Recognizing Determinism in a Time Series

Richard Wayland, David Bromley, Douglas Pickett, and Anthony Passamante
Naval Air Warfare Center, Aircraft Division, Warminster, Pennsylvania 18974-5000
(Received 26 June 1992; revised manuscript received 19 October 1992)

A quantitative measure of determinism in a recurrent time series is developed. Specifically, scalar time series data are used to form a vector series in reconstructed phase space. A statistic is then developed to measure the observed "continuity" of the vector series. The statistic is used as a measure of determinism. Several examples are included to demonstrate the effectiveness of the method.

PACS numbers: 05.45.+b, 05.40.+j

One is often interested in explaining the variability observed in complex time series. Is the variability due to external stochastic forces, internal deterministic dynamics, or to a combination of the two? Answering this question can profoundly affect subsequent model development. Recently, several authors (Sugihara and May [1], and Kaplan and Glass [2]) have developed techniques for measuring the extent to which deterministic dynamics can explain the variability. The Sugihara-May method is based on how well the past can predict the future. The Kaplan-Glass method is based on the parallelness of a certain vector field formed from the data. Here we propose an effective and computationally simple variant on the Kaplan-Glass method, which may be implemented using relatively small sets of data, and which performs well in high levels of uncorrelated noise. Our method uses the "phase space continuity" observed in the time series to measure determinism. Several examples are presented to illustrate the method.

For our purposes, we use the following operational definition of a deterministic time series: A time series $s(1),s(2),\ldots,s(N)$ is *deterministic*, if the sequence of vectors

$$x(j)=(s(j),s(j+L),\ldots,s(j+(E-1)L)), \quad j=1,\ldots,N-(E-1)L \tag{1}$$

can be "accurately" modeled as the iteration of some continuous function f, for an appropriate choice of the embedding dimension E and the time lag L. We refer to this collection of vectors as the "experimental attractor." Since we are working with a finite sequence of vectors, we cannot implement a "standard" calculus test for continuity. Consequently, we develop an empirical test for continuity, applicable to our situation, which is based on the fact that for a continuous map f, points close together will map to points close together.

Let x_0 be a fixed, but otherwise arbitrary, vector on the experimental attractor, and let $x_1,\ldots,x_k$ denote the k nearest neighbors of x_0. Let $y_0,\ldots,y_k$ represent the images of these points. (For continuous-time systems, the sampling rate and lag must be chosen to insure that the points $x_0,\ldots,x_k$ are not temporally correlated.) If the data are deterministic, we expect the translation vectors

$$v_j=y_j-x_j \tag{2}$$

to be nearly equal, provided that x_0 and its k nearest neighbors are contained in a sufficiently small region of E-dimensional space.

To quantify this notion, let

$$\langle v\rangle=\frac{1}{k+1}\sum_{j=0}^{k}v_j \tag{3}$$

denote the average of the translation vectors v_j and compute the *translation error*

$$e_{\text{trans}}=\frac{1}{k+1}\sum_{j=0}^{k}\frac{||v_j-\langle v\rangle||^2}{||\langle v\rangle||^2}, \tag{4}$$

where $||v||$ denotes the Euclidean length of the vector v. The translation error measures the fractional spread in the displacements experienced by $x_0,\ldots,x_k$ relative to the average displacement $\langle v\rangle$. Note that division by the length of $\langle v\rangle$ makes the translation error insensitive to an overall scaling of the original time series. If the time

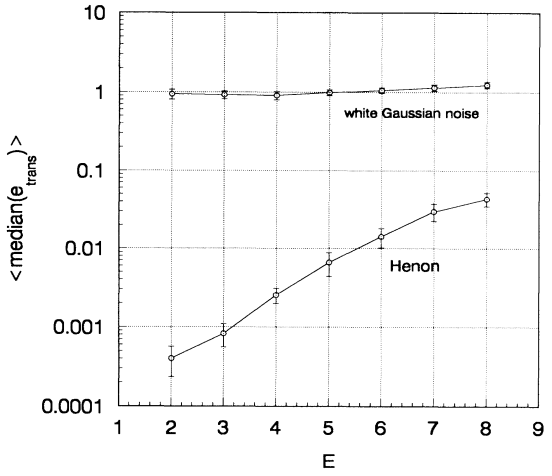

FIG. 1. Plot of the median translation error vs embedding dimension for the Hénon map and for white Gaussian noise. $L=2$, $k=4$, and $N_{\text{rcs}}=100$. In each case, thirty 1024-point realizations of the signals were used and an average of the results was computed. The error bars represent the mean ± 1 standard deviation.

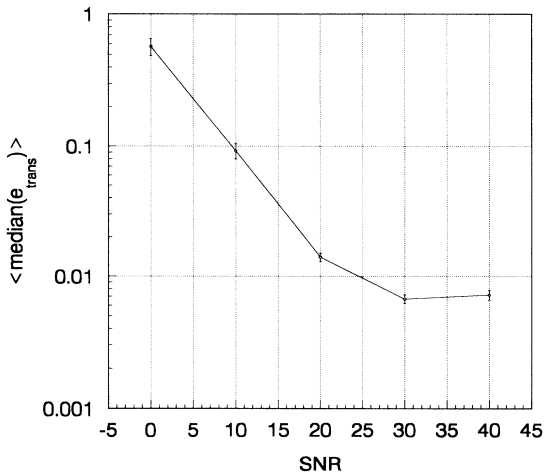

FIG. 2. Plot of the median translation error vs signal-to-noise ratio for Hénon data contaminated by adding white Gaussian noise. $E=5$, $L=2$, $k=4$, and $N_{rcs}=100$. At each signal-to-noise ratio, thirty 1024-point realizations of Hénon and noise were used to calculate the mean and standard deviation of the median translation error.

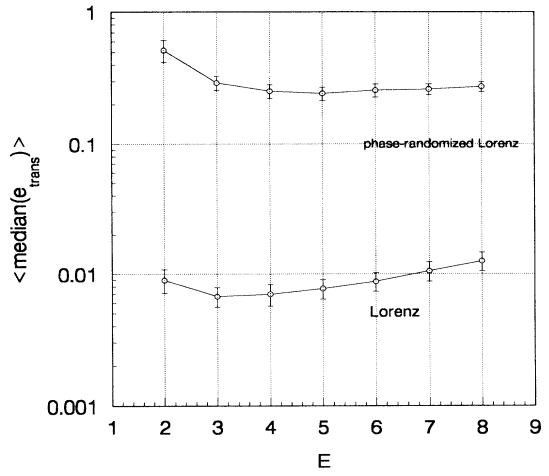

FIG. 3. Plot of the median translation error vs embedding dimension for the Lorenz data and a Gaussian stochastic process having the same average power spectrum as the Lorenz data. $L=4$, $k=4$, and $N_{rcs}=200$. In each case, thirty 2048-point realizations of the signals were used and an average of the results was computed. The error bars represent the mean ± 1 standard deviation.

series is deterministic, then the v_j will be nearly equal, and the translation error will be small.

We extend this local calculation of continuity to a global measure as follows. Choose N_{rcs} random centers from the experimental attractor—this is equivalent to sampling the natural density of the experimental attractor. For each of the centers compute the translation error. The median of these values provides a robust measure of the determinism in the attractor. We now discuss applications of this technique.

We begin with a simple example. In Fig. 1, we compare the translation error for Hénon data [3] to that of a white Gaussian process. Note that spectrally the Hénon map is a white process. For this example, we have chosen $k=4$, $L=2$, $N=1024$, and $N_{rcs}=100$. Thirty realizations of each signal were used to compute an average median translation error. The error bars represent the mean ± 1 standard deviation. Note that the Hénon error exhibits an exponential increase with embedding dimension, while the error for white Gaussian noise is roughly constant.

In Fig. 2, we illustrate the effect of additive noise on the translation error. One 1024-point sample of a Hénon time series was mixed with various levels of white Gaussian noise. We measure the level of noise with the *signal-to-noise ratio* R_{SNR}, where $R_{SNR}=10\log$ (signal variance/noise variance). Thirty realizations of Hénon and noise were used.

As a second example, we consider the Lorenz equations [4]. (Parameters are $r=28$, $b=8/3$, and $\sigma=10$.) A Runge-Kutta method of order 4 with a step size of 0.00314 was used to obtain an approximate solution of the system. The x component was then decimated by 10, resulting in approximately 20 samples per typical oscillation about one of the unstable spiral points. Here, we compare the translation error for Lorenz $x(t)$ data to that of a *surrogate* stochastic process [5]. The surrogate data were created by averaging the magnitude spectrum of 2048-point samples of Lorenz $x(t)$ data. A random phase was assigned to each value of the average magnitude spectrum, and then an inverse discrete Fourier transform was used to create surrogate time series. In this way, we obtain samples of a Gaussian process having the same average power spectrum as the Lorenz data. Figure 3 compares the behavior of the median translation errors for the Lorenz and Lorenz surrogate time series. For this

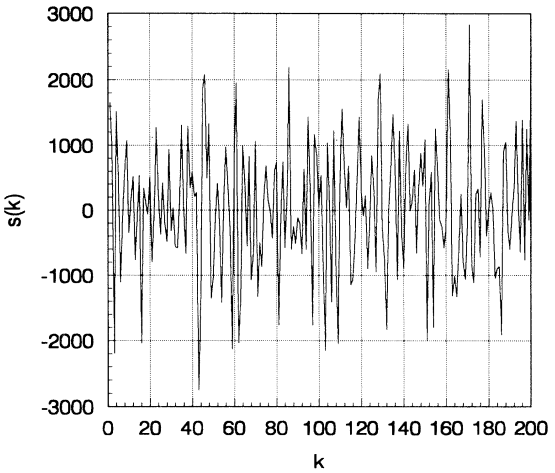

FIG. 4. Sample time series of sea noise data.

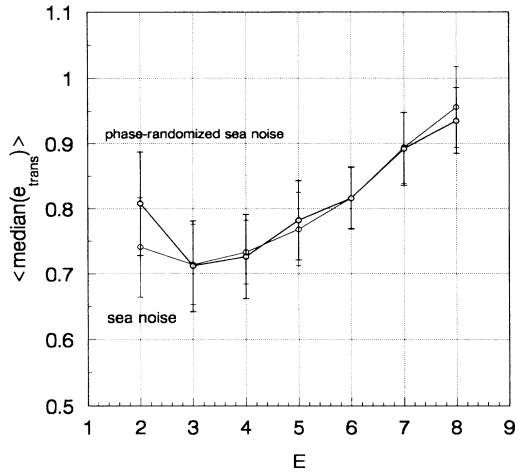

FIG. 5. Plot of median translation error vs embedding dimension for sea noise and a Gaussian stochastic process having the same average power spectrum as the sea noise. $L=3$, $k=4$, and $N_{rcs}=200$. In each case, thirty 2048-point realizations of the signals were used and an average of the results was computed. The error bars represent the mean ± 1 standard deviation.

example, we have chosen $k=4$, $L=4$, $N=2048$, and $N_{rcs}=200$. Thirty realizations of each signal were used to compute an average median translation error. Note that based on the behavior of the translation error, the two processes are clearly distinguishable.

As our final example, we apply our technique to a sample of ambient sea noise. The data were taken in a frequency band from 550 to 750 Hz where one expects the dominant noise source to be surface winds. The original full-band signal was translated down in frequency, low pass filtered, and then decimated to achieve a final sampling rate of 820 Hz. Figure 4 shows a time series from the resulting data. As in our second example, we first construct a surrogate data set having the same average power spectrum as the sea noise. In Fig. 5, we compare the behavior of the median translation errors for the sea noise and the surrogate data. For this example, we have chosen $k=4$, $L=3$, $N=2048$, and $N_{rcs}=200$. Thirty realizations of each signal were used to compute an average median translation error. We remark that a similar calculation with $k=4$, $L=3$, $N=6144$, and $N_{rcs}=600$ produced nearly identical results. Thus, according to our measure of determinism, the sea noise behaves like a Gaussian stochastic process.

In conclusion, we have developed a quantitative measure of the amount of determinism in a time series. The method is intuitively simple, computationally efficient, and gracefully degrades in noise. We have presented several examples of how the technique can be applied. We end by noting that while our method does not provide an absolute test for determinism in a time series (if such a test existed), it does provide a quantitative measure of the appropriateness of deterministic models for a complex time series.

[1] G. Sugihara and R. M. May, Nature (London) **344**, 734 (1990).
[2] D. Kaplan and L. Glass, Phys. Rev. Lett. **68**, 427 (1992).
[3] M. Hénon, Commun. Math. Phys. **50**, 69 (1976).
[4] E. N. Lorenz, J. Atmos. Sci. **20**, 130 (1963).
[5] J. Theiler, J. Galdrikian, A. Longtin, S. Eubank, and J. D. Farmer, in *Nonlinear Modeling and Forecasting*, edited by M. Casdagli and S. Eubank (Addison-Wesley, Reading, MA, 1992).