# Determinism test for very short time series

P.-M. Binder,* Ryu Igarashi, William Seymour, and Candy Takeishi
*Natural Sciences Division, University of Hawaii, Hilo, Hawaii 96720-4091, USA*
(Received 11 June 2004; published 23 March 2005)

A test for determinism suitable for time series shorter than 100 points is presented, and applied to numerical and observed data. The method exploits the linear $d(d_0)$ dependence in the expression $d(t)\sim d_0 e^{\lambda t}$ which describes the growth of small separations between trajectories in chaotic systems.

DOI: 10.1103/PhysRevE.71.036219 PACS number(s): 05.45.Pq

Nonlinear time series analysis is one of the most productive areas in the study of chaotic systems [1–3]. When poorly applied, however, it can result in the spurious identification of chaotic behavior in experimental or observed time series; specific cases are discussed in [4–6]. Hence, it is advisable to test series for the two main characteristics of chaos, nonlinearity, and determinism, before engaging in a full analysis. The surrogate data technique has proved to be a good test for nonlinearity [7]. Determinism tests have also existed for some time [8–11].

To analyze a time series one requires a number of points, typically a thousand or more, that increases with the system's fractal dimensions [3]. Unfortunately, not all observed series are this long. In this paper we propose a method to detect the presence of determinism which has proved to work well with series shorter than 100 points. Most of the existing methods use as a criterion for determinism the flow of nearby trajectories in similar directions, and give more reliable results for longer time series. The method we present, instead, relies on algebraic calculations of the evolution over time of separations between trajectories. In what follows we describe the method, its performance in detecting determinism and discerning chaotic, noisy chaotic and random series, and its application to observed time series from astrophysics, climate and experimental ecology.

Our method exploits the expression $d(t)\sim d_0 e^{\lambda t}$ for the evolution in time of the separation in phase space or reconstructed space $d$ between states that are initially a distance $d_0$ apart in a chaotic system. In contrast, for a random system one expects $d$ to be independent of $d_0$. The $e^{\lambda t}$ dependence is frequently the starting point to estimate Lyapunov exponents.

For short time series of $N$ points, we can generate all $N(N-1)/2$ distances $d_0$ between distinct points. Subsequently, we evolve all states by 1, 2,… time steps, and calculate the distances $d_1, d_2, \ldots, d_j$ after $j$ time steps between all pairs of points. For small enough $j$, we expect the $d_j$ vs $d_0$ dependence to be very different for chaotic and random systems. For chaotic series, we expect a linear relation with a positive slope; the argument is as follows. Suppose we sample pairs of points with a typical separation $\delta$. The iterates after time $t$, $\delta_t$, will separate as $\delta e^{\lambda_L t}$, where the $\lambda_L$ are local separation rates (not rigorously Lyapunov exponents, since the two points will generally not be separated along a principal component). Now, consider pairs of points with a typical separation $k\delta$, $k>1$. These points will be distributed along the attractor with frequencies that roughly follow the invariant distribution, and the ensuing distribution of separation rates will be similar to that of the previous case. Hence, the expected separations after time $t$ will be $\delta_t \sim k\delta e^{\lambda_L t}$, leading to the linear relation between $d_0$ and $d_j$ postulated above.

For random series, we expect no particular correlations between points, and a near-zero slope. These expectations should especially hold as $d_0$ approaches zero. The relation between $d_0$ and $d_j$ is much easier to see if the latter are averaged over small bins of $d_0$, which produces a "skeleton" of the functional form which is easier to visualize (see below, especially Fig. 1).

In order to verify these predictions and the usefulness of the test, we have considered the following time series: periodic and quasiperiodic sinusoidal series, continuous-time chaotic series: the Rössler attractor [12]; discrete-time chaotic series: the Hénon attractor [13]; series resulting from autoregressive (AR) processes [14]; chaotic series with added noise, and finally series drawn from a uniform, random distribution. Unless otherwise specified, we have used an embedding dimension of 2.

Out of the above, we have found that our method has the most difficulties with the first type of series. Average separations fluctuate in periodic and quasiperiodic systems, especially if the sampling frequency is incommensurate with the

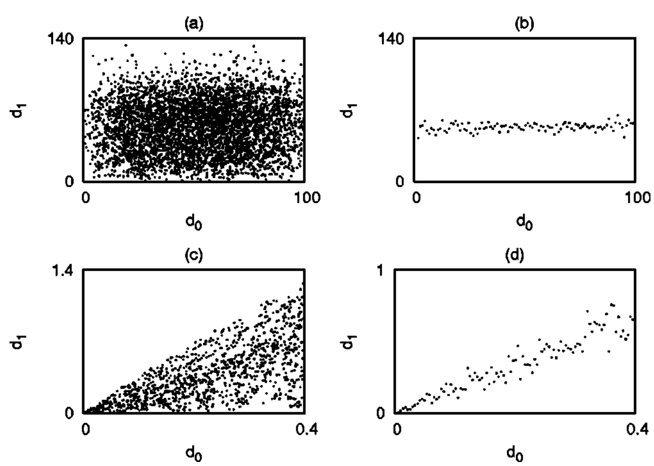


FIG. 1. $d_1$ vs $d_0$ for random series: (a) scatter plot, (b) average; for chaotic (Hénon) series: (c) scatter plot, (d) average.

*Author to whom correspondence should be addressed. Electronic address: pbinder@hawaii.edu

 

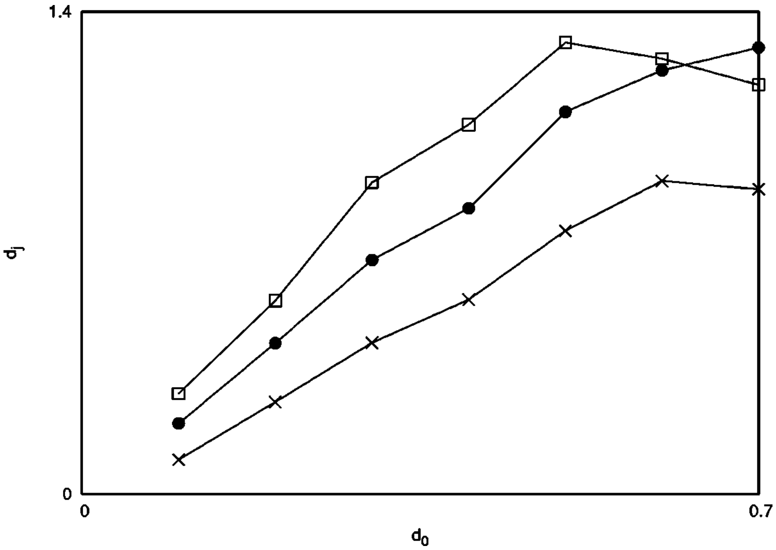


FIG. 2. $d_1$ (crosses), $d_2$ (full circles), and $d_3$ (open squares) vs $d_0$ for the Hénon map time series.

series's frequencies. Fortunately, these are the easiest to identify through standard linear methods such as Fourier analysis, so we will not be further concerned with them.

Figure 1 shows $d_1$ vs $d_0$ plots for two time series. Two-dimensional reconstructions of a series of random numbers, each uniformly distributed between 0 and 100 are shown as (a) a scatter plot and (b) an average over bins of width 2 in $d_0$. For chaotic series (the Hénon map with usual parameters $a=1.4$, $b=0.3$), part (c) is a scatter plot and part (d) is an average over bins of width 0.004. In the random case, the linear correlation $(r)$ coefficient of the skeleton data is 0.196, with slope of 0.031 and vertical intercept of 47.7. As expected, the slope is near zero. For the Hénon data $r=$ $=0.947$, the slope is 1.7 and the intercept is $-0.0023$, also consistent with expectations. The results for the Rössler system will be shown below.

Figure 2 shows averages, in bins of width 0.1, of $d_1$, $d_2$, and $d_3$ vs $d_0$ again for the Hénon map. Two remarks are in order. First, for fixed $d_0$ the separations increase with time $(d_3 > d_2 > d_1)$. This also holds for larger $j$ in the limit $d_0 \to 0$. Second, the $d_j$ are seen to saturate for large enough $d_0$, as they become comparable to the attractor size and no further separation is possible.

Figure 3 shows scatter plots of $d_1$ vs $d_0$ for the Rössler system, with the usual parameter $\mu=5.7$. The data were generated by integrating 300 time steps of size 0.04 with a fourth-order Runge-Kutta method, after discarding transients. Part (a) shows the data for a time difference between $d_0$ and $d_1$ of 16 time steps, comparable to the optimal time delay for attractor reconstruction [15]; part (b) shows the same data for a time difference of 80 time steps, which can also be interpreted as being $d_5$ in the units of part (a). There are two main differences from the results shown in Figure 1. The first is a high degree of internal structure, which persists for times much longer than the optimal reconstruction time. The second is that the slope of $d_1$ does not seem to increase with time; in fact, it is within 1% of 1.0 for all times checked (1 to 80 time steps difference between $d_0$ and $d_1$). Both effects are likely to be a consequence of the continuous-time formulation of this system and the fact that very close trajectories (within 0.1 units) were not found with such a short series. We are unable to offer a detailed explanation for our results. Nevertheless, the linear relation between the bin averaged values of $d_1$ and $d_0$ still exists: the $r$ coefficients for the bin-averaged data are above 0.95.

Next we tested the method with noisy, chaotic series. We studied Hénon series of different lengths (50, 100, and 150 points), each with added uniform noise of amplitudes of 10, 20,…,90% of the attractor size. Typical results for 50 points are shown in Fig. 4. While for 10% noise the $d_1$ vs $d_0$ curve

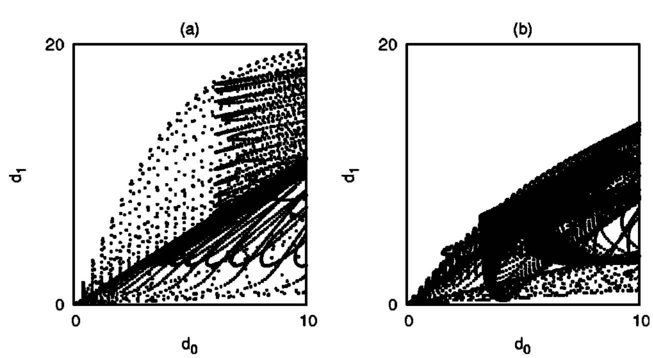


FIG. 3. Rössler time series $d_1$ vs $d_0$ for iteration over (a) 16 and (b) 80 time steps.

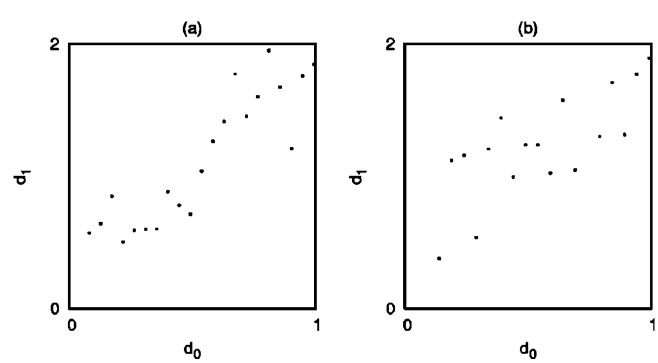


FIG. 4. $d_1$ vs $d_0$ for 50-point Hénon time series with noise amplitude of (a) 10%, (b) 30%.

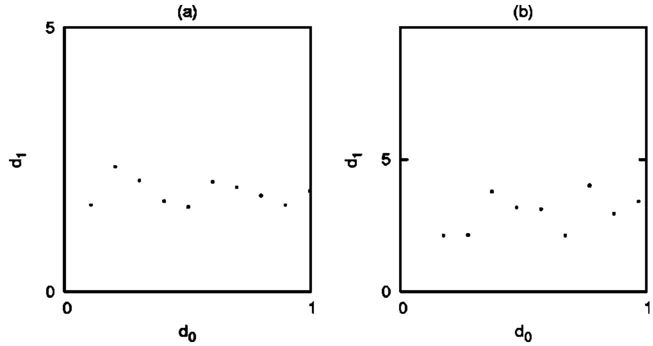


FIG. 5. $d_1$ vs $d_0$ for two autoregressive time series, with (a) coefficients comparable to the noise level, (b) coefficients much larger than the noise level (values given in the text).

appears to go through the origin, for 30% noise the intercept is slightly above 0. The ambiguity continues until for 50% or more noise the intercept is so high that one has to conclude that determinism is not present. To summarize the results, for the 50-point series the method detects determinism with up to 30% noise; for the longer series, determinism is seen with up to 50% added noise. As one would expect, having a longer series helps detect determinism masked by higher levels of noise.

AR processes [14] are important statistical models that contain deterministic and stochastic components. We have studied second-order AR processes, $x_t = a_1 x_{t-1} + a_2 x_{t-2} + \nu_t$, where $\nu_t$ is a Gaussian random variable with mean zero and standard deviation one. In Fig. 5 we present typical results for two cases. In part (a) the coefficients $(a_1, a_2) = (1, -0.5)$ are comparable to the noise, and in part (b) the coefficients $(a_1, a_2) = (10, -9)$ are considerably larger than the noise. In both cases the slope is close to 0 and the intercept is about 3, indicating absence of determinism in AR models. To explain this apparent failure of our method, we note that the noise $\nu$ is part of the dynamics, and hence greatly alters the values of the $x_t$ series, in contrast with the previous case, in which noise is added after the dynamics. This explains why our method fails with this apparently simple, linear process.

We then analyzed two observed data sets with this method. The first series is the light curve for KPD 1930 +2752, a pulsating subdwarf B star (Fig. 9 in [16], lower curve). Short-period variations have been filtered out, and a four-frequency ellipsoidal modulation has been subtracted, which leaves a series of 100 residuals, which are presumably random. We studied the latter with the method described above.

The second is a series of 41 triplets (larvae, pupae, adults), collected every two weeks, of the population of the *Tribolium* flour beetle [17]. The series is replicate 13 in p. 218, with mortality and cannibalism coefficients $(\mu_a, c_{pa})$ adjusted so that the system is chaotic. In addition to "demographic" fluctuations, the measurements in the series are non-negative integers, which coarsens phase space.

The results for these data sets are shown in Fig. 6. Part (a) shows $d_1$ vs $d_0$ for the subdwarf data. Here $r = -0.46$, the slope is $-2.6 \times 10^{-4}$, and the vertical intercept is 0.27. Clearly, the skeleton of $d_1$ is nearly flat, and we can conclude

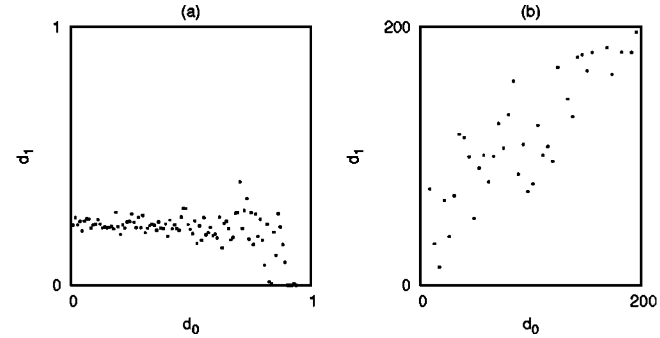


FIG. 6. $d_1$ vs $d_0$ for (a) pulsating subdwarf and (b) flour beetle population time series, as described in the text.

that the data are random. Part (b) shows $d_1$ vs $d_0$ for the flour beetle data. Here $r = 0.42$, the slope is 0.67, and the intercept is 55. Since the slope is much greater than zero, we can conclude that the data set is deterministic, as was established in [17].

The final example conveys two points. The first is to document the method's performance with a series studied by other methods: the Southern Oscillation Index (SOI) series (see [6]). The second is to attempt to provide more rigor in the application and interpretation of this method, and to explore the limitations in doing so.

Figure 7(a) shows the now-familiar $d_1$ vs $d_0$ plot for the SOI series, embedded in three dimensions. The result clearly points to absence of determinism, in agreement with what was reported in Ref. [6], using other methods [8,9]. However, the figure presented here was obtain with only 100 points of the SOI series and relatively little computation time, in contrast with the other two methods, in which well over 1000 points were used.

In what we have shown so far, the method relies on the partially subjective choices of what a suitably "small" $d_0$ should be, and how well the $d_1$ vs $d_0$ curve can be described by a straight line going through the origin. Below, we attempt to adapt the ideas of surrogate data testing [7] to add rigor to our method.

To generate a series with the same invariant distribution but without (deterministic) temporal correlations, we just need to shuffle the data points. We have done this, and show the $d_1$ vs $d_0$ results in part (b) of Fig. 7. If the series were

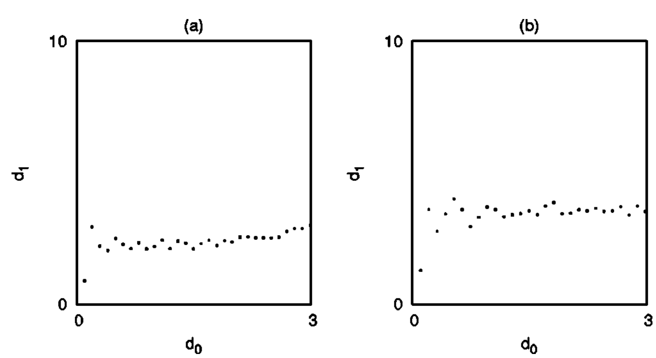


FIG. 7. $d_1$ vs $d_0$ for Southern Oscillation Index series (100 points), embedded in three dimensions: (a) original series, and (b) reconstruction of shuffled series.

deterministic, one would expect this curve to show far more random characteristics than the original one; however, if the series is nondeterministic, as is the case here, the results should not change very much. Indeed, both parts of the figure show a clear nonzero intercept and very shallow slope. In principle, one can compare a given metric (slope, intercept) with that of an ensemble of shuffled time series. One can then attach statistical significance to the metric of the original series by calculating how close it is, in standard deviations, to the mean of the shuffled samples: if it is atypical of a random replicate, it is more likely to be deterministic. However, since the choice of the $d_0$ cutoff is still arbitrary, we recommend at best generating one surrogate data series, as we did in this example.

To conclude, we have developed a method to discern random and chaotic behavior in short time series, with even fewer than 100 points. The method generates all $O(N^2)$ distances $d_0$ between points, and for small enough $d_0$, looks at the evolution of these distances in time. If the separations after $j$ time steps $d_j$ are independent of $d_0$, the series is likely to be random; if there is a linear $d_j$ vs $d_0$ relation with near-zero intercept and a significant slope, the series is likely to be deterministic. The method appears to be robust even if the series has considerable amounts of noise. We have applied the method to three observed series; it seems to discriminate convincingly between random and chaotic behavior. While we have attempted to find a way for the method to yield statistically significant results, a certain degree of subjectivity in its application ultimately precludes such a development.

This work was supported by the Research Corporation. We thank A. Steindamm, R. Figueroa-Centeno, and H. Read for assistance with data processing.

[1] H. D. I. Abarbanel, *Analysis of Observed Chaotic Data* (Springer, New York, 1996).

[2] H. Kantz and T. Schreiber, *Nonlinear Time Series Analysis* (Cambridge University Press, Cambridge, U.K., 1999).

[3] H. D. I. Abarbanel, R. Brown, J. J. Sidorowich, and L. Sh. Tsimring, Rev. Mod. Phys. **65**, 1331 (1993).

[4] D. R. Fredkin and J. A. Rice, Phys. Rev. E **51**, 2950 (1995).

[5] G. B. Pasternack, Adv. Water Resour. **23**, 253 (1999).

[6] P.-M. Binder and C. A. Wilches, Phys. Rev. E **65**, 055207(R), (2002).

[7] T. Schreiber and A. Schmitz, Physica D **142**, 246 (2000).

[8] D. T. Kaplan and L. Glass, Phys. Rev. Lett. **68**, 427 (1992).

[9] R. Wayland, D. Bromley, D. Pickett, and A. Passamante, Phys. Rev. Lett. **70**, 580 (1993).

[10] D. T. Kaplan, Physica D **73**, 38 (1994).

[11] J. S. Lih, J. K. Ko, J. L. Chern, and I. M. Jiang, Europhys. Lett. **40**, 7 (1997).

[12] O. E. Rössler, Z. Naturforsch. A **31**, 259 (1976).

[13] M. Hénon, Commun. Math. Phys. **50**, 69 (1973).

[14] P. J. Brockwell and R. A. Davis, *Introduction to Time Series and Forecasting* (Springer, New York, 1996).

[15] A. M. Fraser and H. L. Swinney, Phys. Rev. A **33**, 1134 (1986).

[16] M. Billères, G. Fontaine, P. Brassard, S. Charpinet, J. Liebert, and R. A. Saffer, Astrophys. J. **530**, 441 (2000).

[17] J. M. Cushing, R. F. Costantino, B. Dennis, R. A. Desharnais, and S. M. Henson, *Chaos in Ecology* (Academic, San Diego, 2003).