Method to distinguish possible chaos from colored noise and to determine embedding parameters

Matthew B. Kennel*
Institute for Nonlinear Science, University of California, San Diego, Mail Code 0402, La Jolla, California 92093-0402 and Department of Physics, University of California, San Diego, Mail Code 0402, La Jolla, California 92093-0402

Steven Isabelle†
Digital Signal Processing Group, Department of Electrical Engineering, Massachusetts Institute of Technology, 36-615, 50 Vassar Street, Cambridge, Massachusetts 01239

(Received 23 September 1991; revised manuscript received 21 April 1992)

We present a computational method to determine if an observed time series possesses structure statistically distinguishable from high-dimensional linearly correlated noise, possibly with a nonwhite spectrum. This method should be useful in identifying deterministic chaos in natural signals with broadband power spectra, and is capable of distinguishing between chaos and a random process that has the same power spectrum. The method compares nonlinear predictability of the given data to an ensemble of random control data sets. A nonparametric statistic is explored that permits a hypothesis testing approach. The algorithm can detect underlying deterministic chaos in a time series contaminated by additive random noise with identical power spectrum at signal to noise ratios as low as 3 dB. With less noise, this method can also be used to get good estimates of the parameters (the embedding dimension and the time delay) needed to perform the standard phase-space reconstruction of a chaotic time series.

PACS number(s): 02.50.+s, 05.45.+b, 02.70.+d

I. INTRODUCTION

Deterministic chaotic processes can produce signals with broadband spectra similar to random processes. The chaotic system would have a finite-dimensional attractor, whereas a truly random process ought to be infinite dimensional. Given observed data, a finite number for the calculated fractal dimension of the supposed attractor has been assumed to imply that a deterministic process produced the time series in question. There are biases, however, in commonly used dimension algorithms, such as the Grassberger-Procaccia (GP) correlation dimension [1]. These algorithms seem to indicate a finite dimension, even when presented with random noise. As an illustrative, but not unique, example, Osborne and Provenzale [2] numerically demonstrate that a stochastic process with power spectrum $1/f^{\alpha}$ with $1 < \alpha < 3$ will result in a correlation dimension which converges to $D = 2/(\alpha - 1)$. Theiler [3] argues that this "anomalous fractal scaling" is observed in some phase-space regimes due to correlations between points that are close in time index. In a colored random process, the short-time correlations will produce an effect that will mimic a finite correlation dimension. If one does not count those points in the GP formula whose time separations are less than the characteristic autocorrelation time, this anomalous fractal scaling will be eliminated. There are known to be other biases in the correlation dimension algorithm, for example, the "edge effect" due to finite data samples in high dimensions that biases downward the observed value of dimension, a problem attacked by Dvorak and Klaschka [4]. This effect can make even white noise appear to have a finite correlation dimension, to say nothing of colored noise. In any case, there is the question of how to extract the dimension from a plot of "mass" versus radius, in other words, finding the proper "scaling region," which for the most part is done *ad hoc* (see Ref. [5]). The Lyapunov exponent for a noise signal ought to be infinite, in principle, but it is not clear how to use this fact to create a practical and robust algorithm for distinguishing noise from chaos. Finally, the number of points necessary to get good dimension estimates can grow exponentially with the fractal dimension, making difficult the distinction between "medium-dimensional" chaotic processes and noise. (See, however, [6].)

All the complications involved in getting a good dimension estimate motivate a search for a method to distinguish noise from chaos in an observed time series without explicitly using a dimension algorithm. Our method does this by using short-term *predictability* as a tool to distinguish deterministic chaos from randomness. This is aesthetically appealing as it emphasizes one of the fundamental distinctions between the two.

In brief, we compare the prediction error of the given time series to the prediction errors, computed the same way, of an ensemble of *random* time series which have the same average power spectral density, and thus the same autocorrelation, as the original data series. If the prediction errors of the real time series are sufficiently smaller than those of the random series we may reject the null hypothesis that the input time series is noise. This comparison with an ensemble of isospectral random

time series, as suggested by Theiler [3], will fail to find any difference when presented with linearly correlated noise, of any power spectrum.

In practice, this depends on the assumptions that one can perform the standard time-delay phase-space reconstruction, and from that construct a predictor that can approximate the dynamics of a supposedly low-dimensional signal. The philosophical question of whether some type of noise is truly "random" or merely high-dimensional chaos is irrelevant for our purposes. Our distinction between "noise" and "chaos" is determined by whether, in a practical setting, a prediction function designed for low-dimensional chaos offers a predictive advantage over a collection of "control" data sets. A deterministic, but sufficiently high-dimensional, data set may not be distinguishable from noise by our algorithm. This means that, for practical purposes, with the given amount of data and the tools used, one would might not find the methods of *deterministic* phase-space analysis significantly better than an analysis based on the hypothesis of a random process.

One can also use this statistic as a criterion to get good values of embedding parameters. One chooses the time delay and embedding dimension so as to maximally enhance predictive ability compared to random time series with the same spectrum. This selects a phase-space representation that maximizes nonlinear predictability, but only those representations that are highly predictable in an "interesting" way.

II. THE PREDICTION ALGORITHM

The nonlinear prediction relies on the now standard time-delay embedding for reconstructing the phase space [7]. Given an input scalar time series $x(n)$, with integer "time" $n = 1, \ldots, N$, one forms vectors in d dimensions:

$$\mathbf{y}(n) = (x(n), x(n+T), \ldots, x(n+(d-1)T)). \quad (1)$$

The integer quantities T and d being the time delay and embedding dimension, respectively. The prediction problem is posed: given an input vector $\mathbf{z} \in \mathbb{R}^d$ that represents the current position in phase space, predict the next scalar component τ time steps ahead of the last component of $\mathbf{z}$. There is no barrier to predicting all the components of the future image of $\mathbf{z}$ some τ time units later, but with a time-delay embedding as above, only the last component provides information about the future state of the system. Local information about the future evolution of the close neighbors of $\mathbf{z}$ is used to predict the future evolution of $\mathbf{z}$. Specifically, in our case, we find the single closest neighbor (both Euclidean and L^∞ metrics work fine) $\mathbf{y}(\eta)$ to the reference point $\mathbf{z}$. The predicted scalar, given a fixed prediction interval $\tau > 0$, is $x(\eta + (d-1)T + \tau)$, the future iterate of the nearest neighbor to $\mathbf{z}$. This locally constant predictor is the simplest form of the local predictors discussed in Farmer and Sidorowich [8], and is nearly too trivial to be called a "prediction algorithm." This choice is somewhat arbitrary and is motivated by a desire for simplicity, reproducibility, and speed. More sophisticated methods could very well provide superior results, though surprisingly, a brief investigation with local linear and quadratic predictors [8], as well as with radial basis function type predictors [9,10], did not show significant improvements for our purposes. Those and other methods require extra free parameters and entail more computation and programming, so we concentrated on the local constant predictor. In fact, using a phase-space predictor as a distinguishing statistic is not *a priori* necessary, but we commend its use as it directly investigates one of the principal differences between chaos and noise, namely determinism.

We use a k-d tree search algorithm [11], as simplified in Ref. [12], to rapidly accomplish the search for nearest neighbors. The k-d tree is an extension to k dimensions of the familiar binary tree data structure of computer science. As suggested by Theiler [3, 13], we ignore those neighbors whose time indices differ from that of the reference point by less than the autocorrelation time. This eliminates the spurious enhancement of predictive ability due to simple short-time correlations and has a similar effect to evaluating the prediction error on a different sample than that used to make the predictions, without having to split a given data set into two or more parts and suffer a loss in power due to fewer numbers of points.

We apply this prediction formula on every vector of the data set $\mathbf{y}(j)$ to obtain a set of prediction errors. At every point $\mathbf{y}(j)$ we compute the error $e(j)$ involved in predicting the next step forward, by first finding the nearest neighbor in the data set of $\mathbf{y}(j)$, denoted by $\mathbf{y}(\eta_j)$, and computing the difference between the neighbor's future iterate and the actual evolution of $\mathbf{y}(j)$:

$$e(j) = |x(\eta_j + (d-1)T + \tau) - x(j + (d-1)T + \tau)|. \quad (2)$$

With N scalar data points, one can compute predictions on $N - (d-1)T - \tau$ vectors.

III. SURROGATE DATA

We create an ensemble of random time series (typically between 10 and 100 in number) of exactly the same length and spectral magnitude, in the ensemble limit, as the original scalar data set. With the input data series denoted as $x(n)$ and its discrete Fourier transform as $X(k)$,

$$X(k) = \sum_{n=0}^{N-1} x(j) \exp 2\pi i n k/N, \quad (3)$$

we make M "faked" data sets, indexed by $r \in [1, M]$,

$$X^r(k) = X(k)[\xi(k) + i\eta(k)], \quad (4)$$

where ξ and η are independent, real Gaussian random numbers with mean 0 and variance 1/2. The randomizing numbers are not totally independent, though: we require $\eta(k) = -\eta(N-k)$ so that $X^r(k) = [X^r(N-k)]^*$, thus ensuring that the inverse transformation

$$x^r(n) = \frac{1}{N}\sum_{k=0}^{N-1} X^r(k)\exp[-2\pi ink/N] \quad (5)$$

is purely real. This is the one-dimensional version of the algorithm in Ref. [14], using simply the squared magnitude of the original signal's Fourier transform as the input power spectrum of the random process to simulate. Each random time series $x^r(n)$ is embedded in the same fashion as before, and from each of these, the predictor errors $e^r(j)$ are computed similarly.

There is a technicality. Due to central-limit behavior, the probability distribution of the points $x^r(n)$ that result from the inverse transform [Eq. (5)] tend to a Gaussian distribution. The method as it stands can distinguish chaos from Gaussian colored noise, which is frequently observed, but if the input signal $x(n)$ is noise, but with a strongly non-Gaussian distribution, the resulting "faked" $x^r(n)$ data sets can sometimes be statistically distinguishable from the original, therefore giving a false positive result.

To alleviate this problem, we perform a "histogram transformation" of the original scalar data set to obtain an approximately Gaussian probability distribution function, before the embedding and calculation of the prediction errors $e(n)$. (Before computing each fake data set, we redo this transformation.)

To accomplish this, we create a sample of Gaussian random numbers the same size as the input data series, and then sort both data sets in numerical order, making a one-to-one correspondence between points with the same index. This defines an invertible nonlinear scalar transformation of the original time series to a new one, with approximately Gaussian probability density. This transformation locally preserves phase-space neighborhoods; therefore, if the original series came from a dynamical system, the dynamics will be preserved by this transformation. Specifically, the Lyapunov exponents can be shown to be invariant [15] under this transformation, and thus by the Kaplan-Yorke formula [16], so ought to be the fractal dimension. If, on the other hand, the original time series is a random process, then so will be the transformed one, but with Gaussian probability distribution.

As a check, we perform a two-sample Kolmogorov-Smirnov test on the transformed input set and each fake data set. If this statistic distinguishes the two (one-dimensional) probability distribution functions with significance greater than 95%, then we discard this fake data set and create a new one. If this occurs very frequently, it usually means that the input data set has a few Fourier components much larger than the others, either because of a very strong periodicity or because large power at low frequencies. This means that our method of noise simulation is not appropriate for the input data, and a new one ought to be developed.

IV. THE STATISTIC

Our prediction formula is specifically designed for deterministic signals, and so should give comparatively better predictions on deterministic data sets than on random ones. The crux of the algorithm comes down to comparing the sizes the prediction errors on the true data set $e(j)$ to the sizes of the errors on the faked data sets $e^r(j)$, $r = 1,\ldots,M$. One could, for example, examine the ratio of the mean error of the true data set to the mean error of all the fake data sets. When this ratio is sufficiently smaller than 1, this implies that the true data are more predictable than the fake data. But the question remains, what is "sufficiently smaller"?

Thus we reformulate the question as follows: Do the two observed sample distributions of prediction errors (from the true data set and from the faked data sets) come from the same underlying distribution (the null hypothesis that the input signal is noise), or are the true prediction errors statistically smaller than the fake prediction errors?

Unfortunately, all conventional statistical tests require that the samples be independent, and this is not strictly true in our circumstance, as, in principle, the prediction at any one point is affected by all others in the same data set. The only independent measurements come from independent data sets and thus, to be entirely rigorous, one could use only a single measurement from the true data and one from each fake data set. A very large number of fake data sets would need to be synthesized and processed to achieve adequate statistical power, entailing a huge computational cost. In practice, however, there does not appear to be much correlation between predictor errors whose reference points are sufficiently separated in time. As an approximation, then we decimate the set of predictor errors $e(j)$ by a factor corresponding to the autocorrelation time of the predictor errors (i.e., where it drops effectively to zero), and then assuming that the remaining samples are independent. For nearly all cases we studied, the autocorrelation time of the predictor errors is much smaller than the autocorrelation time of the data set itself. It seems that a rigorous justification of this procedure may be difficult, but it appears to be physically reasonable and its success is borne out by experiment. If this approximation is worrisome and computation cheap, then one may always increase the time interval for the decimation and then increase the number of surrogate data sets to make up for the loss of power due to smaller numbers.

With the decimated set of prediction errors on the real data set denoted A, and the union of all decimated prediction errors on fake data sets denoted B we form the Mann-Whitney rank-sum statistic

$$U = \sum_{i=1}^{N_2}\sum_{j=1}^{N_3}\Theta(A_i - B_j), \quad (6)$$

with N_2 number of elements in A, $N_3 = MN_2$ the number of elements in B, and Θ the Heaviside step function: $\Theta(x) = 1$ for $x > 0$ and $\Theta(x) = 0$ for $x \le 0$. This statistic may be evaluated efficiently [$O(N \ln N)$ time] by use of a quick sort algorithm, and only depends on the relative ranks of the samples of each set. For large N_2 (which in practical terms means a few tens), the quantity

$$z = \frac{U - N_2 N_3/2}{\sqrt{\frac{1}{12} N_2 N_3 (N_2 + N_3 + 1)}} \tag{7}$$

is normally distributed with zero mean and unit variance, under the *null hypothesis* that the two observed samples came from the same distribution. If on the other hand, the samples in set A were smaller than those of B, in a statistically significant way, we would observe small values of z. Thus, for example, if we observe a single z value less than -2.33, we can state that we disprove the null hypothesis at the 99% confidence level. In other words, the probability that the two sets A and B were random samples of the *same* (unknown) distribution is less than $\alpha = 0.01$. This statistic serves the equivalent purpose as Student's classical t-test, but does not assume an underlying normal distributions for each set. This is appropriate for our situation, as justified by numerical experiment. (We have successfully used other distinguishing statistics, such as the Kolmogorov-Smirnov test. That test does not indicate which of the two distributions has the larger or smaller elements, though.)

If one were to repeatedly compute the z values, corresponding to a range of time delays and embedding dimension, the probability that *any* of the z values exceed that threshold under the null hypothesis is clearly larger than 1%. Therefore, to maintain a set confidence level, the limits on z get more stringent. With K repetitions of the statistic the corresponding one-variable confidence limit $\alpha = 0.01$ would need to be lowered. We conservatively approximate the new the rejection threshold z_0,

$$\int_{-\infty}^{z_0} \frac{1}{\sqrt{2\pi}} e^{-z^2} dz = \alpha/K. \tag{8}$$

Bearing in mind the approximations made in the method, one ought to refrain from designating a given time series as having structure different from noise unless results indicated significantly enhanced predictability over a range of time delays and embedding dimensions, as well as increasing M, the number of "control" data sets, even though technically a single statistic below the threshold derived above indicates rejection.

V. NUMERICAL RESULTS

We apply the method to "data" generated from a deterministic three-dimensional model of Lorenz [17]:

$$\begin{aligned} \dot{x} &= -y^2 - z^2 - a(x - F), \\ \dot{y} &= xy - bxz - y + G, \\ \dot{z} &= bxy + xz - z. \end{aligned} \tag{9}$$

We use the values $a = 0.25, b = 4.0, F = 8.0$, and $G = 1.0$ where Lorenz points out that irregular behavior is encountered. The attractor has a numerically computed fractal dimension d_A slightly greater than 2.5. This is a more complicated attractor than the classical Lorenz model [24]. We produced the data with a variable order Adams integrator with a sampling time $\delta t = 0.05$, with $N = 16384$ samples of the x variable used as the univariate time series. Figure 1 shows the z statistic [Eq. (7)]

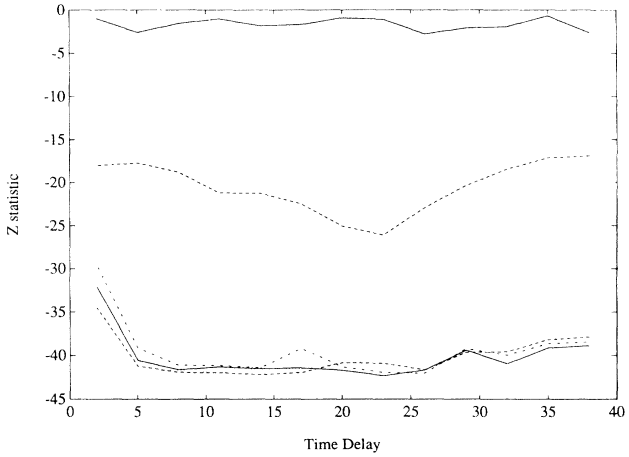

FIG. 1. Statistic for Lorenz [17] data set, no added noise. X axis is in multiples of the sampling time $T = 0.05$. Each curve is for a different embedding dimension: $d = 1$ is the top curve, with d increasing for successively lower curves.

as a function of time delay (in multiples of the sampling time) and embedding dimension. We choose prediction interval τ to be 10 in these time units, about half the mutual information time. This is the time for which the average mutual information (see Fraser and Swinney [18]) reaches its first minimum, which is often used as a good guess for a characteristic time scale, say as the time delay for embedding. The method is not sensitive to the exact choice of τ, within reason. The large negative z values indicate that this data set is unquestionably deterministic. The statistic converges to its lowest value for $d = 5$, which is consistent with other estimates of embedding dimension [19], although using $d = 4$ would probably cause little harm. There is a broad minimum of acceptable time delays, perhaps from 10 to 30, that result in the same level of predictability.

Figure 2 displays the same statistic when 10% (by amplitude) noise is added to the input data set. To demonstrate the algorithm's power, the noise signal has the

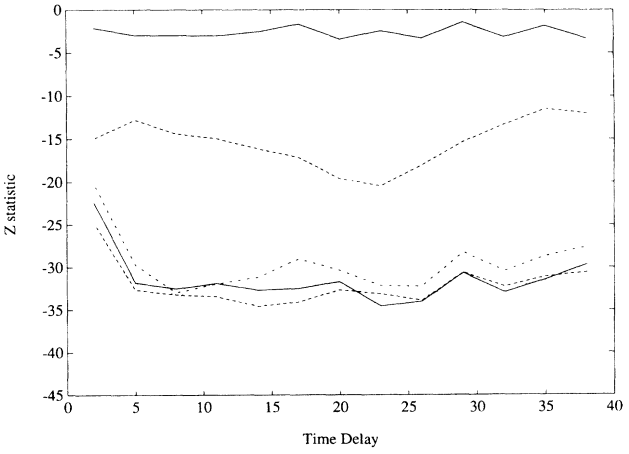

FIG. 2. Statistic for Lorenz [17] data set, 10% added noise. Noise has the same power spectrum as the original signal.

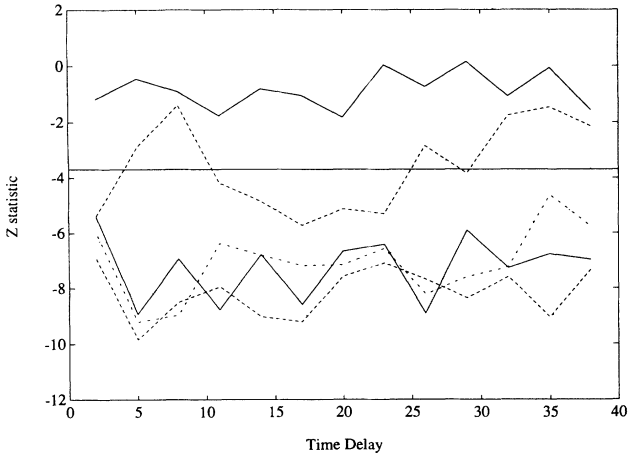

FIG. 3. Statistic for Lorenz [17] data set, 50% added noise. Horizontal line is at 99% confidence level for rejection.

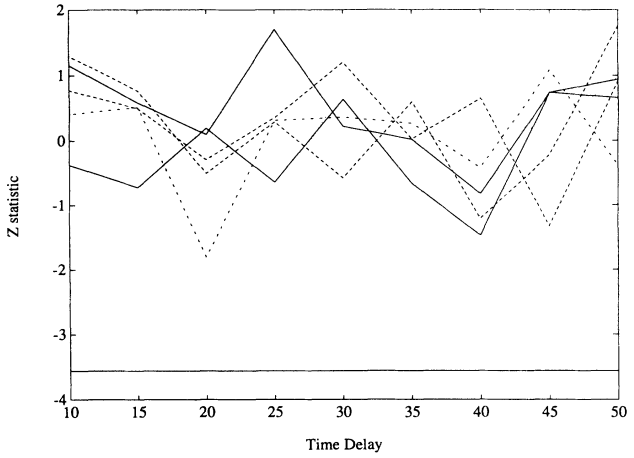

FIG. 5. Statistic for colored noise data set. Horizontal line is at 99% confidence level for rejection.

same power spectrum as the original (generated by the method previously described), thus this noise cannot be distinguished or removed from the original signal using spectral approaches used in traditional linear signal processing. The bottom plateau is raised above its previous level, but almost all qualitative features are maintained. Figures 3 and 4 show the cases of 50% and 75% noise, respectively, where the horizontal line in these figures displays the 99% rejection threshold. For the first case, the rejection of the null hypothesis is still significant for most $d > 1$, over a wide range of time delays. The equal power noise case shows the limit of the algorithm. There are enough cases that exceed the rejection threshold to indicate a deterministic element in the input series, although the time delay and embedding dimension information is lost. The second case, 75% noise, shows approximately highest noise power at which one can still reject the null in this system over a range of time delays and embedding dimensions. To improve the performance for this run, we used 30 fake data sets, three times the usual number, for each point. With much more noise (negative signal-to-noise ratios) we cannot reject the null hypothesis as the random noise swamps out the signal.

We have observed similar successes using other attractors used in the dynamics literature, such as the Hénon [20] and Ikeda [21] maps, as well as the Rössler [22], Mackey-Glass [23], and Lorenz [24] attractors. In general, the algorithm can detect determinism up to noise levels of about 50–100 %, depending on the system, the type of noise, and how much computation (in terms of the number of fake data sets) one is willing to expend.

In comparison, we calculate this same statistic on data from a first-order autoregressive random process:

$$x(n+1) = 0.9x(n) + \xi(n), \tag{10}$$

where the $\xi(n)$ are samples of white Gaussian random noise, zero mean and unit variance. We use $N = 16\,384$ points, $M = 10$ fake data sets, a prediction time $\tau = 20$, and an autocorrelation cutoff of 200. These last two quantities are estimated from the average mutual information and autocorrelation plot in the same way that one would do with a chaotic data set. Results are not sensitive to the exact values, however. Figure 5 shows

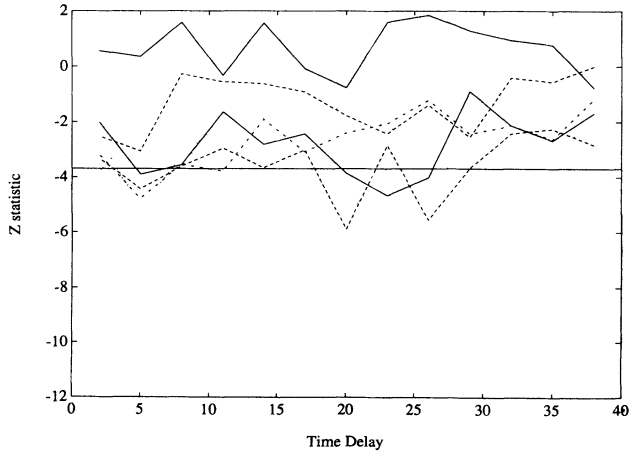

FIG. 4. Statistic for Lorenz [17] data set, 75% added noise. Horizontal line is at 99% confidence level for rejection.

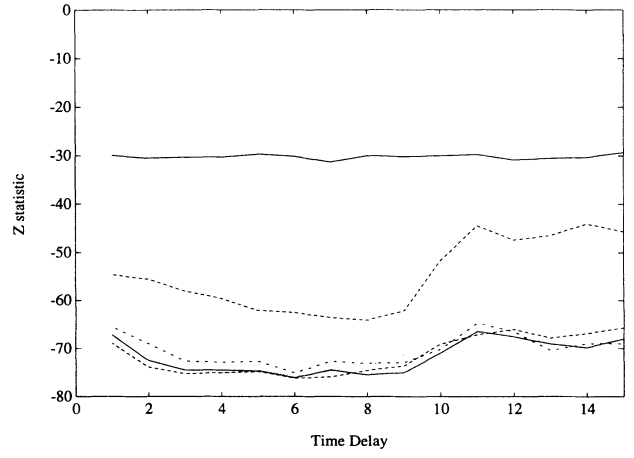

FIG. 6. Statistic for hysteretic nonlinear circuit data.

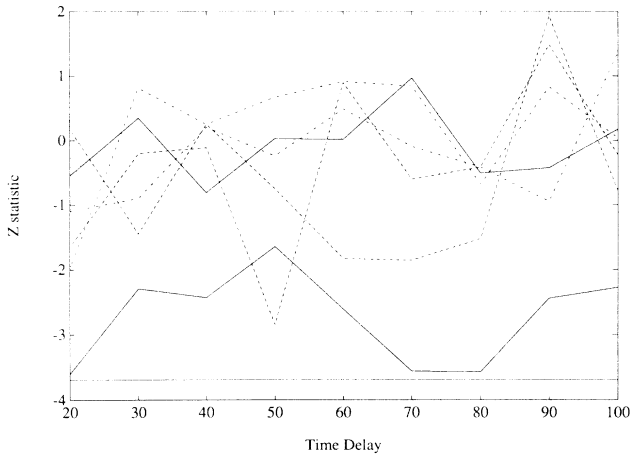

FIG. 7. Statistic for wind velocity data. Straight horizontal line is 99% rejection threshold.

the statistic as a function of time delay and embedding dimension. Clearly, we do not reject the null hypothesis that the data set is noise. We have obtained similar results (acceptance of the null) with a number of other random data sets, such as white uniform and Gaussian numbers, $1/f^{\alpha}$ noise as considered by Osborne and Provenzale [2] and Theiler [3], and colored random processes with highly non-Gaussian probability distribution functions, such as log-normal. For the sake of brevity we omit the graphs, all of which qualitatively resemble Fig. 5.

Figure 6 shows the result of applying this technique to 32 768 experimental data points from a nonlinear electronic circuit. The test emphatically distinguishes the time series from noise, indicating a clean, low-dimensional ($d_E \approx 4$ or 5) chaotic system. Finally, Fig. 7 shows the application to a time-series measurement of wind velocity at a fixed position [25]. Here there is no rejection of the null hypothesis. Interestingly, though, the statistic for embedding dimension 1 is lower than for all higher embedding dimensions. Perhaps this indicates possible nonlinear structure of the noise, which could potentially also be distinguished from linearly correlated noise in our test, though not clearly significant determinism. These results are meant only as illustrations of the technique, and are not the last word on these actual physical systems.

VI. DISCUSSION AND CONCLUSION

In this work, we have demonstrated an *objective* test to determine whether a given time series has deterministic structure or is purely random, though possibly with linear correlation. We term a time series to be deterministic if we can practically observe more phase-space predictability, in a statistically significant sense, than in an ensemble of random time series with the same power spectrum. This will point out where traditional linear, spectral methods of analysis are not optimal, and where a phase-space approach may be superior. The statistic automatically adapts to changing sample size, requiring a larger predictive advantage for a smaller data set in order to reject at a given significance level.

One may use the statistic as a criterion for choosing embedding dimension and time delay (or possibly any other relevant parameters) for phase-space reconstruction: choose the reconstruction that maximizes the predictability, as defined by our algorithm. For the purposes of embedding dimension, the essential mechanism is the same as explored other recent work [19, 26]. If one were to choose an embedding dimension that is too small, there will be some regions of phase space where truly separated points of the attractor, in terms of the original dynamical system, are accidentally made close by the embedding process. This results in what Kennel, Brown, and Abarbanel [19] term "false nearest neighbors." The predictions made using these false neighbors will be quite erroneous, and will contribute to increasing the prediction error, and hence the z statistic. Once the proper dimension has been reached, these errors will not occur, and the statistic will stop dropping dramatically with increasing embedding dimension.

For the purposes of determining time delay, the mechanism is not as clear. According to the mathematical theorems [27, 28] that underlie the time-delay embedding process, *any* time delay is valid, though of course rigorously true only for the unphysical case of infinite amounts of noise-free data. In a practical sense, though, a range of time delays ought to work well, and this is often reflected in the broad minimum of our statistic as a function of time delay. Using this method to obtain time delays is simply a heuristic to satisfy some arbitrary, but potentially useful, condition (maximize predictive ability compared to isospectral random processes) similar to other proposed means, such as finding the first minimum of the average mutual information [18], or the first zero crossing of the autocorrelation function. In a general sense, one can imagine how for very large time delays, a chaotic signal can look more like noise, resulting in a higher statistic, and how for very short time delays, the signal can be mimicked by noise with finite autocorrelation, again resulting in higher values of the statistic. It has perhaps the advantage of treating time delay and embedding dimension, and possibly any other parameters of a more sophisticated reconstruction, simultaneously. Liebert, Pawelzik, and Schuster's method [26] also makes claim to the same advantage, though their method does not specifically address the question of noise.

Casdagli [9] has advocated using nonlinear prediction (he specifically uses radial basis functions) to detect chaos. He proposes examining the prediction error as a function of embedding dimension. A clean chaotic signal ought to show a large drop in prediction error once the correct embedding dimension has been reached, whereas a noise signal will not, in general. We have observed, however, that prediction error can decrease with embedding dimension even on random time series, and the error for a noise-contaminated deterministic time series will not drop quite as radically as for a clean one, and hence the distinction between the two may not always be obvious. More recently [10], he proposes using a predictor function possessing a parameter that can adjust the fit-

ting model from a highly local predictor to a global one. A noise signal should do as well with a global model as with a local one, but a deterministic chaotic signal should be more predictable with a local model.

Sugihara and May [29] have also used prediction as a means of distinguishing chaos from noise. Essentially, they examine prediction error as a function of time interval ahead that one predicts, iterating a one-step predictor. A deterministic, chaotic time series will exhibit increasing prediction error with increasing time step, whereas for a noise signal (they only consider white noise) the prediction error will be approximately constant. Distinguishing the two is done only heuristically. It seems that this test will fail on correlated noise, as even in this case, prediction error can be better for short times than for long.

Brock [30] derives a statistical test on values of the correlation integral to accept or reject the hypothesis of white noise. Explicitly comparing an observed time series with noise that has similar properties, such as probability distribution and autocorrelation, is starting to become a useful diagnostic tool for the analysis of observed data. Early on, Grassberger [31] passed noise through a first-order linear filter and showed it produced similar correlation dimension statistics as a geophysical time series previously claimed to be low dimensional.

Scheinkman and LeBaron [32] compare an observed time series with temporally scrambled versions of the same series, thus testing the null hypothesis of white noise, but with the same probability distribution as the original signal. More recently, Kaplan and Glass [33] use the averaged directional vectors in a partitions of phase space as a distinguishing statistic from autocorrelated random data sets and an analytic criterion based on a random-walk hypothesis, though they do not test their method's sensitivity to noise. The methods outlined by Theiler *et al.* [34] are the most similar to those presented here, the major difference being in our use of predictor errors from many points in phase space instead of the average prediction error as a distinguishing statistic. Because the resulting increase in statistical power, we should be able to achieve good results with substantially less computer time, though theoretically there is a potential for a false positive result. In addition, the surrogate data sets used by Theiler *et al.* do not simultaneously possess the same probability distribution function and power spectrum as the observed set used to calculate the statistic, though their transformed versions do. Their use of local linear or polynomial predictors may be overkill as a simpler local constant predictors appears to give similar performance. Theiler *et al.* also consider correlation dimension as a statistic, but it would seem that the extraction of a consistently good "scaling region" might be too problematic to be easily automated.

Potential future enhancements to this method lie in the area of developing new methods of generating surrogate data, such as for multidimensional data sets (generate random streams with identical autocorrelation and cross correlation, as in Ref. [14]), or for new null hypotheses such as noisy periodicity (extract a periodic wave form and add noise).

In sum, we have presented a practical method to distinguish an observed data set from a linear stochastic process with broadband power spectrum. The calculations can be performed rapidly on common workstations and show robustness in detecting determinism in the presence high levels of additive observational noise. The statistic is also useful in optimizing reconstruction parameters, such as, but not limited to, time delay and embedding dimension in such a way that maximizes "nontrivial" nonlinear predictability.

ACKNOWLEDGMENTS

We thank the members of INLS, the MIT Digital Signal Processing group, and Cory Myers of Lockheed/Sanders for discussions on this subject. The comments and assistance of Henry Abarbanel, Reggie Brown, and of J. J. Sidorowich were especially helpful. This work was supported in part under the DARPA—University Research Initiative, URI Contract No. N00014-86-K-0758, and in part by Code 121 of the Office of Naval Research through a contract to Lockheed/Sanders, MIT, and UCSD.

* Electronic address (internet): mbk@inls1.ucsd.edu
† Electronic address (internet): stevi@allegro.mit.edu

[1] P. Grassberger and I. Procaccia, Physica D **35**, 189 (1983).
[2] A. R. Osborne and A. Provenzale, Physica D **35**, 357 (1989).
[3] J. Theiler, Phys. Lett. A **155**, 480 (1991).
[4] I. Dvorak and J. Klaschka, Phys. Lett. A **145**, 225 (1990).
[5] J. Theiler, Phys. Rev. A **41**, 3038 (1990).
[6] N. Gershenfeld, Physica D **55**, 135 (1992).
[7] J. P. Eckmann and D. Ruelle, Rev. Mod. Phys. **57**, 617 (1985).
[8] J. D. Farmer and J. J. Sidorowich, Phys. Rev. Lett. **59**, 845 (1987).
[9] M. C. Casdagli, Physica D **35**, 335 (1989).
[10] M. C. Casdagli, J. R. Stat. Soc. B **54**, 303 (1991).
[11] J. H. Friedman, J. L. Bentley, and R. A. Finkel, ACM Trans. Math. Software **3**, 209 (1977).
[12] R. F. Sproull, Algorithmica **6**, 579 (1991).
[13] J. Theiler, Phys. Rev. A **34**, 2427 (1986).
[14] L. E. Wittig and A. K. Sinha, J. Acoust. Soc. Am. **58**, 630 (1975).
[15] H. D. I. Abarbanel, R. Brown, and M. B. Kennel, Int. J. Mod. Phys. B **5**, 1347 (1991).
[16] J. L. Kaplan and J. A. Yorke, in *Functional Differential Equations and Approximation of Fixed Points*, edited by H. O. Peitgen and H. O. Walther, Lecture Notes in Mathematics Vol. 730 (Springer, Berlin, 1979), p. 204.
[17] E. N. Lorenz, Tellus **36A**, 98 (1984).
[18] A. M. Fraser and H. L. Swinney, Phys. Rev. A **33**, 1134

(1986); A. M. Fraser, IEEE Trans. Inf. Theory **35**, 245 (1989); A. M. Fraser, Physica D **34**, 391 (1989).
[19] M. B. Kennel, R. Brown, and H. D. I. Abarbanel, Phys. Rev. A **45**, 3403 (1992).
[20] M. Hénon, Commun. Math. Phys. **50**, 69 (1976).
[21] K. Ikeda, Opt. Commun. **30**, 257 (1979).
[22] O. E. Rössler, Phys. Lett. A, **57**, 397 (1976).
[23] M. C. Mackey and L. Glass, Science **197**, 287 (1977); D. Farmer, Physica D **4**, 366 (1982).
[24] E. N. Lorenz, J. Atmos. Sci. **20**, 130 (1963).
[25] We are grateful to T. Carroll and L. Pecora from the U.S. Naval Research Laboratory for providing the circuit data, and to Cory Myers of Lockheed/Sanders for the wind data.
[26] W. Liebert, K. Pawelzik, and H. G. Schuster, Europhys. Lett. **14**, 521 (1991).
[27] R. Mañe, in *Dynamical Systems and Turbulence,* edited by D. Rand and L. S. Young, Lecture Notes in Mathematics Vol. 898 (Springer, Berlin, 1981), p. 230.
[28] F. Takens, in *Dynamical Systems and Turbulence* (Ref. [27]), 366.
[29] G. Sugihara and R. M. May, Nature **344**, 734 (1990).
[30] W. A. Brock, J. Econ. Theory **40**, 168 (1986).
[31] P. Grassberger, Nature **323**, 609 (1986).
[32] J. Sheinkman and B. LeBaron, J. Bus. **62**, 311 (1989).
[33] D. T. Kaplan and L. Glass, Phys. Rev. Lett. **68**, 427 (1992).
[34] J. Theiler, B. Galdrikian, A. Longtin, S. Eubank, and J. D. Farmer, Los Alamos Report No. LA-UR-91-2615, 1991 (unpublished).