PHYSICAL REVIEW
LETTERS

VOLUME 59 | 24 AUGUST 1987 | NUMBER 8

Predicting Chaotic Time Series

J. Doyne Farmer and John J. Sidorowich[a]
Theoretical Division and Center for Nonlinear Studies, Los Alamos National Laboratory, Los Alamos, New Mexico 87545
(Received 22 April 1987)

We present a forecasting technique for chaotic data. After embedding a time series in a state space using delay coordinates, we "learn" the induced nonlinear mapping using a local approximation. This allows us to make short-term predictions of the future behavior of a time series, using information based only on past values. We present an error estimate for this technique, and demonstrate its effectiveness by applying it to several examples, including data from the Mackey-Glass delay differential equation, Rayleigh-Bénard convection, and Taylor-Couette flow.

PACS numbers: 05.45.+b, 02.60.+y, 03.40.Gc

One of the central problems of science is forecasting: Given the past, how can we predict the future? The classic approach is to build an explanatory model from first principles and measure initial data. Unfortunately, this is not always possible. In fields such as economics, we still lack the "first principles" necessary to make good models. In other cases, such as fluid flow, the models are good, but initial data are difficult to obtain. We can derive partial differential equations that allow us to predict the evolution of a fluid (at least in principle), but specification of an initial state requires the measurement of functions over a three-dimensional domain. Acquisition of such large amounts of data is usually impossible. Typical experiments employ only a few probes, each of which produces a single time series. Partial differential equations simply cannot operate on such data. In either case, when we lack proper initial data or when we lack a good model, we must resort to alternative approaches.

Such an alternative is exemplified by the work of Yule,[1] who in 1927 attempted to predict the sunspot cycle by building an *ad hoc* linear model directly from the data. The modern theory of forecasting[2] as it has evolved since then views a time series $x(t_i)$ as a realization of a random process. This is appropriate when effective randomness arises from complicated motion involving many independent, irreducible degrees of freedom.

An alternative cause of randomness is chaos,[3] which can occur even in very simple deterministic systems. While chaos places a fundamental limit on long-term prediction,[3] it suggests possibilities for *short-term* prediction: Random-looking data may contain simple deterministic relationships, involving only a few irreducible degrees of freedom. In chaotic fluid flows, for instance, experimental[4] and theoretical results[5] indicate that in some cases the state space collapses onto an attractor of only a few dimensions.

In this paper we present a method to make predictions about chaotic time series. These ideas were originally inspired by efforts to beat the game of roulette, in collaboration with Packard.[6]

If the data are a single time series, the first step is to embed it in a state space. Following the approach introduced by Packard *et al.*,[7] and put on a firm mathematical basis by Takens,[7] we create a state vector $\mathbf{x}(t)$ by assigning coordinates $x_1(t)=x(t)$, $x_2(t)=x(t-\tau)$, ..., $x_d(t)=x(t-(d-1)\tau)$, where τ is a delay time. If the attractor is of dimension D, a minimal requirement is that $d \geq D$.

The next step is to assume a functional relationship between the current state $\mathbf{x}(t)$ and the future state $\mathbf{x}(t+T)$,

$$\mathbf{x}(t+T)=f_T(\mathbf{x}(t)). \tag{1}$$

We want to find a predictor F_T which approximates f_T.

If the data are chaotic, then f_T is necessarily nonlinear. There are several possible approaches: One can assume a standard functional form, such as an mth-order polynomial in d dimensions, and fit the coefficients to the data set using least squares.[8] Forecasts for longer times $2T, 3T, \ldots$, can then be made by composing F_T with itself. This approach has the disadvantage that errors in approximation grow exponentially with composition. An alternative is to fit a new function F_T for each time T. This has the advantage that global approximation techniques only work well for smooth functions—and higher iterates of chaotic mappings are *not* smooth. Yet another approach is to recast Eq. (1) as a differential equation and write $\mathbf{x}(t+T)$ as its integral. All of these approaches suffer from the problem that the number of free parameters for a general polynomial is $(m+d)!/(m!d!) \approx d^m$, which is intractable for large d.

Our preliminary results suggest that a more effective approach is the *local approximation*, using only nearby states to make predictions. To predict $x(t+T)$ we first impose a metric on the state space, denoted by $\| \ \|$, and find the k nearest neighbors of $\mathbf{x}(t)$, i.e., the k states $\mathbf{x}(t')$ with $t' < t$ that minimize $\|\mathbf{x}(t) - \mathbf{x}(t')\|$. We then construct a local predictor, regarding each neighbor $\mathbf{x}(t')$ as a point in the domain and $x(t'+T)$ as the corresponding point in the range. The simplest approach to construct a local predictor is approximation by nearest neighbor, or *zeroth-order approximation*, i.e., $k=1$ and $x_{\text{pred}}(t,T) = x(t'+T)$. A superior approach is the first-order, or linear, approximation, with our taking k greater than d, and fitting a linear polynomial to the pairs $(\mathbf{x}(t'), x(t'+T))$. For convenience we usually treat the range as a scalar, mapping d-dimensional states into one-dimensional values, although for some purposes it is desirable to let the range be d dimensional. The fit can be made in any of several ways; for the work reported here we did least squares by singular-value decomposition. When $k=d+1$ this is equivalent to linear interpolation, but to ensure stability of the solution it is frequently advantageous to take $k > d+1$. We have also experimented with approximation using higher-order polynomials, but in higher dimensions our results are not significantly better than those obtained with first order.

If done in the most straightforward manner, finding a neighboring value in a data base of N points requires the order of N computational steps. This can be reduced to $\log N$ by the partitioning of the data in a decision tree.[9] Furthermore, once the neighbors are found predictors for many times T can be computed in parallel. With these speedups the computations reported here can be done on small computers.

To facilitate the comparison of results, in this paper we simply build the data base from the first part of the time series, and hold it fixed as we make predictions on the remainder. Alternatively, it is possible for one to optimize the performance with respect to either memory or data limitations by dynamically updating the data base.

To evaluate the accuracy of our predictions, we compute the root-mean-square error, $\sigma_\Delta(T) = \langle [x_{\text{pred}}(t,T) - x(t+T)]^2 \rangle^{1/2}$. For convenience we normalize this by the rms deviation of the data $\sigma_x = \langle (x - \langle x \rangle)^2 \rangle^{1/2}$, forming the *normalized error* $E = \sigma_\Delta(T)/\sigma_x$. If $E=0$, the predictions are perfect; $E=1$ indicates that the performance is no better than a constant predictor $x_{\text{pred}}(t,T) = \langle x \rangle$. To estimate E we make as many predictions as we need for reasonable convergence, typically on the order of 1000.

We have applied our method to several artificial and experimental systems, including the logistic map,[3] the Hénon map,[3] the Mackey-Glass delay-differential equation,[10] Taylor-Couette flow,[11] and Rayleigh-Bénard convection in an ^{3}He-^{4}He mixture.[12] Our results are summarized in Table I.

An illustration of the performance of the local linear approximation is given in Fig. 1, with use of convection data obtained by Haucke and Ecke.[12] The dimension of this time series is $D \approx 3.1$ (see Ref. 12). To compare with a "standard forecasting technique," we also show

TABLE I. A summary of forecasts using local linear approximation for several different data sets. D is an estimate of the attractor dimension, d is the embedding dimension, N is the number of data points used, $T_{\max}$ is the rough prediction time (Ref. 13) at which the normalized error approached 1, and t_{char} is the "characteristic time" for the time series estimated as the inverse of the mean frequency in the power spectrum. In comparison, a standard forecasting technique (global linear autoregression) gave $T_{\max}$ values that were typically about one characteristic time.

	Differential delay[a] $t_d =$				Rayleigh-Bénard[b] $R/R_c =$			Couette[c] $R/R_c =$		
	17	23	30	100	10.55	12.19	12.24	10.2	12.9	13.7
D	2.1	2.7	3.5	10	2.0	2.6	3.1	2.0	2.7	3.1
d	4	4	6	18	5	6	6	10	6	6
N	10^4	2×10^4	2×10^4	10^5	10^4	10^4	3×10^4	10^4	3×10^4	3×10^4
$T_{\max}$	600	300	300	150	∞	1000 s	100 s	∞	3 s	1 s
t_{char}	50	55	60	65	3 s	2 s	1.5 s	0.5 s	0.4 s	0.1 s

[a]Reference 10.
[b]Reference 12.
[c]Reference 11.

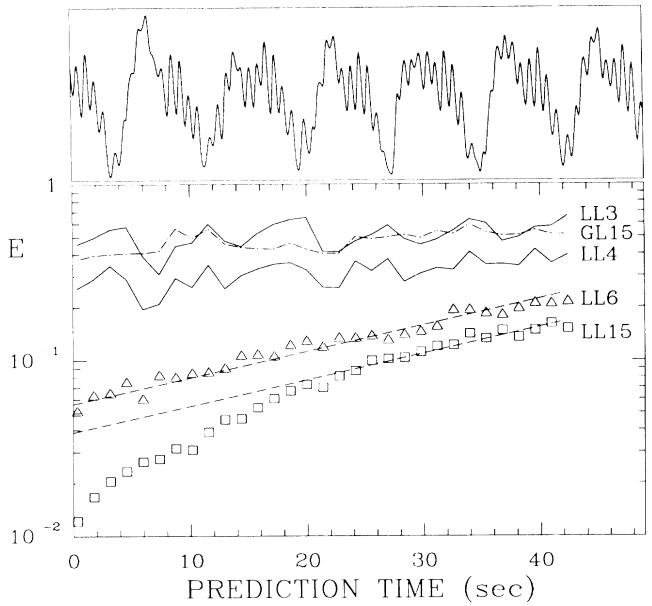

FIG. 1. Top: A time series obtained from Rayleigh-Bénard convection in an ^{3}He-^{4}He mixture (Ref. 12), with Rayleigh number $R/R_c = 12.24$, and dimension $D \approx 3.1$. Bottom: The normalized error $E(T) = \sigma_\Delta(T)/\sigma_x$. The top and bottom time scales are the same. We show results for the local linear (LL) and global linear (GL) methods; numbers following the initials indicate the embedding dimension. The dashed lines are from Eq. (2), with k equal to the computed metric entropy from Ref. 12, and C determined by a least-squares fit.

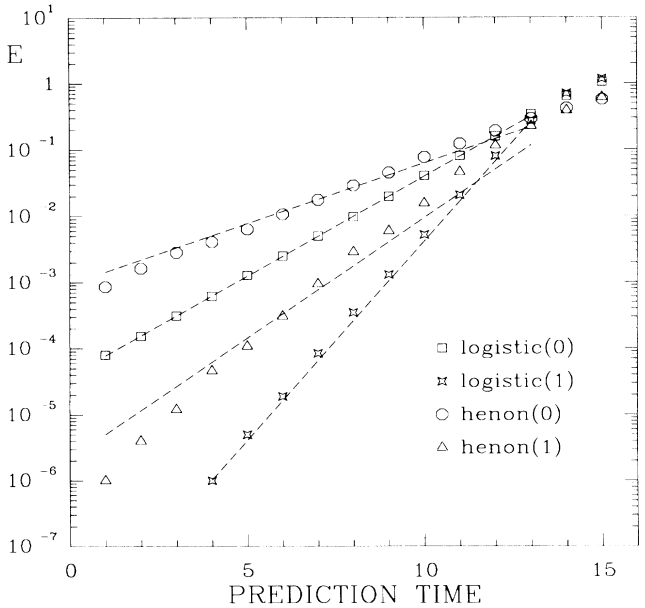

FIG. 2. The normalized error as a function of the prediction time T, for the logistic and Hénon maps (Ref. 3). Results are shown using zeroth-order (0) and first-order (1) local approximation. The dashed lines are from Eq. (2), with a least-squares fit for C and the positive Lyapunov exponents, $k = \log 2$ for the logistic map, and $k \approx 0.42$ for the Hénon map (Ref. 14).

results obtained using a global linear approximation (linear autoregression[2]). When $d < D$, the quality of prediction for the local approximation is roughly the same as that obtained with the global linear approach, but for d sufficiently large the predictions are significantly better.

How well does this local approximation work? This depends on the parameters of the problem, including the number of data points N, the attractor dimension D, the metric entropy h, the signal-to-noise ratio S, and the prediction time T. There are two distinct regimes: If the typical spacing between data points, $\epsilon \approx N^{-1/D} < S^{-1}$, then the forecast is limited by noise. Following Shaw,[3] the average information in a prediction is $\langle I(T) \rangle \approx \ln S - hT$. For a narrowly peaked distribution with $E \ll 1$, to first order $\langle I(T) \rangle$ is proportional to $-\ln E$.

The second regime occurs when $\epsilon > S^{-1}$ and the accuracy of forecasts is limited by the number of data points. In this case, providing d is sufficiently greater than D, in the limit that $E \ll 1$ we propose the following error estimate:

$$E \approx C e^{(m+1)kT} N^{-(m+1)/D}, \tag{2}$$

where m is the order of approximation and C is a constant. k equals the largest Lyapunov exponent when $m = 0$, and equals the metric entropy otherwise. The arguments leading to this formula are too involved to report here, but they are based on the following facts: The error of interpolation in one dimension is proportional to $f^{(m+1)}\epsilon^{m+1}$; to leading order the mth derivative grows under iteration as the mth power of the first derivative, and the average derivatives along the unstable manifold grow according to the positive Lyapunov exponents. Detailed arguments leading to this result will be presented elsewhere.[14]

The scalings predicted by Eq. (2) are illustrated in Figs. 2 and 3. The exponential increase of $E(T)$ is demonstrated in Fig. 2, for numerical experiments on the logistic and Hénon maps.[3] For the logistic map the slopes are very close to those predicted. For the Hénon map, a least-squares fit gives slopes about 10% greater than those expected with the positive Lyapunov exponent, indicating a possible correction[14] to Eq. (2). For the convection data on Fig. 1, agreement with computed values of the metric entropy[12] is very good as indicated in the figure.

Note that setting $E(T_{\max}) = 1$ in Eq. (2) yields $T_{\max} = (\ln N)/kD$, independent of the order m. Thus zeroth-order interpolation is less effective than first order, except when E is the order of 1. Equation (2) suggests that higher-order polynomial interpolation might be more effective, but this is difficult in more than two dimensions.

The power-law variation of E with N is illustrated in

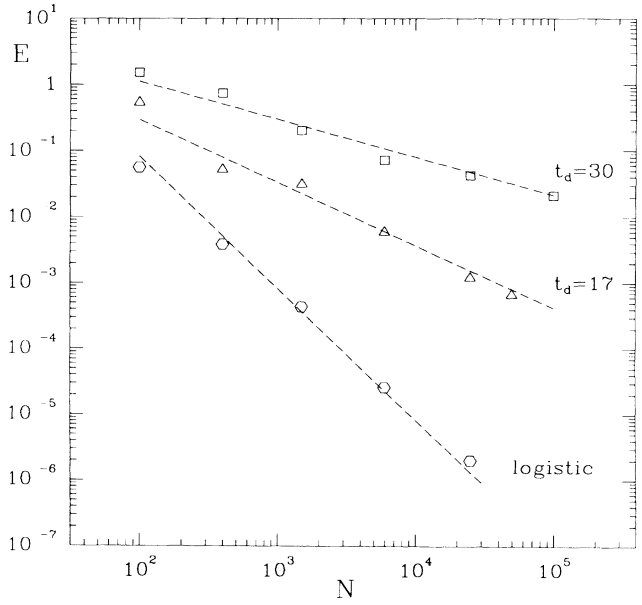

FIG. 3. The normalized error with use of local linear approximation as a function of the number of data points N, at fixed prediction time T. For the logistic map (Ref. 3) $r=4$ and $T=3$; for the Mackey-Glass delay-differential equation (Ref. 10) $T=40$, with two values of the delay parameter t_d. The dashed lines are from Eq. (2), with $D=1$ for the logistic map and $D=D_L$, the Lyapunov dimension, from Ref. 10 for Mackey-Glass equation.

Fig. 3, where we show the behavior for the logistic map and the Mackey-Glass equation. The agreement of the slopes with the expected values of $(m+1)/D$ based on computations of the Lyapunov dimension[10] is quite good. This scaling law breaks down for large N, on account of an approach to the noise floor (when $\epsilon < S^{-1}$).

In addition to the obvious practical applications of forecasting, the construction of approximate models can be a useful diagnostic tool to investigate chaos. Equation (2) demonstrates how forecasting can be used to estimate dimension and entropy; Lyapunov exponents are also easily obtained. Forecasting provides a way to determine whether the resulting numerical values of dimension and entropy are reliable. Ultimately, the ability to forecast successfully with deterministic methods may be the strongest test of whether or not low-dimensionality chaos is present.

At this point, this work is still in a preliminary stage and many possibilities remain to be investigated. In a future paper we plan to compare the local approximation in more detail to some other approaches, in particular recursive–ordinary-differential-equation models such as neural nets.

In this paper we have shown that a forecasting approach based on deterministic chaos can be quite effective in predicting low- to moderate-dimensionality time series. Furthermore, this can be done with reasonable amounts of data and computer time. The most important point is not the specific technique, but rather the demonstration that approaches based on deterministic chaos can be effective; we expect that new and better techniques will emerge rapidly as more attention is focused on this problem. Such methods should be effective for problems in fluid dynamics, control theory, artificial intelligence, and possibly even economics.

We would like to thank Scott Konishi for assistance on numerical procedures, Glenn Carter for assistance in reading nonstandard computer tape formats, and Ian Percival, Josh Deutsch, and particularly Norman Packard for valuable conversations. We would also like to thank Steve Omohundro for making us aware of the literature on decision trees. In a different context, he has recently independently proposed a local linear approximation as an efficient means of performing artificial-intelligence tasks.[15]

[a]Permanent address: Department of Physics, University of California at Santa Cruz, Santa Cruz, CA 95064.

[1]G. U. Yule, Philos. Trans. Roy. Soc. London A **226**, 267 (1927).

[2]For example, S. M. Pandit and S.-M. Yu, *Time Series and System Analysis with Applications* (Wiley, New York, 1983).

[3]For reviews see, for example, R. S. Shaw, Z. Naturforsch. **36a**, 80 (1981); J. P. Crutchfield, J. D. Farmer, N. H. Packard, and R. S. Shaw, Sci. Am. **254** (No. 12), 46 (1986); J. Ford, Phys. Today **36** (No. 4), 40 (1983).

[4]For example, G. Mayer-Kress, *Dimensions and Entropies in Chaotic Systems* (Springer-Verlag, Berlin, 1986).

[5]N. Aubry, P. Holmes, J. L. Lumley, and E. Stone, Cornell University Report No. FDA-86-15, 1986 (to be published).

[6]T. A. Bass, *The Eudaemonic Pie* (Houghton-Mifflin, New York, 1985).

[7]N. H. Packard, J. P. Crutchfield, J. D. Farmer, and R. S. Shaw, Phys. Rev. Lett. **45**, 712 (1980); F. Takens, in *Dynamical Systems and Turbulence,* edited by D. A. Rand and L.-S. Young (Springer-Verlag, Berlin, 1981).

[8]D. Gabor, W. P. Wilby, and R. Woodcock, Proc. IEEE **108B**, 422 (1960).

[9]J. L. Bentley and J. H. Friedman, A. C. M. Comput. Surv. **11**, 297 (1979).

[10]M. C. Mackey and L. Glass, Science **197**, 287 (1977); J. D. Farmer, Physica (Amsterdam) **4D**, 366 (1982).

[11]A. Brandtstater, J. Swift, H. L. Swinney, A. Wolf, J. D. Farmer, E. Jen, and J. P. Crutchfield, Phys. Rev. Lett. **51**, 1442 (1983); A. Brandtstater and H. L. Swinney, Phys. Rev. A **35**, 2207 (1987).

[12]H. Haucke and R. Ecke, Physica (Amsterdam) **25D**, 307 (1987).

[13]T_{max} is a crude estimate of our ability to forecast. In some cases $E(T)$ reaches a plateau at a level less than 1. In these cases we estimate T_{max} by extrapolating the initial rate of increase. (In some cases such as GL15 in Fig. 1, it is necessary to expand the T axis to see the initial increase.)

[14]Detailed arguments suggest that there may be a correction to the T scaling in Eq. (2), but we have not yet resolved this.

[15]S. Omohundro, "Efficient algorithms with neural network behavior" (to be published).