Physica D 58 (1992) 77–94
North-Holland

PHYSICA D

Testing for nonlinearity in time series: the method of surrogate data

James Theiler[a,b], Stephen Eubank[a,b,c,d], André Longtin[a,b], Bryan Galdrikian[a,b] and J. Doyne Farmer[a,c,d]

[a]*Theoretical Division, Los Alamos National Laboratory, Los Alamos, NM 87545, USA*
[b]*Center for Nonlinear Studies, Los Alamos National Laboratory, Los Alamos, NM 87545, USA*
[c]*Santa Fe Institute, 1660 Old Pecos Trail, Santa Fe, NM 87501, USA*
[d]*Prediction Company, 234 Griffin Street, Santa Fe, NM 87501, USA*

Received 4 October 1991
Revised manuscript received 11 February 1992
Accepted 3 March 1992

We describe a statistical approach for identifying nonlinearity in time series. The method first specifies some linear process as a null hypothesis, then generates surrogate data sets which are consistent with this null hypothesis, and finally computes a discriminating statistic for the original and for each of the surrogate data sets. If the value computed for the original data is significantly different than the ensemble of values computed for the surrogate data, then the null hypothesis is rejected and nonlinearity is detected. We discuss various null hypotheses and discriminating statistics. The method is demonstrated for numerical data generated by known chaotic systems, and applied to a number of experimental time series which arise in the measurement of superfluids, brain waves, and sunspots; we evaluate the statistical significance of the evidence for nonlinear structure in each case, and illustrate aspects of the data which this approach identifies.

1. Introduction

The inverse problem for a nonlinear system is to determine the underlying dynamical process in the practical situation where all that is available is a time series of data. Algorithms have been developed which can in principle make this distinction, but they are notoriously unreliable, and usually involve considerable human judgement. Particularly for experimental data sets, which are often short and noisy, simple autocorrelation can fool dimension and Lyapunov exponent estimators into signalling chaos where there is none. Most authors agree that the methods contain many pitfalls, but it is not always easy to avoid them. While some data sets very cleanly exhibit low-dimensional chaos, there are many cases where the evidence is sketchy and difficult to evaluate. Indeed, it is possible for one author to claim evidence for chaos, and for another to argue that the data is consistent with a simpler explanation [1–4].

The real complication arises because low-dimensional chaos and uncorrelated noise are not the only available alternatives. The erratic fluctuations that are observed in an experimental time series owe their dynamical variation to a mix of various influences: chaos, nonchaotic but still nonlinear determinism, linear correlations, and noise, both in the dynamics and in the measuring apparatus. While we are motivated by the prospect of ultimately disentangling these influences, we take as a more modest goal the detection of nonlinear structure in a stationary time series. (We will not attempt to characterize non-stationary time series – see refs. [5–9] for a

discussion of some of the problems arising in the estimation of nonlinear statistics from nonstationary data.)

Positive identification of chaos is difficult; the usual way to detect low-dimensional behavior is to estimate the dimension and then see if this value is small. With a finite time series of noisy data, the dimension estimated by the algorithm will at best be approximate and often, outright wrong. One can guard against this by attempting to identify the various sources of error (both systematic and statistical), and then putting error bars on the estimate (see, for example, refs. [10–18]). But this can be problematic for nonlinear algorithms like dimension estimators: first, assignment of error bars requires some model of the underlying process, and that is exactly what is not known; further, even if the underlying process were known, the computation of an error bar may be analytically difficult if not intractable.

The goal of detecting nonlinearity is considerably easier than that of positively identifying chaotic dynamics. Our approach is to specify a well-defined underlying linear process or null hypothesis, and to determine the distribution of the quantity we are interested in (dimension, say) for an ensemble of surrogate data sets which are just different realizations of the hypothesized linear stochastic process. Then, rather than estimate error bars on the dimension of the original data, we put error bars on the value given by the surrogates. This can be done reliably because we have a model for the underlying dynamics (the null hypothesis itself), and because we have many realizations of the null hypothesis, we can estimate the error bar numerically (from the standard deviation of all estimated dimensions of the surrogate data sets) and avoid the issue of analytical tractibility altogether.

While this article elaborates on preliminary work described in an earlier publication [19], our aim is to make this exposition self-contained. In section 2, we express the problem of detecting nonlinearity in terms of statistical hypothesis testing. We introduce a measure of significance, develop various null hypotheses and discriminating statistics, and describe algorithms for generating surrogate data. Section 3 demonstrates the technique for several computer-generated examples under a variety of conditions: large and small data sets, high and low-dimensional attractors, and various levels of observational and dynamical noise. In section 4, we illustrate the application of the method to several real data sets, including fluid convection, electroencephalograms (EEG), and sunspots. With real data, there is always room for human judgment, and we argue that besides formally rejecting a null hypothesis, the method of surrogate data can also be useful in an informal way, providing a benchmark, or control experiment, against which the actual data can be compared.

2. Statistical hypothesis testing

The formal application of the method of surrogate data is expressed in the language of statistical hypothesis testing. This involves two ingredients: a null hypothesis against which observations are tested, and a discriminating statistic. The null hypothesis is a potential explanation that we seek to show is inadequate for explaining the data; and the discriminating statistic is a *number* which quantifies some aspect of the time series. If this number is different for the observed data than would be expected under the null hypothesis, then the null hypothesis can be rejected.

It is possible in some cases to derive analytically the distribution of a given statistic under a given null hypothesis, and this approach is the basis of many existing tests for nonlinearity (e.g., see refs. [20–26]). In the method of surrogate data, this distribution is estimated by direct Monte Carlo simulation. An ensemble of surrogate data sets are generated which share given properties of the observed time series (such as mean, variance, and Fourier spectrum) but are

otherwise random as specified by the null hypothesis. For each surrogate data set, the discriminating statistic is computed, and from this ensemble of statistics, the distribution is approximated.

While this approach can be computationally intensive, it avoids the analytical derivations which can be difficult if not impossible. This leads to increased flexibility in the choice of null hypotheses and discriminating statistics; in particular, the hypothesis and statistic can be chosen independently of each other. The method of surrogate data is basically an application of the "bootstrap" method of modern statistics. These methods have by now achieved widespread popularity for reasons that are well described in Efron's 1979 manifesto [27]. A more recent reference, which applies the bootstrap in a context very similar to ours is by Tsay [28].

2.1. Computing significance

Let Q_D denote the statistic computed for the original time series, and Q_{H_i} for the ith surrogate generated under the null hypothesis. Let μ_H and σ_H denote the (sample) mean and standard deviation of the distribution of Q_H.

If multiple realizations are available for the observational data, then it may be possible to compare the two distributions (observed data and surrogate) directly, using for instance the Kolmogorov–Smirnov or Mann–Whitney test, which compare the full distributions, or possibly a Student-t test which only compares their means. For the present purposes, however, we consider that only one experimental data set is available[#1], and we use a kind of t test.

[#1] Of course, it is always possible to create several realizations out of that single set by chopping up the data; we have not tried this approach, but just as the convergence of numerical algorithms like correlation dimension and Lyapunov exponent estimation are compromised by shortened data sets, so we suspect will be their power to reject a null hypothesis. This is only a suspicion, however; it would be worthwhile to compare the relative power of several short data sets versus that of one long data set.

We define our measure of "significance" by the difference between the original and the mean surrogate value of the statistic, divided by the standard deviation of the surrogate values:

$$\mathcal{S} \equiv \frac{|Q_D - \mu_H|}{\sigma_H} . \tag{1}$$

The significance is properly a dimensionless quantity, but it is natural to call the units of $\mathcal{S}$ "sigmas". If the distribution of the statistic is gaussian (and numerical experiments indicate that this is often a reasonable approximation), then the p-value is given by $p = \mathrm{erfc}(\mathcal{S}/\sqrt{2})$; this is the probability of observing a significance $\mathcal{S}$ or larger if the null hypothesis is true.

A more robust way to define significance would be directly in terms of p-values with rank statistics. For example, if the observed time series has a statistic which is in the lower one precentile of all the surrogate statistics (and at least a hundred surrogates would be needed to make this determination), then a (two-sided) p-value of $p = 0.02$ could be quoted. We have used eq. (1) for the investigations reported here because the computational effort in that case is not as severe.

2.1.1. Estimating error bars on significance

Our plots of significance include error bars; these are meant only as a rough guide and are computed assuming that the statistics are distributed as a gaussian.

We write the error bar on $\mathcal{S}$ as $\Delta\mathcal{S}$, and it is computed by standard propagation of errors methodology. Here

$$\left(\frac{\Delta\mathcal{S}}{\mathcal{S}}\right)^2 = \left(\frac{\Delta|\mu_H - \mu_D|}{|\mu_H - \mu_D|}\right)^2 + \left(\frac{\Delta\sigma_H}{\sigma_H}\right)^2$$

$$= \frac{(\Delta\mu_H)^2 + (\Delta\mu_D)^2}{(\mu_H - \mu_D)^2} + \left(\frac{\Delta\sigma_H}{\sigma_H}\right)^2 . \tag{2}$$

Now the error of the sample mean based on N observations is given by $(\Delta\mu)^2 = \sigma^2/N$, and the error of the sample standard deviation is

$(\Delta\sigma)^2 = \sigma^2/2N$, so we can write

$$\left(\frac{\Delta\mathscr{S}}{\mathscr{S}}\right)^2 = \frac{\sigma_H^2/N_H + \sigma_D^2/N_D}{(\mu_H - \mu_D)^2} + \frac{1}{2N_H}. \quad (3)$$

The absolute error bar is then given by

$$\Delta\mathscr{S} = \sqrt{(1 + \tfrac{1}{2}\mathscr{S}^2)/N_H + (\sigma_D/\sigma_H)^2/N_D}. \quad (4)$$

When only a single realization of the time series is available, we take $\sigma_D = 0$ and ignore the second term in the above equation. This reports the error bar on the significance of the specific realization.

In our numerical experiments, we use several realizations of the time series under question. However, the significance we report is not based on the collective evidence of the several, but is the average significance of each realization taken individually. The error bar in that case describes the expected error of our estimate of this average. Note that this differs from the error reported for a single realization.

2.2. *Toward a hierarchy of null hypotheses*

The null hypothesis defines the nature of the candidate process which may or may not adequately explain the data. Our null hypotheses usually specify that certain properties of the original data are preserved – such as mean and variance – but that there is no further structure in the time series. The surrogate data is then generated to mimic these preserved features but to otherwise be random. There is some latitude in choosing which features ought to be preserved: certainly mean and variance, and possibly also the Fourier power spectrum. If the raw data is discretized to integer values, then the surrogate data should be similarly discretized.

Ultimately we envision a hierarchy of null hypotheses against which time series might be compared. Beginning with the simplest hypotheses, and increasing in generality, the following sections outline some of the possibilities that we have considered.

2.2.1. *Temporally independent data*

The first (and easiest) question to answer about a time series is whether there is evidence for any dynamics at all. The null hypothesis in this case is that the observed data is fully described by independent and identically distributed (IID) random variables. If the distribution is further assumed to be gaussian, then surrogate data can be readily generated from a standard pseudorandom number generator, normalized to the mean and variance of the original data.

To test the hypothesis of IID noise *with arbitrary amplitude distribution* in an analysis of stock market returns, Schienkman and LeBaron [29] generated surrogate data by shuffling the time-order of the original time series. The surrogate data is obviously guaranteed to have the same amplitude distribution as the original data, but any temporal correlations that may have been in the data are destroyed. Breeden and Packard also used a shuffling process along with a sophisticated nonlinear predictor to prove that there was some dynamical structure to a time series of quasar data which were sampled nonuniformly in time [30].

2.2.2. *Ornstein–Uhlenbeck process*

A very simple case of non-IID noise is given by the Ornstein–Uhlenbeck process [31]. A discrete sampling of this process yields a model of the form

$$x_t = a_0 + a_1 x_{t-1} + \sigma e_t, \quad (5)$$

where e_t is uncorrelated gaussian noise of unit variance. The coefficients a_0, a_1, and σ collectively determine the mean, variance, and autocorrelation time of the time series. In fact, the autocorrelation function is exponential in this case,

$$A(\tau) \equiv \frac{\langle x_t x_{t-\tau}\rangle - \langle x_t\rangle^2}{\langle x_t^2\rangle - \langle x_t\rangle^2} = e^{-\lambda|\tau|},$$

where $\langle\ \rangle$ denotes an average over time t, and $\lambda = -\log a_1$.

To make surrogate data sets, the mean μ, variance v, and first autocorrelation $A(1)$ are estimated from the original time series; from these the coefficients are fit: $a_1 = A(1)$, $a_0 = \mu(1 - a_1)$, and $\sigma^2 = v(1 - a_1^2)$. Finally, one generates the surrogate data by iterating eq. (5), using a pseudorandom number generator for the unit variance gaussian e_t.

2.2.3. Linearly autocorrelated gaussian noise

We can generalize the above null hypothesis by extending eq. (5) to arbitrary order. This leads to the hypothesis that is generally associated with linearity. We emphasize that we are discussing linear *gaussian* processes here (see Tong [26, pp. 13, 14] for a brief description of some of the surprising properties of linear nongaussian processes); Section 2.2.4 describes one approach toward a nongaussian null hypothesis. The model is described by fitting coefficients a_k and σ to a process

$$x_t = a_0 + \sum_{k=1}^{q} a_k x_{t-k} + \sigma e_t \,, \tag{7}$$

which mimics the original time series in terms of mean, variance, and the autocorrelation function for delays of $\tau = 1, \ldots, q$. This is an autoregressive (AR) model; a more general model includes a moving average (MA) of time delayed noise terms as well, and the combination is called an ARMA model. For large enough q, the models are essentially equivalent. The null hypothesis in this case is that all the structure in the time series is given by the autocorrelation function, or equivalently, by the Fourier power spectrum.

One algorithm for generating surrogate data under this null hypothesis is again to iterate eq. (7), where the coefficients have been fit to the original data. We describe an alternative algorithm in section 2.4.1 which involves randomizing the phases of a Fourier transform. (To our knowledge, this algorithm was first suggested in this context by Osborne et al. [5], and independently in refs. [15, 32].) The alternative algorithm generates surrogate data which by construction has the same Fourier spectrum as the original data. While the two algorithms are essentially equivalent, we use the Fourier transform method because it is numerically stabler. If the values of the coefficients in eq. (7) are mis-estimated slightly, it is possible that iterating the equation will lead to a time series which diverges to infinity; this is particularly problematic if the raw time series is nearly periodic or highly sampled continuous data.

We remark that this is the null hypothesis that is associated with residual-based tests for nonlinearity. For instance, see refs. [22–24, 33, 34]. In these tests, a model of the form of eq. (7) is fit to the data, and the residuals

$$\epsilon_t = x_t - \left(a_0 + \sum_{k=1}^{q} a_k x_{t-k} \right) \tag{8}$$

are tested against a null of temporally independent noise. In ref. [19], we argue that it is usually preferable to use the method of surrogate data on the raw data directly, rather than working with residuals.

2.2.4. Static nonlinear transform of linear gaussian noise

One way to generalize the above null hypothesis to cases where the data is nongaussian is to suppose that although the dynamics is linear, the observation function may be nonlinear. In particular, we hypothesize that there exists an "underlying" time series $\{y_t\}$, consistent with the null hypothesis of linear gaussian noise, and an observed time series $\{x_t\}$ given by

$$x_t = h(y_t) \,. \tag{9}$$

Since x_t depends only on the current value of y_t and not on derivatives or past values, the filter $h(\,)$ is said to be "static" or "instantaneous". To permit the generation of surrogate data, we must further assume (as part of the null hypothesis)

that the observation function $h(\ )$ is effectively invertible.

In section 2.4.3, an algorithm for generating surrogate data corresponding to this null hypothesis is described. Its effect is to shuffle the time-order of the data but in such a way as to preserve the linear correlations of the underlying time series $y_t = h^{-1}(x_t)$. One advantage of shuffling over, for example, a smooth fit to the function $h(\ )$, is that any discretization that was present in the original data will be reflected in the surrogate data.

Note that time series in this class are strictly speaking nonlinear, but that the nonlinearity is not in the dynamics. Most conventional tests for nonlinearity would (correctly) conclude that the time series is nonlinear, but would not indicate whether the nonlinearity was in the dynamics or in the amplitude distribution. By using surrogate data tailored to this specific null hypothesis, it becomes possible to make such fine distinctions about the nature of the dynamics.

2.2.5. *More general null hypotheses*

Eventually, we would like to extend this list to consider more general cases. A natural next step is a null hypothesis that the dynamics is a noisy limit cycle. Such time series cannot be described by a linear process, even if viewed through a static nonlinear transform. Yet it is often of great interest, particularly in systems driven by seasonal cycles, to determine the nature of the inter-seasonal variation.

There is another large class of nonlinear stochastic processes which are not predictable even in the mean; among these are the conditional heteroscedastic models (for which the variance is conditioned on the past, but not the mean) in favor among economists. While there is definite nonlinear structure in these time series, it is not manifested in enhanced predictability by nonlinear models. (For instance, it may be possible to predict the magnitude $|x_t|$ from past values of x, but not the sign.)

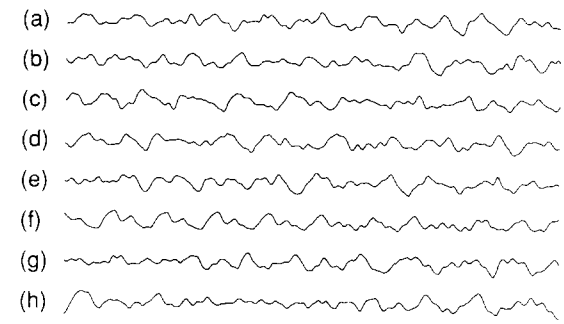

Fig. 1. Shown is a time series from the Mackey–Glass equation with $\tau = 30$, which is known to be low-dimensional and chaotic, and seven surrogate time series generated by the WFT algorithm. It is often not obvious by eye which is the actual data set and which are the surrogates. In this case it is series (f) which is the real one.

2.3. *Battery of discriminating statistics*

The method of surrogate data can in principle be used with virtually any discriminating statistic. Formally, all that is required to reject a null hypothesis is that the statistic have a different distribution for the data than for the surrogates. However, the method is more useful if the statistic actually provides a good estimate of a physically interesting quantity; in that case, one may not only formally reject a null hypothesis, but also informally characterize the nature of the nonlinearity.

Since we were motivated by the possibility that the underlying dynamics may be chaotic, our original choices for discriminating statistics were the correlation dimension, Lyapunov exponent, and forecasting error. Ideally, dimension counts degrees of freedom, Lyapunov exponent quantifies the sensitivity to initial conditions, and forecasting error tests for determinism. One of the ultimate aims in this project is to understand the conditions in which one or the other of these methods will be more effective.

We should remark that a danger in using a battery of statistics is that one of them, by chance, will show up as significant. This effect can be formally accounted for by keeping strict count of the number of tests used, and increasing

the threshold of significance accordingly. The formal approach tends to be more conservative than necessary, since the tests are not really independent of each other, but it is still a recommended practice to maintain a reasonably high threshold of significance.

2.3.1. Correlation dimension, ν

Dimension is an exponent which characterizes the scaling of some bulk measure with linear size. A number of algorithms are available [17, 35] for estimating the dimension of an underlying strange attractor from a time series; we chose a box-assisted variation [36] (see Grassberger [37] for an elegant alternative) of the Grassberger–Procaccia–Takens algorithm [38–40] to compute a correlation integral, and the best estimator of Takens [12] for the dimension itself. To compute a dimension, it is necessary to choose some range of sizes over which the scaling is to be estimated. The Takens estimator requires only an upper cutoff size; we used one-half of the rms variation in the time series for this value. (See Ellner [41] for an estimator that takes both an upper and a lower cutoff.)

We will concede that this choice is a bit arbitrary; one might prefer a more sophisticated algorithm for choosing a good scaling range. L. Smith (personal communication) has suggested choosing the range "by eye" for the raw data and then keeping this range for the surrogates. From the point of view of the formal test, it does not really matter, but if we are to ask for insight as well as a rejected null, then it is important to use a good dimension estimator. In the $N \to \infty$ limit, the estimator we describe will not converge to the actual precise dimension of the attractor, but we note that it will converge fairly rapidly to a number which is often reasonably close to actual dimension (of course, one can always contrive counterexamples); in particular, it will properly indicate low-dimensionality when it sees it. While we do not claim that this is the optimal dimension estimator, we believe that it is a useful one.

2.3.2. Forecasting error, ϵ

A system is deterministic if its future can be predicted. A natural statistic in this case is some average of the forecasting errors obtained from nonlinear modeling. The method we use entails first splitting the time series into a fitting set of length N_f, and a testing set of length N_t, with $N_f + N_t = N$, the length of the time series; then fitting a local linear model [42] to the fitting set, locality given by the number of neighbors k; and finally, using this model to forecast the values in the testing set, and comparing them with the actual values.

The prediction error $e_t = x_t - \hat{x}_t$ is the difference between the actual value of x and the predicted value, $\hat{x}$; we define our discriminating statistic as the log median absolute prediction error.

Several modeling parameters must be chosen, including the partitioning of the data set into fitting (N_f) and testing (N_t) segments, the number of steps ahead to predict (T), and number of neighbors (k) used in the local linear fit. We arbitrarily chose to divide the fitting and testing sets equally, with $N_f = N_t = \frac{1}{2}N$, and to predict one step ahead, so $T = 1$. For oversampled continuous data, a larger T would be more appropriate. The choice of k is also important. For the results in this article, we set $k = 2m$, which is twice the minimum number needed for a fit, but we note that this is often not optimal. Indeed, Casdagli [43, 44] has advocated sweeping the parameter k in a local linear forecaster as an exploratory method to look for nonlinearity in the first place.

2.3.3. Estimated Lyapunov exponent, λ

Following standard practices [45–47], we compute Lyapunov exponents by multiplying Jacobian matrices along a trajectory, with the matrices computed by local linear fits, and we use QR decomposition to maintain orthogonality.

We have found that numerical estimation of even the largest Lyapunov exponent can be problematic in the presence of noise. Indeed, for

our surrogate data sets, for which the linear dynamics is contracting, we often obtain positive Lyapunov exponents. This indicates that our Lyapunov exponent estimator (which, as we have described, is fairly standard) is seriously flawed, something we might not have noticed had we not tested with linear stochastic time series. We are aware of at least one group whose Lyapunov exponent estimator explicitly considers the effects of noise [48–50]. While our estimator is arguably still useful as a statistic which formally distinguishes original data from surrogate data, it would be better to use a discriminating statistic which correctly quantifies some feature of the dynamics. For that reason, we have avoided using the Lyapunov exponent estimator in this article.

2.3.4. *Other discriminating statistics*

We have found that using the correlation integral ($C(r)$ for some value of r) directly as a discriminating statistic generally provides a more powerful discrimination than the estimated dimension itself, but of course it is less useful as an informal tool. L. Smith (personal communication) has suggested using a statistic which characterizes the linearity of a $\log C(r)$ versus $\log r$ curve. We have also considered but not implemented two-sided forecasting – predicting x_t from the past *and* future values: $x_{t-1}, \ldots$; $x_{t+1}, \ldots$, instead of the usual forecasting which uses only the past (this was inspired by the simple noise reduction technique suggested by Schreiber and Grassberger [51]). In our forecasting, we are careful to distinguish the "training" set from the "testing" set, so that the forecasting statistic is an out-of-sample error; but the in-sample fitting error may also suffice as a discriminating statistic. We have found that the BDS test [33], which was designed to test for any temporal correlation at all – linear or nonlinear, can readily be extended to test other null hypotheses; we use the same statistic, but we do not pre-whiten the data, and instead of relying on an analytical derivation of the distribution function, we use surrogate data. Higher and cross moments provide another class of discriminating statistic; in fact, many of these are the basis of traditional tests for nonlinearity in a time series (e.g., see refs. [22–24]). We have found that a simple skewed difference statistic, defined by $Q = \langle (x_{t+m} - x_t)^3 \rangle / \langle (x_{t+m} - x_t)^2 \rangle$, is both rapidly computable and often quite powerful. Informally, this statistic indicates the asymmetry between rise and fall times in the time series. The most direct example we know is due to Brock, Lakonishok, and LeBaron [52], who used technical trading rules as discriminating statistics for financial data; here there is no difficulty interpreting the informal meaning of the statistic: it is how much money you should have made using that rule in that market.

2.4. *Algorithms for generating surrogate data*

In this section, we describe algorithms we use for generating surrogate data. The first two are consistent with the hypothesis of linearly correlated noise described in section 2.2.3, and the third adjusts for the possibility of a static nonlinear transform as discussed in section 2.2.4.

2.4.1. *Unwindowed Fourier transform (FT) algorithm*

This algorithm is based on the null hypothesis that the data come from a linear gaussian process. The surrogate data are constructed to have the same Fourier spectra as the raw data. The algorithm is described in more detail in ref. [19], but we briefly note the main features. First, the Fourier transform is computed for positive and negative frequencies $f = 0, 1/N, 2/N, \ldots, 1/2$, and without the benefit of windowing. Although windowing is generally recommended when it is the power spectrum which is of ultimate interest [53], we originally chose not to use windowing because what we wanted was for the real and surrogate data to have the same power spectrum; we were not concerned with the spectrum, *per se*. The Fourier transform has a complex am-

plitude at each frequency; to randomize the phases, we multiply each complex amplitude by $e^{i\phi}$, where ϕ is independently chosen for each frequency from the interval $[0, 2\pi]$. In order for the inverse Fourier transform to be real (no imaginary components), we must symmetrize the phases, so that $\phi(f) = -\phi(-f)$. Finally, the inverse Fourier transform is the surrogate data.

One limitation of this algorithm is that it does not reproduce "pure" frequencies very well. What happens is that nearby frequencies in Fourier space are "contaminated" and then because their phases are randomized, they end up "beating" against each other and producing spurious low-frequency effects. (We are grateful to S. Ellner for pointing this out to us.) This may not be too surprising since it is difficult to make a linear stochastic process with a long coherence time. Put another way, the time series should not only be much larger than the dominant periodicities but also much longer than the coherence time of any given frequency if one is to try and model it with a linear stochastic process.

A second problem, which is most evident for highly sampled continuous data, is that spurious high frequencies can be introduced. This can be understood as an artifact of the Fourier transform which assumes the time series is periodic with period N. This means that there is a jump-discontinuity from the last to the first point. We recommend tailoring the length N of the data set so that $x[0] \approx x[N-1]$. This should not be a problem if the time series is stationary and much longer than its dominant frequency. We have done this for the experimental results in this article.

2.4.2. *Windowed Fourier transform (WFT) algorithm*

The problem of spurious high frequencies can also be addressed by windowing the data before taking the Fourier transform. The time series is multiplied by a function $w(t) = \sin(\pi t/N)$ which vanishes at the endpoints $t = 0$ and $t = N$. This suppresses the jump discontinuity from the last to the first point, and seems to effectively get rid of the high frequency effect. However, it also introduces a spurious low-frequency from the power spectrum of $w(t)$ itself. We have done experiments where we simply set the magnitude of the offending frequency ($f = 1/N$) to zero; this seems to work well for stationary time series, but if there is significant power at that frequency in the original data, it too will be suppressed.

2.4.3. *Amplitude adjusted Fourier transform (AAFT) algorithm*

The algorithm in this section generates surrogate data sets associated with the null hypothesis in section 2.2.4, that the observed time series is a monotonic nonlinear transformation of a linear gaussian process. The idea is to first rescale the values in the original time series so they are gaussian. Then the FT or WFT algorithm can be used to make surrogate time series which have the same Fourier spectrum as the rescaled data. Finally, the gaussian surrogate is then rescaled back to have the amplitude distribution as the original time series.

Denote the original time series by $x[t]$, with $t = 0, \ldots, N-1$. The first step is to make a gaussian time series $y[t]$, where each element is generated independently from a gaussian pseudorandom number generator. Next, we re-order the time sequence of the gaussian time series so that the ranks of both time series agree; that is, if $x[t]$ is the nth smallest of all the x's, then $y[t]$ will be the nth smallest of all the y's. Therefore, the re-ordered $y[t]$ is a time series which "follows" the original time series $x[t]$ and which has a gaussian amplitude distribution. Using the FT or WFT algorithm, a surrogate, call it $y'[t]$, of the gaussian time series can be created. If the original time series $x[t]$ is time re-ordered so that it follows $y'[t]$ in the sense that the ranks agree, then the time-re-ordered time series provides a surrogate of the original time series which matches its amplitude distribution. Further, the

"underlying" time series ($y[t]$ and $y'[t]$) are gaussian and have the same Fourier power spectrum.

3. Experiments with numerical time series

To properly gauge the utility of the surrogate data approach will eventually require many tests with data from both numerical and laboratory experiments. In this section we illustrate several aspects of the method with data whose underlying dynamics is known. In the next section, we consider several examples with real data.

3.1. Linear gaussian data

First, we note that a time series which actually is generated by a linear process should by construction give a negative result (that is, the null hypothesis should *not* be rejected). We checked this by generating some time series with two simple linear processes, a moving average

$$x_t = e_t + ae_{t-1} \tag{10}$$

and an autoregressive

$$x_t = ax_{t-1} + e_t . \tag{11}$$

We used an embedding dimension $m = 3$ and computed correlation dimension from $N = 4096$ points. The "correct" dimension for both processes is $\nu = m = 3$, though as fig. 2 shows, the estimates were always biased low. The bias increases for data which are more highly autocorrelated ($|a|$ larger) but the point we wish to make is that the bias is the same for the original data and for the surrogates. The null hypothesis is not rejected.

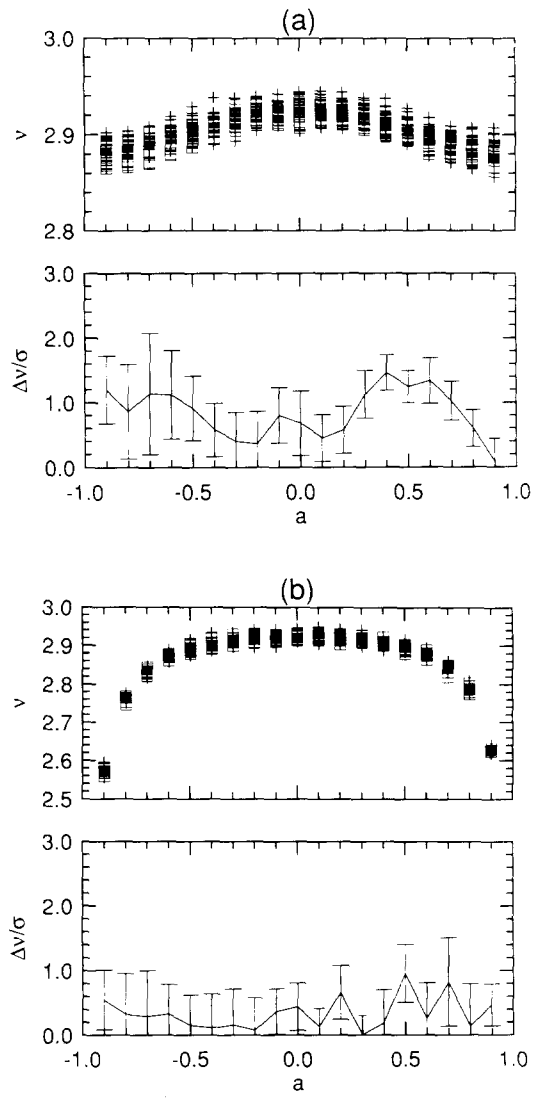

Fig. 2. Significance of evidence for nonlinearity for linear gaussian time series generated by (a) a moving average process, and (b) an autoregressive process. The coefficient in each case is a. The estimated dimension is shown for five realizations of the linear process (□) and thirty realizations of surrogate data (+). Note that the dimension does not distinguish the original from the surrogate data. The value we obtain for significance is shown in the lower panels and in neither case is significant.

3.2. Variation with number of data points and complexity of attractor

Using the FT algorithm, we showed in ref. [19] that increasing the number of points in a time series increases the significance with which nonlinearity can be detected in a time series that is known to be chaotic; and increasing the complexity of the chaotic time series decreases the ability to distinguish from linearity. This basic

point is illustrated again here using the AAFT algorithm (see fig. 3); while the significance is not as large using this more general null hypothesis, the qualitative behavior is the same. Time series are generated by summing n independent trajectories of the Hénon map [54]; such time series will have a dimension nv where $v \approx 1.2$ is the dimension of a single Hénon trajectory. For the largest data sets, with $N = 8192$ points, our dimension estimator obtained correlation dimensions of 1.215 ± 0.008, 2.279 ± 0.014, 3.48 ± 0.02, and 4.81 ± 0.06 using embedding dimensions $m = 3$, 4, 5, and 6, for $n = 1$, 2, 3, and 4, respectively.

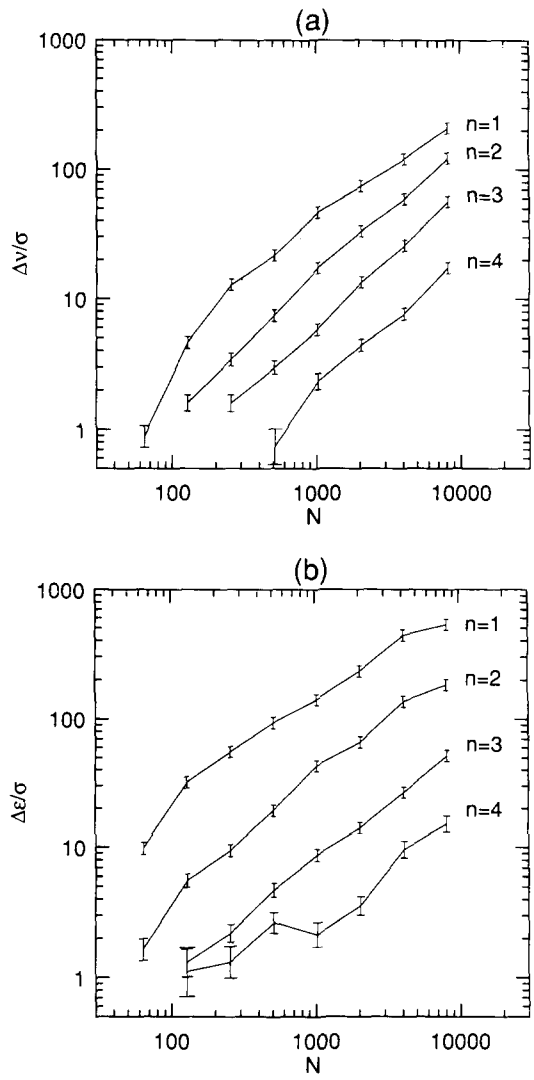

Fig. 3. Using the AAFT algorithm to generate surrogate data, the significance as a function of the number N of data points is computed for time series obtained by summing n independent trajectories of the Hénon map. For both (a) correlation dimension and (b) forecasting error, the significance increases with number of data points and decreases with the complexity of the system.

3.3. Effect of observational and dynamical noise

To test whether nonlinear determinism can be detected even when it is mixed with noise, we added both dynamical (η) and observational (ε) noise to the cosine map: $y_t = \lambda \cos(\pi y_{t-1}) + \eta_t$; $x_t = y_t + \varepsilon_t$. We chose a value $\lambda = 2.8$ which is in the chaotic regime when the external noise is zero. (The cosine map was used instead of the Hénon map because it does not "blow up" in the presence of too much dynamical noise.) In fig. 4,

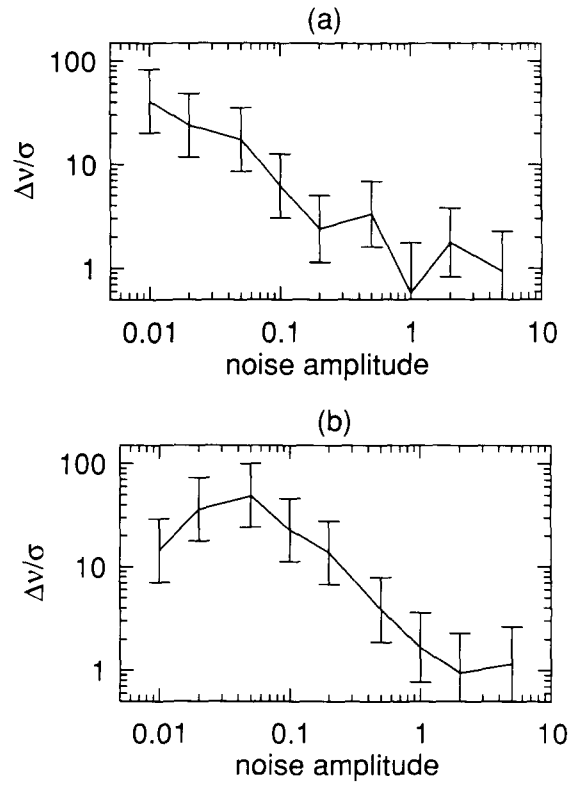

Fig. 4. Effect of noise on significance for a short time serie of $N = 512$ points, derived from the cosine map with $\lambda = 2.8$ (a) observational noise; (b) dynamical noise.

we plot significance as a function of noise level for both dynamical and observational noise. As expected, significance decreases with increasing noise level, though we remark that the nonlinearity is still observable even with considerable noise. In the absence of noise, the rms amplitude of the signal is 0.36; thus we are able to detect significant nonlinearity even with a signal to noise ratio of one, using a time series of length $N = 512$. We also note that the decrease in significance with increased dynamical noise is not always monotonic; low levels of dynamical noise can make the nonlinearity more evident.

3.4. Continuous data

In most experiments, data is better described as a flow than a map. Although there is a formal equivalence, data which arise from processes that are continuous in time are often sampled at a much faster rate than is characteristic of the underlying dynamics. For these data sets, the effects of autocorelation can be quite large, and the importance of testing against a null hypothesis that includes autocorrelation becomes paramount.

We illustrate this point with numerical experiments on data obtained from the Mackey–Glass differential delay equation [55]

$$\frac{dx}{dt} = -bx(t) + \frac{ax(t-\tau)}{1+[x(t-\tau)]^{10}}, \tag{12}$$

with $a = 0.2$, $b = 0.1$, and $\tau = 30$. Grassberger and Procaccia [39] compute a correlation dimension of $\nu \approx 3.0$ for these parameters.

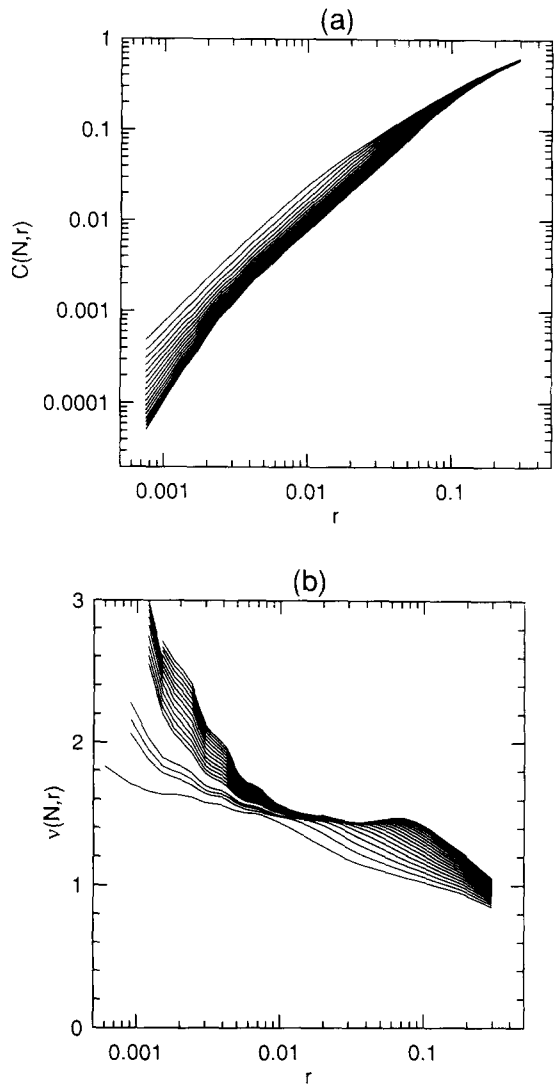

Fig. 5. (a) Correlation integral $C(N, r)$ for $N = 4096$ points and embedding dimensions $m = 3, \ldots, 19$ from oversampled Mackey–Glass data. (b) Estimated correlation dimension ν according to Takens estimator as a function of cutoff r.

3.4.1. A poor embedding

We oversample the data ($\Delta t = 0.1$) and use a deliberately poor embedding strategy – straight time-delay coordinates with a lag time of one sample period. We estimate correlation dimension with $N = 4096$ points and compute distances between all pairs of distinct vectors (despite the advice in refs. [2, 56]). Fig. 5 shows the correlation integral and estimated dimension as a function of the upper cutoff value R. There is about a decade of roughly constant slope, which might be taken to indicate convergence to a low correlation dimension.

For this example, the dimension statistic was computed as the Takens best estimator [12] at an upper cutoff of $R = 0.02$. (For comparison, the

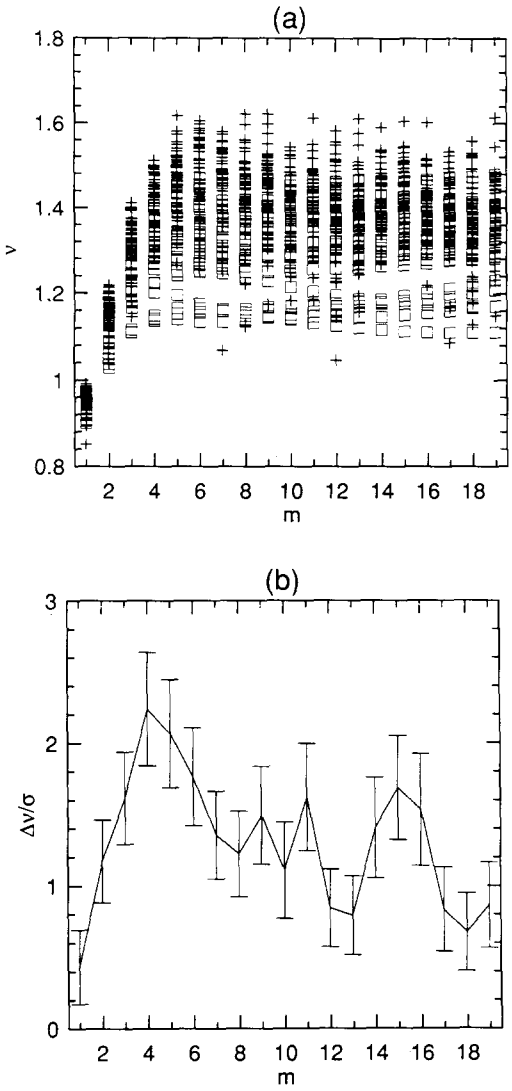

Fig. 6. (a) Estimated correlation dimension versus embedding dimension for oversampled ($\Delta t = 0.1$) Mackey–Glass data (□) and for surrogates generated using the WFT algorithm (+). (b) Significance of nonlinearity in no case exceeds three sigmas.

RMS value for this time series is $R_{\text{rms}} = 0.25$.) Fig. 6a shows an apparent convergence of the estimated dimension as a function of embedding dimension. A naive interpretation of this figure is that the time series arises from a low-dimensional strange attractor. However, as fig. 6a shows, the surrogate data also converge to a low dimension; the convergence is evidently an artifact of the autocorrelation. Indeed, fig. 6b shows that the dimension statistic in this case does not even provide evidence for nonlinearity.

3.4.2. *A better embedding*

From the same Mackey–Glass process, we recompute correlation dimension and the signifi-

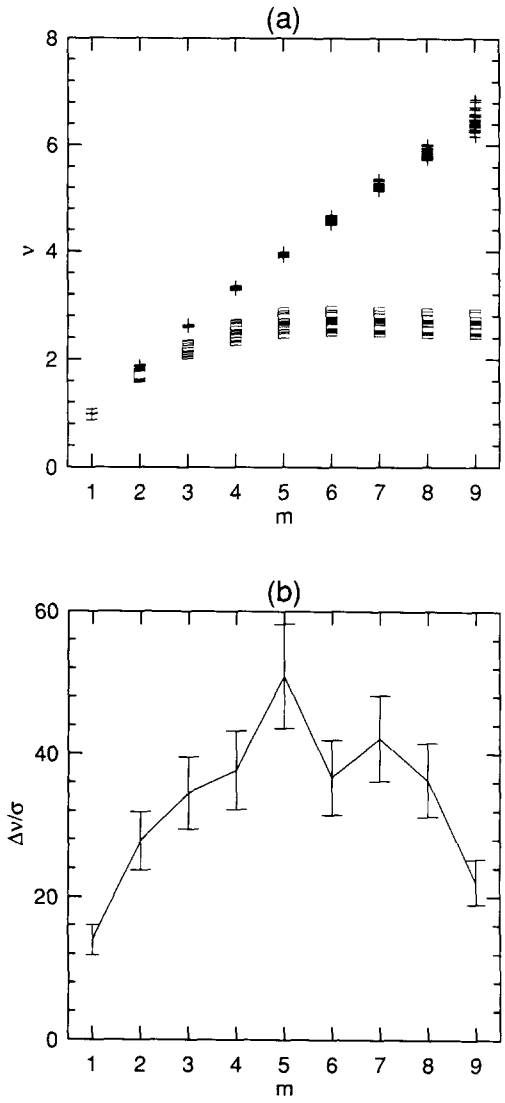

Fig. 7. Same as previous figure, except that a better embedding and a better algorithm were used for estimating the dimension. Not only is the evidence for nonlinearity extremely significant in this case, but it is also evident that the process is low-dimensional.

cance of evidence for nonlinearity using a better (though probably still not optimal) choice of embedding. We sample at a much lower rate, $\Delta t = 3.0$, and again use straight delay coordinates with lag time of one sample period. We estimate the correlation dimension as described in section 2.3.1 with $N = 4096$ points, and we avoid pairs of points which are closer together in time than one hundred sample periods. In fig. 7, we see that the evidence for nonlinearity is extremely significant. Indeed, we also see positive evidence of low-dimensional behavior (the estimated dimension ν converges with m) which we know is not an artifact of autocorrelation.

4. Examples with real data

We report some results on experimental time series from several sources. These results should be taken as illustrative, and not necessarily typical of the class which they represent. In particular, we have not yet attempted to "normalize" our findings with others that have previously appeared in the literature.

4.1. *Rayleigh–Benard convection*

Data from a mixture of ^{3}He and superfluid ^{4}He in a Rayleigh–Benard convection cell [57] provides an example where the evidence for nonlinear structure is extremely significant. The significance as obtained with the dimension and forecasting statistics from a time series of $N = 2048$ points are shown in fig. 8. Further, the dimension statistic indicates that the flow is in fact low-dimensional; while the measured dimension of 3.8 may be due to an artifact of some kind, we are at least assured that it is not an artifact of autocorrelation or of nongaussian amplitude distribution. Farmer and Sidorowich [42] used this data to demonstrate the enhanced predictability using nonlinear rather than linear predictors.

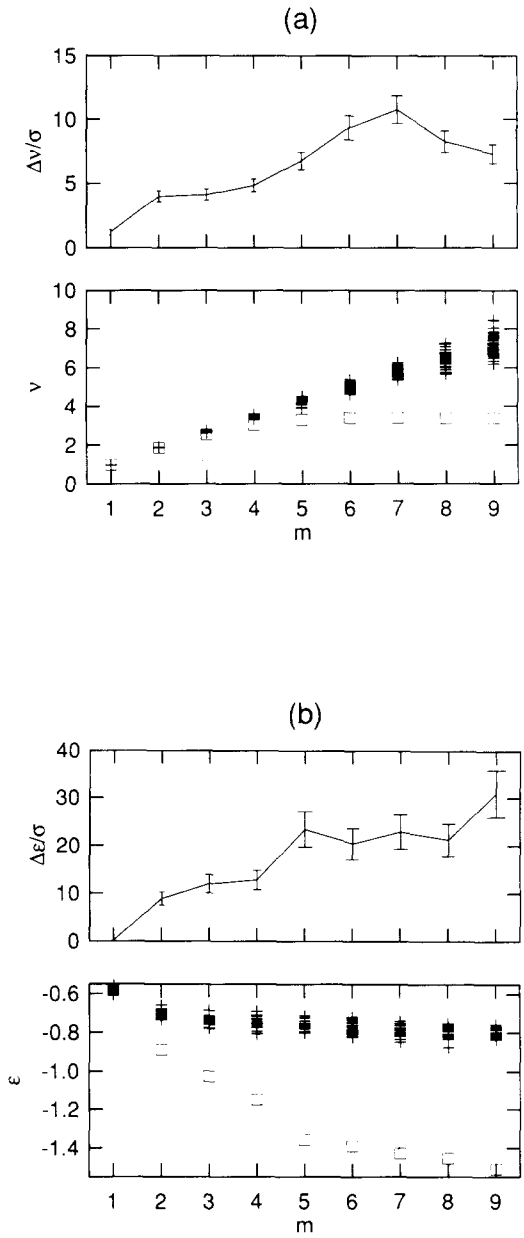

Fig. 8. Data from a fluid convection experiment exhibits very significant nonlinear structure, using (a) dimension, and (b) forecasting error. The top panel in these figures show the significance, measured in "sigmas", and the bottom panel shows the values of the statistics, with squares (□) for the original data and pluses (+) for the AAFT-generated surrogates. Both panels plot these statistics against the embedding dimension m. Not only is the evidence for nonlinear structure statistically significant, but the estimated dimension of about $\nu = 3.8$ suggests that the underlying dynamics is in fact low-dimensional chaos.

4.2. The human electroencaphalogram (EEG)

The electroencephalogram (EEG) is to the brain what the electrocardiogram (EKG) is to the heart. It has become a widely used tool for the monitoring of electrical brain activity, and its potential for diagnosis is still being explored. A number of researchers have applied the methods that were developed for the analysis of chaotic time series to EEG time series. While it was hoped that the characterization of deterministic structure in EEG would eventually lead to insights about the workings of the brain, the shorter term goal was to use the nonlinear properties of the time series as a diagnostic tool [58, 59].

Although we feel a more systematic survey is in order, we have not examined any EEG data which gives positive evidence of low-dimensional chaos. However, we have found examples where nonlinear structure was evident. We present here two cases, one positive and one negative. The two time series are from the same individual, eyes closed and resting; one is from a probe at the left occipital (O1), and the other from the left central (C3) part of the skull. The sampling rate is 150 Hz, and $N = 2048$ time samples are taken. The two time series are not necessarily contemporaneous. Using the dimension statistic, the first data set shows no significant evidence for nonlinearity, but the second data set exhibits about eight sigmas. Even in the significant case, we do not see any evidence that the time series is in fact low-dimensional (the correlation dimension ν does not converge with increasing embedding dimension m). We are formally able to reject the null hypothesis that the data arise from a linear stochastic process, but by comparing the surrogate data to the real data, we see no reason to expect that the "significant" data arises from a low-dimensional chaotic attractor.

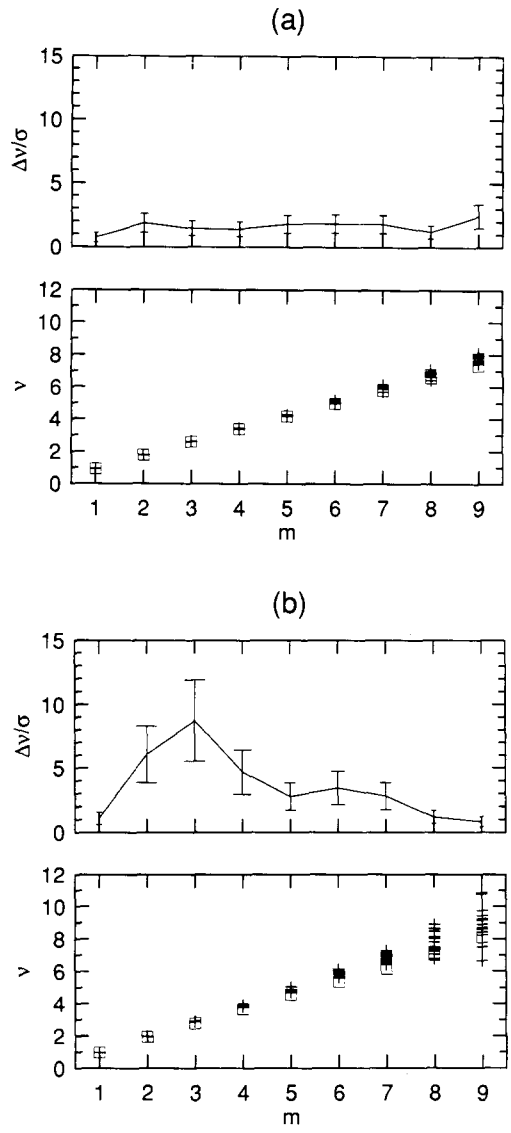

Fig. 9. Data from two electroencephalogram (EEG) time series. Using the dimension statistic, the first (a) shows no nonlinear structure, while the second (b) exhibits significant nonlinear structure at the eight sigma level. The evidence for low-dimensional chaos, however, is weak, since the estimated dimension increases almost as rapidly with embedding dimension for the original time series as it does for the surrogates.

4.3. The sunspot cycle

Our final example is the well known and much studied eleven year sunspot cycle [44, 60–66]. First, we used the FT algorithm for generating surrogate data, but we were careful to use a length of time series ($N = 287$) for which the first and last data point both corresponded to minima; this avoids introducing the spurious high

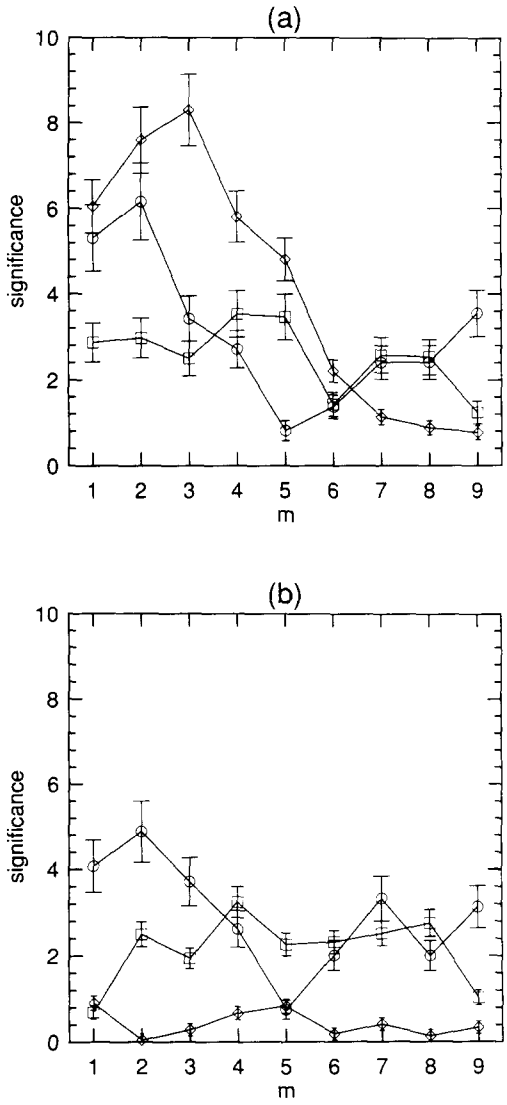

Fig. 10. Significance of nonlinearity in the annual sunspot series; (a) against the null hypothesis of linear gaussian noise (surrogates generated by FT algorithm), and (b) the null hypothesis of amplitude corrected linear gaussian noise (surrogates generated by AAFT algorithm). For both plots, the discriminating statistics are estimated dimension (◇), log median prediction error (□), and the skew statistic described in the text (○).

frequencies that we discussed in section 2.4.1. As fig. 10a shows, it is possible to quite confidently reject the null hypothesis of linear gaussian noise; this is in agreement with the numerous authors who obtained better agreement using nonlinear models instead of linear models. However, when we expand the null hypothesis to include a static nonlinear observation of an underlying linear gaussian process, the evidence for dynamical nonlinear structure is less dramatic. Using the dimension statistic, there is no significance; the prediction statistic gives that the evidence is just significant; the cubed difference statistic $Q = \langle (x_{t+m} - x_t)^3 \rangle / \langle (x_{t+m} - x_t)^2 \rangle$, which is a measure of the time irreversibility of the data, provides a more significant rejection of the hypothesis of static nonlinear filter of an underlying linear process.

5. Comparison to other work

Numerous authors have carefully compared their dimension estimates for real data against similar estimates for white noise. A few have extended this informal control to other forms of correlated noise. Grassberger [2] showed that a reported dimension for climate data could be reproduced with data from an Ornstein–Uhlenbeck process. Osborne et al. [5], criticized the Grassberger–Procaccia algorithm on the basis that the low dimension it gave to nonstationary data on ocean currents it also gave to data generated by randomizing the phases of the Fourier transform. Kaplan and Cohen [32] argued that fibrillation was not usefully described as chaotic, again since randomly phased data gave similar dimensions. Smith [67] has used the FT algorithm to generate surrogates which are used to assess the predictability of geophysical time series. Weiss [62], described a comparison of the sunspot time series against a particular stochastic model. Brock et al. [52] used technical trading rules to distinguish stock market data from surrogates generated by several stochastic models. And Ellner [68] showed that a variety of "plausible alternatives" might adequately explain measles and chickenpox data, despite earlier claims of chaos.

Brock and coworkers in particular [33, 52,

69–71], and the economics community in general [29, 34, 72, 73], have been extremely active in the development of statistical tools for time series analysis. While the choice of null hypotheses for financial time series tends to be different than for more physical time series (autocorrelation plays a lesser role, for example), the overall methodologies are quite similar. Classical statisticians [20–25, 28] have long considered tests for nonlinearity, and are becoming increasingly aware of low-dimensional chaos (just as physicists are becoming increasingly aware of the importance of the statistical approach); we cite Tong [26] as *the* review which most neatly and comprehensively ties these two fields together.

6. Conclusion

In this article, we have described an approach for evaluating the statistical significance of evidence for nonlinearity in a stationary time series. The test properly fails to find nonlinear structure in linear stochastic systems, and correctly identifies nonlinearity in several well-known examples of low-dimensional chaotic time series, even when contaminated with dynamical and observational noise. We illustrated the method with several experimental data sets, and confirmed the evidence for nonlinear structure in some systems, while failing to see such structure in other time series.

Acknowledgements

We are pleased to acknowledge Martin Casdagli, Xiangdong He, Willian Brock, Blake LeBaron, Steve Ellner, Peter Grassberger, Mark Berge, Tony Begg and Bette Korber for useful discussions. We are especially grateful to Lenny Smith for illuminating disagreements and a careful reading of the manuscript. We also thank Bob Ecke for providing convection data, and Paul Nuñez and Arden Nelson for providing EEG data. This work was partially supported by the National Institute for Mental Health under grant 1-R01-MH47184-01, and performed under the auspices of the Department of Energy.

References

[1] C. Nicolis and G. Nicolis, Nature 311 (1984) 529.
[2] P. Grassberger, Nature 323 (1986) 609.
[3] C. Nicolis and G. Nicolis, Nature 326 (1987) 523.
[4] P. Grassberger, Nature 326 (1987) 524.
[5] A.R. Osborne, A.D. Kirwin, A. Provenzale and L. Bergamasco, Physica D 23 (1986) 75.
[6] G. Mayer-Kress, in: Directions In Chaos, Vol. I, ed. B.-L. Hao (World Scientific, Singapore, 1988), pp. 122–147.
[7] A.R. Osborne and A. Provenzale, Physica D 35 (1989) 357.
[8] A. Provenzale, A.R. Osborne R. Soj, Physica D 47 (1991) 361.
[9] J. Theiler, Phys. Lett. A 155 (1991) 480.
[10] J.B. Ramsey and J.-J. Yuan, Phys. Lett. A 134 (1989) 287.
[11] J. Theiler, Phys. Rev. A 41 (1990) 3038.
[12] F. Takens, in: Dynamical Systems and Bifurcations (Groningen, 1984), eds. B.L.J. Braaksma, H.W. Broer and F. Takens, Lecture Notes in Mathematics, Vol. 1125 (Springer, Berlin, 1985), pp. 99–106.
[13] W.E. Caswell and J.A. Yorke, in: Dimensions and Entropies in Chaotic Systems – Quantification of Complex Behavior, Springer Series in Synergetics, Vol. 32 (Springer, Berlin, 1986), pp. 123–136.
[14] J. Holzfuss and G. Mayer-Kress, in: Dimensions and Entropies in Chaotic Systems – Quantification of Complex Behavior, Springer Series in Synergetics, Vol. 32 (Springer, Berlin, 1986), pp. 114–122.
[15] J. Theiler, Quantifying Chaos: Practical Estimation of the Correlation Dimension, Ph.D. thesis (Caltech, 1988).
[16] M. Möller, W. Lange, F. Mitschke, N.B. Abraham and U. Hübner, Phys. Lett. A 138 (1989) 176.
[17] J. Theiler, J. Opt. Soc. Am. A 7 (1990) 1055.
[18] R.L. Smith, in: Nonlinear Modeling and Forecasting, eds. M. Casdagli and S. Eubank, of SFI Studies in the Sciences of Complexity, Vol. XII (Addison-Wesley, Reading, MA, 1992) pp. 115–136.
[19] J. Theiler, B. Galdrikian, A. Longtin, S. Eubank and J.D. Farmer, in: Nonlinear Modeling and Forecasting, eds. M. Casdagli and S. Eubank, SFI Studies in the Sciences of Complexity, Vol. XII (Addison-Wesley, Reading, MA, 1992) pp. 163–188.
[20] T. Subba Rao and M.M. Gabr, J. Time Series Anal. 1 (1980) 145.
[21] M.J. Hinich, J. Time Series Anal. 3 (1982) 169.
[22] A.I. McLeod and W.K. Li, J. Time Series Anal. 4 (1983) 269.

[23] D.M. Keenan, Biometrika 72 (1985) 39.
[24] R.S. Tsay, Biometrika 73 (1986) 461.
[25] R.S. Tsay, Stat. Sin. 1 (1991) 431.
[26] H. Tong, Non-linear Time Series: A Dynamical System Approach (Clarendon Press, Oxford, 1990).
[27] B. Efron, SIAM Rev. 21 (1979) 460.
[28] R.S. Tsay, Appl. Stat. 41 (1992) 1.
[29] J.A. Scheinkman and B. LeBaron, J. Business 62 (1989) 311.
[30] J.L. Breeden and N.H. Packard, Nonlinear analysis of data sampled nonuniformly in time, Physica D 58 (1992) 273, these Proceedings.
[31] G.E. Uhlenbeck and L.S. Ornstein, Phys. Rev. 36 (1930) 823; reprinted in: Noise and Stochastic Processes, ed. N. Wax (Dover, New York, 1954).
[32] D.T. Kaplan and R.J. Cohen, Circulation Res. 67 (1990) 886.
[33] W.A. Brock, W.D. Dechert and J. Scheinkman, A test for independence based on the correlation dimension, Social Systems Research Institute, University of Wisconsin at Madison, technical report 8702 (1986).
[34] T.-H. Lee, H. White and C.W.J. Granger, Testing for neglected nonlinearity in time series models: A comparison of neural network methods and alternative tests, J. Econometrics, to appear.
[35] E.J. Kostelich and H.L. Swinney, in: Chaos and Related Natural Phenomena, eds. I. Procaccia and M. Shapiro (Plenum, New York, 1987), p. 141.
[36] J. Theiler, Phys. Rev. A 36 (1987) 4456.
[37] P. Grassberger, Phys. Lett. A 148 (1990) 63.
[38] P. Grassberger and I. Procaccia, Phys. Rev. Lett. 50 (1983) 346.
[39] P. Grassberger and I. Procaccia, Physica D 9 (1983) 189.
[40] F. Takens, Invariants related to dimension and entropy, in: Atas do 13° Colóqkio Brasiliero de Matemática (1983).
[41] S. Ellner, Phys. Lett. A 133 (1988) 128.
[42] J.D. Farmer and J.J. Sidorowich, in: Evolution, Learning and Cognition, ed. Y.C. Lee (World Scientific, Singapore, 1988), pp. 277–330.
[43] M. Casdagli, in: Modeling Complex Phenomena, eds. L. Lam and V. Naroditsky (Springer, New York, 1992), p. 131.
[44] M. Casdagli, Chaos and deterministic versus stochastic nonlinear modeling, J. R. Stat. Soc. B 54 (1992) 303.
[45] M. Sano and Y. Sawada, Phys. Rev. Lett. 55 (1985) 1082.
[46] J.-P. Eckmann and D. Ruelle, Rev. Mod. Phys. 57 (1985) 617.
[47] J.-P. Eckmann, S.O. Kamphorst, D. Ruelle and S. Ciliberto, Phys. Rev. A 34 (1986) 4971.
[48] S. Ellner, A.R. Gallant, D. McCaffrey and D. Nychka, Phys. Lett. A 153 (1991) 357.
[49] D. McCaffrey, S. Ellner, A.R. Gallant and D. Nychka, Estimating the Lyapunov exponent of a chaotic system with nonparametric regression, J. Am. Stat. Assoc., to appear.
[50] D. Nychka, S. Ellner, D. McCaffrey and A.R. Gallant, J. R. Stat. Soc. B 54 (1992) 399.
[51] T. Schreiber and P. Grassberger, Phys. Lett. A 160 (1991) 411.
[52] W.A. Brock, J. Lakonishok and B. LeBaron, Simple technical trading rules and the stochastic properties of stock returns, J. Finance, to appear.
[53] R.B. Blackman and J.W. Tukey, The Measurement of Power Spectra (Dover, New York, 1959).
[54] M. Hénon, Commun. Math. Phys. 50 (1976) 69.
[55] M.C. Mackey and L. Glass, Science 197 (1977) 287.
[56] J. Theiler, Phys. Rev. A 34 (1986) 2427.
[57] H. Haucke and R. Ecke, Physica D 25 (1987) 307.
[58] P.E. Rapp, I.D. Zimmerman, A.M. Albano, G.C. deGuzman, N.N. Greenbaum and T.R. Bashore, in: Nonlinear Oscillations in Biology and Chemistry, ed. H.G. Othmer (Springer, Berlin, 1986), pp. 175–205.
[59] P.E. Rapp, T.R. Bashore, J.M. Martinerie, A.M. Albano, I.D. Zimmerman and A.I. Mees, Brain Topography 2 (1989) 99.
[60] G.U. Yule, Philos. Trans. R. Soc. London A 226 (1927) 267.
[61] H. Tong and K.S. Lim, J. R. Stat. Soc. B 42 (1980) 245.
[62] N.O. Weiss, Phil. Trans. R. Soc. London A 330 (1990) 617.
[63] J. Kurths and A.A. Ruzmaikin, Solar Phys. 126 (1990) 407.
[64] A. Weigend, B. Huberman and D. Rummelhart, Intern. J. Neural Systems 1 (1990) 193.
[65] M. Mundt, W.B. Maguire II and R.B. P. Chase, J. Geophys. Res. 96 (1991) 1705.
[66] A. Weigend, B.A. Huberman and D.E. Rummelhart, Predicting sunspots and exchange rates with connectionist networks, in: Nonlinear Modeling and Forecasting, eds. M. Casdagli and S. Eubank, SFI Studies in the Sciences of Complexity, Vol. XII (Addison-Wesley, Reading, MA, 1992), pp. 397–434.
[67] L. Smith, Identification and prediction of deterministic dynamical systems, this volume.
[68] S. Ellner, Detecting low-dimensional chaos in population dynamics data: a critical review, in: Chaos and Insect Ecology, eds. J.A. Logan and F.P. Hain (University of Virginia Press, Blacksburg, VA, 1991), pp. 65–92.
[69] W.A. Brock and C.L. Sayers, J. Monetary Econ. 22 (1988) 71.
[70] W.A. Brock and W.D. Dechert, Statistical inference theory for measures of complexity in chaos theory and nonlinear science, in: Measures of Complexity and Chaos, eds. N. Abraham et al. (Plenum, New York, 1989), pp. 79–98.
[71] W.A. Brock and S.M. Potter, in: Nonlinear Modeling and Forecasting, eds. M. Casdagli and S. Eubank, SFI Studies in the Sciences of Complexity, Vol. XII (Addison-Wesley, Reading, MA, 1992), pp. 137–162.
[72] D.A. Hsieh, J. Business 62 (1989) 339.
[73] D.A. Hsieh, J. Finance 46 (1991) 1839.