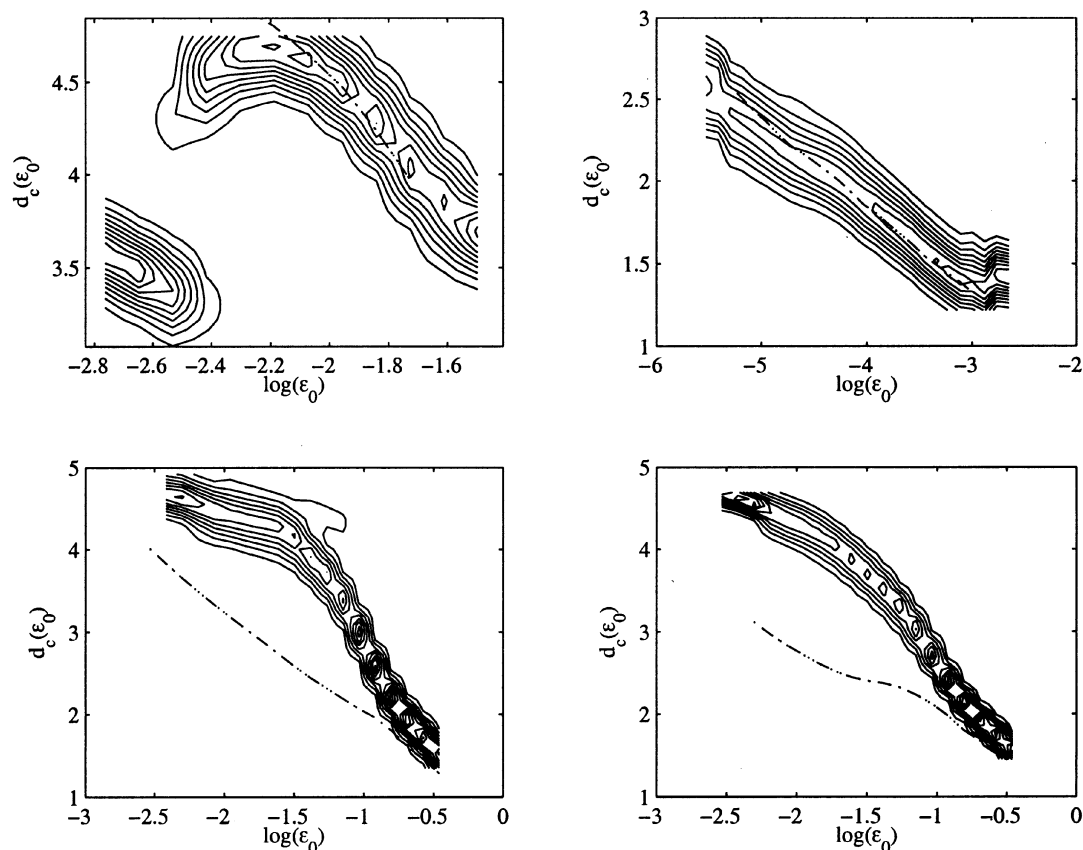


Fig. 5. Probability distribution (shown as a contour plot) of correlation dimension estimates for 50 PPS algorithm surrogates and the original data for data from four different systems is shown. Any value beyond the lowest (outside) isocline would indicate an outlier of the surrogate distribution. The four systems considered are: a colored noise process (top left), a monotonic nonlinear transformation of colored noise (top right), and Chua circuit in chaotic (bottom left) and periodic (bottom right) regime. Note that, although the correlation dimension estimate and contour plot appear to differ in the top left plot this is only because the ensemble of surrogate data sets failed to provide an estimate of d_c for these values of ε_0 . That is, for values of ε_0 for which data and surrogate dimension estimates differ, the number of available surrogates is low. Furthermore, in the case of significant deviation between data and surrogates we expect the correlation dimension of the data to be *lower* than that for the surrogates. Results such as those in the top right panel are more typical.

TABLE I
MAXIMUM NUMBER OF STANDARD DEVIATIONS (OF SURROGATE DISTRIBUTION) SEPARATING THE MEAN OF THE SURROGATE DISTRIBUTION FROM THE DATA ESTIMATE OF $d_c(\varepsilon_0)$

System	Alg. 0	Alg. 1	Alg. 2	PPS
i.i.d. noise	1.9515	8.5915	0.83616	-0.32588
coloured noise	85.577	0.8214	0.47828	0.61221
monotonic transformation of coloured noise	25.558	12.654	1.0058	0.47261
noisy Ikeda map	57.958	11.78	32.909	0.7936
Chua's circuit: single band chaos	5.011	5.5891	5.9146	12.023
Chua's circuit: period 2	2.6582	2.5655	3.5315	9.0051
Chua's circuit: double scroll	6.4708	6.9028	10.806	15.084

clearly distinct from the nonlinear systems (Chua's circuit during broad band chaos and periodic dynamics). Note that this result is as one expects: Chua's circuit in periodic motion was simulated with both observational and dynamic noise. The dynamic noise means that the system is inconsistent with the hypothesis of periodic motion with uncorrelated noise proposed for PPS data [5].

B. Extracting Determinism

In this section, we demonstrate that surrogate data calculations may be employed as a form of model selection criteria.

Minimum description length and other information theoretic measures offer a well founded model selection criteria. However, such measures are often very computationally expensive or analytically complex [9]. Surrogate data provides a straightforward and robust alternative.

We take as our test system 4000 points of the Chua circuit equations during "double-scroll" type chaos ($a = 9$ and $b = 14(2/7)$), sampled at 40 Hz with observational and dynamic noise of $N(0, 0.05)$ and $N(0, 0.1)$ (added to each component of the vector time series) respectively. Although this observation noise level is modest, this level of dynamic noise is sufficient to have a substantial impact on the observed dynamics — especially for a chaotic system. A representative time series and a delay reconstruction of the underlying attractor are depicted in Fig. 6. It is clear from Fig. 6 that the observation and dynamic noise have a substantial impact on the observed time series.

We applied the radial basis modeling scheme described in [7] and [9] to these data. Plots of description length and root-mean-square prediction error are given in Fig. 3. From Fig. 3 the optimal model size is 22 parameters. Although the minimum in the description length curve is only slight, it is significant. For all models of size larger than 22, the description length increases.

For each model of size (i.e., $m + n$) 0 to 36 we applied Algorithm 0 surrogates to the model residuals (prediction errors). The

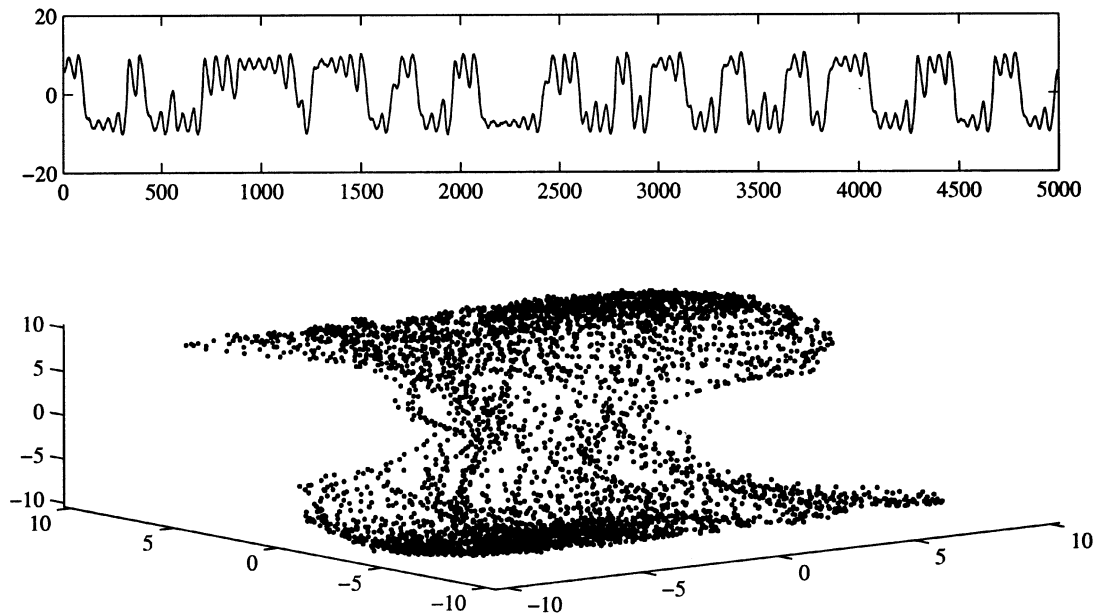


Fig. 6. Computational simulation of 4000 points of Chua's circuit with observation and dynamic noise (standard deviation of 0.1 and 0.05, respectively) is shown in the upper panel. The sampling rate is 40 Hz. The lower panel depicts the attractor reconstructed from this data in three dimensions with a lag of 20.

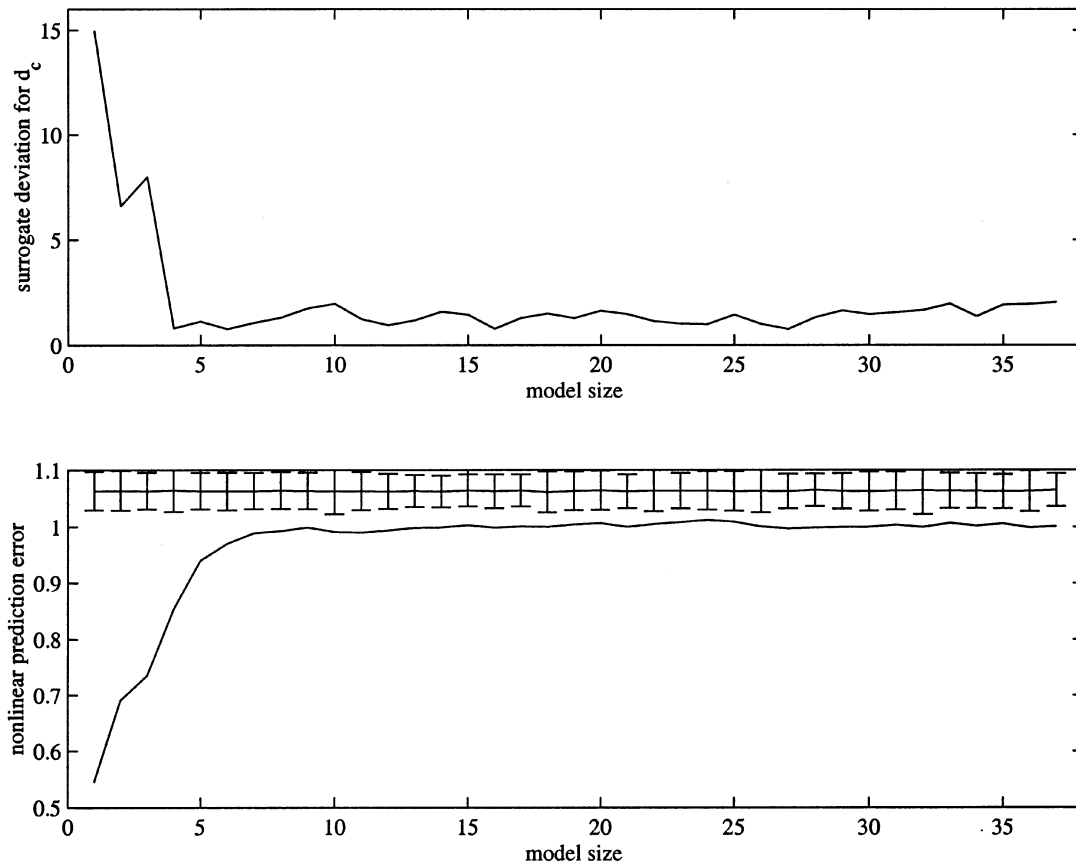


Fig. 7. Top panel shows the number of standard deviations (of the surrogate distribution) separating the mean of the Algorithm 0 surrogate distribution from the value of correlation dimension d_c estimated for the model residuals. The bottom panel is a plot of nonlinear prediction error applied to the residuals as a function of model size. Also shown in this lower plot is the mean and the five standard deviation range of this same quantity for Algorithm 0 surrogates.

results of this calculation are shown in Fig. 7. One can clearly see that this criterion does not indicate the necessity for the large model built according to description length. For all model sizes larger than 4, there is no significant difference between the

model residuals and i.i.d. noise (Algorithm 0 surrogates). However, correlation dimension is not the best statistic to employ in this case. Correlation dimension is insensitive to fine time dependence in the system. Indistinguishable correlation dimen-

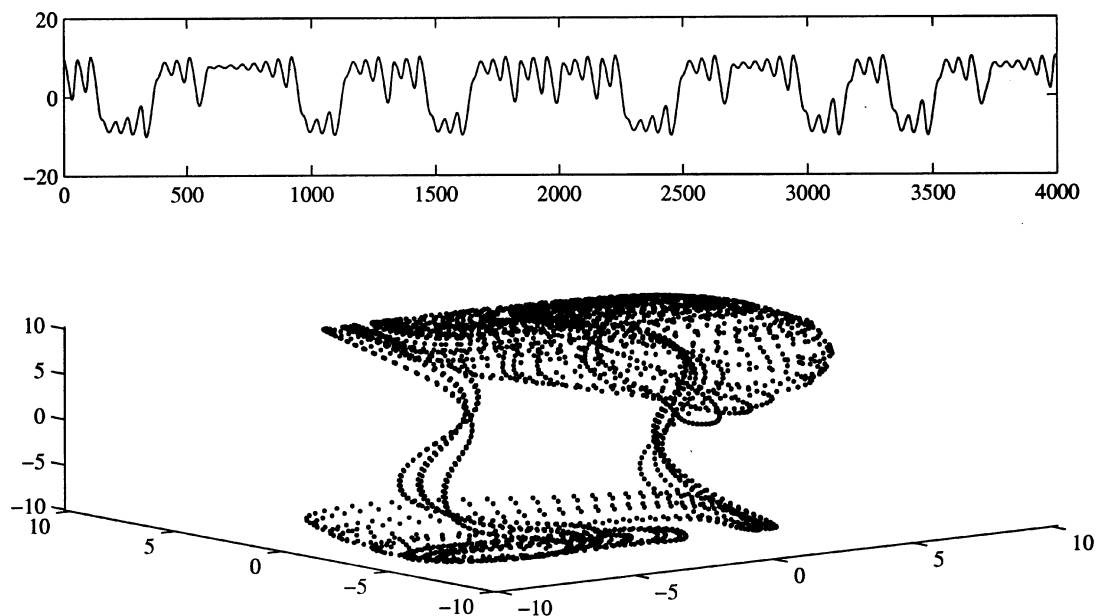


Fig. 8. Attractor reconstructed from a noise free model simulation for the MDL-best model of the data in Fig. 6. The attractor is reconstructed from this data in three dimensions with a lag of 20. Note that the chaotic dynamics exhibited by this model trajectory mimic the behavior of Chua’s circuit very well. All the essential features appear to be preserved, but the noise in the original data (Fig. 6) is absent. This is good evidence that the modeling algorithm described in this paper is fitting the system dynamics without overfitting.

sion estimates only imply that topologically the embedded time series (for model residual and surrogates) are equivalent. As an alternative, we employ a test statistic that is much more sensitive to temporal dependencies in the data: nonlinear prediction error [18].

Nonlinear prediction error is a measure of the amount of determinism in the time series. For a delay embedding such as (1) one predicts the future behavior of state x_t by building a linear model based on the behavior of the spatial neighbors of x_t [19]. Nonlinear prediction of one indicates no temporal determinism, less than one indicates nonlinear determinism. The second panel of Fig. 7 shows the results of this calculation for the model described in Fig. 3. Algorithm 0 surrogates were also used to compute the expected distribution of nonlinear prediction if the residual were i.i.d. noise. That is, the surrogates are employed to act as error bounds on the estimate of nonlinear prediction. One can clearly see that this computation indicates that the residuals of *all model sizes* contain significant temporal determinism. However, the model size for which the nonlinear prediction was least significant corresponds to a model of size 24—very close to the MDL estimate provided by Fig. 3.

Significant nonlinear determinism is observed in models of all sizes. But minimum description length indicates that this determinism is not sufficient to be modeled adequately by (2): the determinism is either very small (in magnitude) or very complex (in terms of radial basis functions).

In Fig. 8, we show the attractor reconstructed from noise free simulations produced by the MDL-best model. It is clear from this simulation that this model is adequate to capture the essential behavior of the original system. It is not meaningful to extract a “true” trajectory in the presence of dynamic and observational noise. However, this simulation shows that equivalent trajectories (in the absence of noise) may be produced by models of the noisy data. Furthermore, the models produced ex-

hibit equivalent dynamics, irrespective of whether an MDL or nonlinear prediction model selection criteria is employed.

Because the data considered in this example contain both dynamic and observational noise, it is not obvious how one may measure the signal-to-noise ratio. Indeed, because the system dynamics are corrupted by the dynamic noise there is no underlying “clean” trajectory. However, a comparable quantity may be estimated as follows. The standard deviation of *the* clean signal may be approximated by the standard deviation of a clean trajectory of Chua’s circuit $\sigma_s \approx 1.34$. The system noise is the net effect of the observational noise (standard deviation 0.1) and dynamic noise (standard deviation of 0.05). However, the dynamic noise is actually a vector (in \mathbf{R}^3) and propagated by the chaotic dynamics of the system. Therefore, we can only estimate the magnitude of the noise from the root-mean-square prediction error of the optimal model⁵ $\sigma_n \approx 0.16$. Hence, we obtain an estimate of signal-to-noise ratio of approximately

$$20 \log \frac{\sigma_s}{\sigma_n} \approx 18 \text{ dB.}$$

IV. CONCLUSION

We have described the existing surrogate data techniques that have been widely applied in the literature to identify *possible* nonlinearity in experimental systems. A new algorithm, the PPS algorithm, may also be applied to experimental time series and has been shown to be more useful for data exhibiting some non-stationarity or determinism. In such systems, this algorithm is able to mimic existing large-scale determinism while destroying fine features, such as deterministic chaos. We demonstrated the

⁵For systems with only observational noise we have found this quantity to be very close to the standard deviation of the noise.

application of these four algorithms to seven test systems and correctly identified the origin of the signal in each case.

To date, surrogate techniques have largely been applied within the nonlinear time series community to screen data prior to analysis with nonlinear methods. Data that is likely to be consistent with these linear algorithms is ignored in favor of more interesting time series. Once such set of data is shown to be distinct from a monotonic nonlinear transformation of linearly filtered noise, it is usual to proceed to the application of more sophisticated data analysis techniques. In an engineering context, these techniques may be applied to observed data to detect significant determinism. One can apply these methods to search for deterministic signals hidden within an apparently random background. For example, surrogate techniques should prove useful for identifying a communication signal encoded with chaotic shift keying: unmasking chaotic masking.

However, to unmask a chaotic signal one needs to know more than its existence. For this reason, we described a nonlinear modeling methods that can separate deterministic and stochastic components in a time series. A common problem for all modeling regimes is how large should one make the model. The algorithm described in this paper addressed this problem with the information theoretic minimum description length. For the data we considered here we found that nonlinear prediction error and surrogate data methods could be employed to provide an equivalent model selection criteria: the best model is achieved when the model prediction errors (residuals) are indistinguishable from i.i.d. variates. We showed that either method could be employed to recover a chaotic attractor from a time series contaminated with dynamic and observational noise.

Apart from the application considered here, a variant on this model selection algorithm may be employed to reduce the residuals to a less trivial form. If the objective were to separate the observed time series into a nonlinear component and a colored noise component one could apply the modeling algorithm until the model residuals are indistinguishable from Algorithm 1 or 2 surrogates. Such a technique could be useful when one is aware of the form of the noise but is only interested in more complex nonlinear structure. However, successful implementation of this would rely on ensuring that the modeling algorithm did not attempt to fit the linear stochastic structure first.

REFERENCES

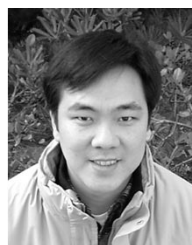
- [1] M. Small, K. Judd, M. Lowe, and S. Stick, "Is breathing in infants chaotic? Dimension estimates for respiratory patterns during quiet sleep," *J. Appl. Physiol.*, vol. 86, pp. 359–376, 1999.
- [2] J. Theiler, S. Eubank, A. Longtin, B. Galdrikian, and J. D. Farmer, "Testing for nonlinearity in time series: The method of surrogate data," *Phys. D*, vol. 58, pp. 77–94, 1992.
- [3] M. Small and K. Judd, "Correlation dimension: A pivotal statistic for nonconstrained realizations of composite hypotheses in surrogate data analysis," *Phys. D*, vol. 120, pp. 386–400, 1998.
- [4] T. Schreiber and A. Schmitz, "Surrogate time series," *Phys. D*, vol. 142, pp. 346–382, 2000.
- [5] M. Small, D. J. Yu, and R. G. Harrison, "A surrogate test for pseudo-periodic time series data," *Phys. Rev. Lett.*, vol. 87, p. 188101, 2001.
- [6] M. Small and C. K. Tse, "Applying the method of surrogate data to cyclic time series," *Phys. D*, vol. 164, pp. 182–202, 2002.
- [7] K. Judd and A. Mees, "On selecting models for nonlinear time series," *Phys. D*, vol. 82, pp. 426–444, 1995.

- [8] J. Rissanen, *Stochastic Complexity in Statistical Inquiry*. Singapore: World Scientific, 1989.
- [9] M. Small and K. Judd, "Comparison of new nonlinear modeling techniques with applications to infant respiration," *Phys. D*, vol. 117, pp. 283–298, 1998.
- [10] M. Small, D. J. Yu, and R. G. Harrison, "Period doubling bifurcation route in human ventricular fibrillation," *Int. J. Bifurcation Chaos*, vol. 13, 2003, to be published.
- [11] M. Small, K. Judd, and A. Mees, "Modeling continuous processes from data," *Phys. Rev. E*, vol. 65, p. 046704, 2002.
- [12] F. Takens, "Detecting nonlinearities in stationary time series," *Int. J. Bifurcation Chaos*, vol. 3, pp. 241–256, 1993.
- [13] J. Theiler and D. Prichard, "Constrained-realization Monte-Carlo method for hypothesis testing," *Phys. D*, vol. 94, pp. 221–235, 1996.
- [14] K. Judd, "An improved estimator of dimension and some comments on providing confidence intervals," *Phys. D*, vol. 56, pp. 216–228, 1992.
- [15] —, "Estimating dimension from small samples," *Phys. D*, vol. 71, pp. 421–429, 1994.
- [16] H. D. I. Abarbanel, *Analysis of Observed Chaotic Data*. New York: Springer-Verlag, 1996.
- [17] F. Takens, "Detecting strange attractors in turbulence," *Lect. Notes Math.*, vol. 898, pp. 366–381, 1981.
- [18] R. Hegger, H. Kantz, and T. Schreiber, "Practical implementation of nonlinear time series methods: The TISEAN package," *Chaos*, vol. 9, pp. 413–435, 1999.
- [19] G. Sugihara and R. M. May, "Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series," *Nature*, vol. 344, pp. 737–740, 1990.



Michael Small (M'01) received the B.Sc. (Hons) degree in pure mathematics and the Ph.D. degree in applied mathematics from the University of Western Australia, Perth, Australia, in 1994 and 1998, respectively.

He is presently an Assistant Professor with the Hong Kong Polytechnic University, Hong Kong, and his research is in the area of nonlinear dynamics and nonlinear time series analysis. His research emphasizes the application of nonlinear time series techniques in a diverse range of areas including: infant respiratory patterns, human cardiac arrhythmia, financial analysis and telecommunications. Prior to his current appointment he has held research posts at Hong Kong Polytechnic University, Hong Kong, Heriot-Watt University, Edinburgh, Scotland, Stellenbosch University, Stellenbosch, South Africa, and the University of Western Australia.



Chi K. Tse (M'90–SM'97) received the B.Eng. (Hons.) degree with first class honors in electrical engineering and the Ph.D. degree from the University of Melbourne, Melbourne, Australia, in 1987 and 1991, respectively.

He is presently a Professor with the Hong Kong Polytechnic University, Hong Kong, and his research interests include chaotic dynamics, power electronics, and chaos-based communications. He is the author of *Linear Circuit Analysis* (London, U.K.: Addison-Wesley, 1998) and *Complex Behavior of Switching Power Converters* (Boca Raton, FL: CRC Press, 2003), coauthor of *Chaos-Based Digital Communication Systems* (Heidelberg, Germany: Springer-Verlag, 2003), and coholder of a U.S. patent.

Prof. Tse served as an Associate Editor of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS PART I—FUNDAMENTAL THEORY AND APPLICATIONS, from 1999 to 2001, and since 1999, he has been an Associate Editor of the IEEE TRANSACTIONS ON POWER ELECTRONICS. In 1987, he was awarded the L.R. East Prize by the Institution of Engineers, Australia, and in 2001, the IEEE TRANSACTIONS ON POWER ELECTRONICS, Prize Paper Award. While with the university, he received twice the President's Award for Achievement in Research, the Faculty's Best Researcher Award and a few other teaching awards. Since 2002, he has been appointed as Advisory Professor by the Southwest China Normal University, Chongqing, China.