

## Predicting physical variables in time-delay embedding

Henry D. I. Abarbanel\*

*Department of Physics and Marine Physical Laboratory, Scripps Institution of Oceanography, University of California, San Diego, Mail Code 0402, La Jolla, California 92093-0402*

T. A. Carroll and L. M. Pecora

*U.S. Naval Research Laboratory, Code 6341, Washington, D.C. 20375-5000*

J. J. Sidorowich and L. S. Tsimring

*Institute for Nonlinear Science, University of California, San Diego, Mail Code 0402, La Jolla, California 92093-0402*

(Received 30 September 1993)

In time-delay reconstruction of chaotic attractors we can accurately predict the short-term future behavior of the observed variable  $x(t) = x(n) = x(t_0 + \tau_s n)$  without prior knowledge of the equations of motion by building local or global models in the state space. In many cases we also want to predict variables other than the one which is observed and require methods for determining models to predict these variables in the same space. We present a method which takes measurements of two variables  $x(n)$  and  $z(n)$  and builds models for the determination of  $z(n)$  in the phase-space made out of the  $x(n)$  and its time lags. Similarly we show that one may produce models for  $x(n)$  in the  $z(n)$  space, except where special symmetries prevent this, such as in the familiar Lorenz model. Our algorithm involves building local polynomial models in the reconstructed phase space of the observed variable of low order (linear or quadratic) which approximate the function  $z(n) = F(\mathbf{x}(n))$  where  $\mathbf{x}(n)$  is a vector constructed from a sequence of values of observed variables in a time delay fashion. We train the models on a partial data set of measured values of both  $x(n)$  and  $z(n)$  and then predict the  $z(n)$  in a recovery set of observations of  $x(n)$  alone. In all of our analyses we assume that the observed data alone are available to us and that we possess no knowledge of the dynamical equations. We test this method on the numerically generated data set from the Lorenz model and also on a number of experimental data sets from electronic circuits.

PACS number(s): 05.45.+b, 84.40.-x

### I. INTRODUCTION

The analysis of data from chaotic systems rests on the use of the embedding theorem which allows the reconstruction of the system phase space (or state space) from scalar observations. In this method [1–5] a scalar measurement  $x(n) = x(t_0 + n\tau_s)$  and its time delays  $x(n + kT)$ ;  $k = 1, 2, \dots, d-1$  are used to make  $d$ -dimensional vectors

$$\mathbf{y}(n) = \left[ x(n), x(n+T), \dots, x(n+(d-1)T) \right], \quad (1)$$

whose components provide a coordinate system in which one can identify the attractor structure associated with the observations. The delay  $T\tau_s$  is an integer multiple of the sampling time  $\tau_s$ . We choose the integer  $T$  by considerations of average mutual information [6].

It is now quite a routine effort to make models of the future behavior of the variable  $x(n)$  by working within the  $d$ -dimensional embedding space of the vectors  $\mathbf{y}(n)$ . These models are either global in phase space or consist of local neighborhood-to-neighborhood maps in the space  $\mathbb{R}^d$ . If one knows that the global space is larger than the dynamical space of the local dynamics  $d_L$ , then the maps

are locally  $d_L$  dimensional [7]. The reason for making models with  $d_L$  degrees of freedom, but in a  $d_E \geq d_L$ -dimensional space, is so that the trajectory is unfolded from its projection on the observation axis  $x(n)$ , and we can unambiguously identify points on the orbit which are dynamical neighbors. These models are of the form  $\mathbf{y}(n+1) = F(\mathbf{y}(n))$  with  $F(\cdot)$  a map from  $\mathbb{R}^{d_E}$  to itself.

In physical settings one is almost always interested in additional information beyond the measured variable  $x(n)$ , though the ability to measure other variables along with the measurements of  $x(n)$  may often be quite circumscribed. An example which comes readily to mind is the measurement of the atmospheric pressure at some location from which one would like to be able to infer the horizontal wind velocity at that location or perhaps at another location. Other examples are easy to construct.

If one has measured only the  $x(n)$ , then absent any knowledge of the connection between this variable and the others there seems no clear way to estimate the behavior of another dynamical variable, call it  $z(n)$ , of the system. In typical experimental situations the equations of motion are either not available or not reliable. Even if given, they usually have a form of some set of differential equations, ordinary or partial, which describes the evolution in a space which is often much larger than that required to embed the observed data, and therefore is difficult to deal with.

In this paper we explore the idea that if one were to

\*Also at Institute for Nonlinear Science, University of California, San Diego, La Jolla, CA 92093-0402.

measure another dynamical variable or dynamical variables, call them  $z(n)$ , in addition to the single scalar  $x(n)$ , then one could make a model of the connection between  $z(n)$  and the time-lagged vector  $\mathbf{y}(n)$ , which are the data as seen in the global reconstructed embedding space. This possibility follows from the natural connection among all the variables in a nonlinear system, and the embedding theorem [2–4] then suggests that if we have used the proper number of time delays to establish coordinates for the attractor, then other variables should be dependent on the  $\mathbf{y}(n)$  in that coordinate system. Thus we are led to consider a connection

$$z(n) = g(\mathbf{y}(n)), \quad (2)$$

which we can deduce from the data by working in the  $d$ -dimensional space of the  $\mathbf{y}(n)$ . Since we typically do not have the differential equations of the system or, more to the point, we may be trying to learn differential equations from the data, we have no other path to the connection between one physical variable, the  $z(n)$ , and measurements of the  $\mathbf{y}(n)$ .

To establish the connection  $z(n) = g(\mathbf{y}(n))$  we must have some measurements of  $z(n)$  along with  $x(n)$ . So our scheme will be based on the scenario where we initially have some measurements of *both*  $x(n)$  and  $z(n)$ . From these measurements we establish the phase space using the  $x(n)$ , and in that space we determine the function  $g(\mathbf{y}(n))$  which gives  $z(n)$ . With this construction we may predict from future measurements of the  $x(n)$  alone the

$$\mathbf{y}(n) = (x(n - DT), x(n - (D - 1)T), \dots, x(n), \dots, x(N + (D - 1)T)), \quad (3)$$

while for  $d$  odd,  $d = 2D + 1$ , we use

$$\mathbf{y}(n) = (x(n - DT), x(n - (D - 1)T), \dots, x(n), \dots, x(n + DT)). \quad (4)$$

Although the reconstruction theorem is insensitive to this choice, in practice time-symmetric embedding gives more accurate results as compared to a standard backward time delay embedding in forecasting models. The reason for this is that in the time-symmetric embedding, the loss of correlation between  $z(n)$  and the last components of the vector  $\mathbf{y}(n)$  is reduced.

Denoting each of these vectors as  $\mathbf{y}(n) = (y_1(n), y_2(n), \dots, y_d(n))$ , we write the local polynomial relationship as

$$z(n) = \sum_{i_1, \dots, i_d} a_{i_1, \dots, i_d} y_1^{i_1}(n) \dots y_d^{i_d}(n), \quad (5)$$

where  $\sum_{j=1}^d i_j \leq M$ , and  $M$  is the maximum order of the polynomial. The number of parameters to be found from the data is  $\sim d^M$ . There is always a trade-off between the order of the polynomials and the state space locality for the maps. Indeed, the larger is  $M$ , the more parameters must be determined by the local fit; hence with a fixed number of data points in the training set the larger the neighborhoods around the orbit points  $\mathbf{y}(n)$  must be to accurately establish the coefficients  $a_{i_1, \dots, i_d}$ .

In fact, for *local* reconstruction there is no need for very high-order polynomials, and we will use only linear

evolution of  $z(n)$ . This is just it would be if we knew the underlying differential equations or map.

The nonlinear functional reconstruction techniques which can be used for this purpose are essentially the same as those employed for conventional nonlinear model building [5,8]. They are roughly divided into local and global techniques. Local methods represent a nonlinear function  $g(\mathbf{y}(n))$  as a collection of local polynomial (or other) maps, different for different neighborhoods in the phase space [9]. The parameters of the individual local maps are determined by locally fitting the points in a data set contained within a small region in the phase space. More precisely, we use the phase-space neighbors of the orbit points  $\mathbf{y}(n)$  and make maps from the neighborhood of  $\mathbf{y}(n)$  to the neighborhoods of  $\mathbf{y}(n + 1)$ . The parameters in the local map are determined by a least-squares fit to the evolution of whole neighborhoods of points. There are now well-developed methods for finding neighbors in multidimensional phase space [5,10]. In contrast with these local neighborhood-to-neighborhood maps, global methods provide a single nonlinear function for the entire data set [11,12]. Global models are smaller and thus more convenient to deal with; however, local models are usually more accurate.

In this paper we use a local polynomial representation of the function  $z(n) = g(\mathbf{y}(n))$ . Further, instead of vectors built on time delays solely from the future or the past, we take advantage of the knowledge we have of both past and future. For  $d$  even,  $d = 2D$ , we use vectors

or quadratic maps. For a linear map we write the relation  $z(n) = g(\mathbf{y}(n))$  in the form

$$z(n) = a_0(n) + \mathbf{a}(n) \cdot \mathbf{y}(n), \quad (6)$$

emphasizing that the local constants depend on the phase-space location. For this local linear model we have  $d + 1$  parameters at each local neighborhood in phase space. We typically would use  $N_B = 2(d + 1)$  total neighbors to make the determination of the constants  $a_0$  and  $\mathbf{a}$ .

Our algorithm works as follows. We present the program with two parallel streams of scalar measurements  $x(n)$  and  $z(n)$ .  $N$  pairs of  $d$ -dimensional vectors  $\mathbf{y}(n)$  and corresponding  $z(n)$  are made from the  $x(n)$  as just described. First  $N_T$  of the total pairs  $\{\mathbf{y}(n), z(n)\}$  are used for “training” the local models as we describe in a moment. A  $kd$  tree [10] for these  $N_T$  is now made so the search for nearest neighbors can be done quickly. On the remaining  $N_R = N - N_T$  pairs we wish to check recovery of the value of the  $z(j)$ ;  $j = N_T + 1, N_T + 2, \dots, N$  from the data stream of  $x(j)$ ;  $j = N_T + 1, N_T + 2, \dots, N$  alone. To start this recovery for point  $z(j)$ , we search our  $kd$  tree to find the  $N_B$  nearest neighbors  $\mathbf{y}^{(k)}(j)$ ;  $k = 1, 2, \dots, N_B$  of  $\mathbf{y}(j)$  which come from the first  $N_T$  data. Using these neighbors, a local linear or quadra-

tic map is constructed by a least-squares fit employing all  $N_B$  required neighbors to find a map of the form  $z(\cdot)=g(\mathbf{y}(\cdot))$  which minimizes residuals over all  $N_B$  neighbors from the training data. This means minimize

$$\sum_{k=1}^{N_B} |z^{(k)}(j) - g_j(\mathbf{y}^{(k)}(j))|^2 \quad (7)$$

at each "time"  $j$ .  $N_B$  is taken to be twice the number of parameters, so  $N_B=2(d+1)$  for a linear fit, and  $N_B=(d+1)(d+2)$  for a quadratic fit. From this local map we determine the recovered  $z_r(j)$  by

$$z_r(j) = g_j(\mathbf{y}(j)) . \quad (8)$$

Since we also have recorded the true value of  $z(j)$  for each of the  $N_R$  values of  $j$ , direct comparison of our accuracy in recovery is possible. Let us emphasize again that the coefficients  $a_0$  and a constituting the model  $g_j$  are different for different data vectors  $\mathbf{y}(j)$ .

If we can measure both variables all the time, then it is sometimes more useful to use time delays of both variables for the coordinates of the state space [1]. This approach was tested by Casdagli and Weigend [13], who used previous values of chest volume *and* heartbeat rate to forecast future values of the heart rate. The main difference in our algorithm is that at the stage of forecasting no use is made of information about the previous dynamics of the variable to be predicted. We envision situations where in a laboratory or other controlled setting one can measure both  $x(n)$  and  $z(n)$ . However, in some application in the field or other environment, only one variable can accurately or efficiently or reliably be measured. The connection between the variables determined in the controlled setting then allows us to determine the other dynamical quantity from the first one in other situations where we may not be able to observe both quantities.

For example, we can suggest a setting where in a laboratory we are able to measure the pressure at the wall of a turbulent boundary layer flow and the horizontal velocity magnitude at some interesting point just above the wall. If we measure both and establish the dynamical connection between them on the attractor, then in future observations where only the wall pressure may be available, we can deduce the flow velocity without further measurement and with some confidence [14].

This paper deals with the implementation of these ideas on the Lorenz attractor taken as a standard model of low-dimensional dynamics, and on data from several nonlinear circuits built and run in our laboratories.

## II. LORENZ MODEL

The Lorenz model [15]

$$\begin{aligned} \dot{x} &= \sigma(y - x) , \\ \dot{y} &= -xz + rx - y , \\ \dot{z} &= xy - bz , \end{aligned}$$

with parameter values  $\sigma=16$ ,  $b=4$ , and  $r=45.92$  serves as a paradigm for low-dimensional chaotic systems. We

will show how one can recover the behavior of the variable  $z(t)$  from a model constructed in a state space spanned by vectors with components made from values of  $x(t)$ .

We proceed by taking measured values of both  $x(n)$  and  $z(n)$ , with the sampling time being the time step introduced in the integration of the Lorenz system of magnitude  $\tau_s=0.02$  (see Fig. 1). Using average mutual information we easily establish that  $T=5$  is appropriate. With this value of the time lag, we learn from a global false nearest-neighbor calculation that a global embedding dimension of  $d_E=3$  will unfold the attractor with time-lagged values of  $x(n)$  as coordinates in the space.

Now taking simultaneous data from  $x(n)$  and  $z(n)$ , we make a quadratic local neighborhood to neighborhood map in  $d_E=5$ . Of course,  $d_E=3$  would do, but it is known that taking an embedding dimension larger than the minimum suggested by false nearest neighbors can often improve the prediction, especially if noise is present [16]. This is also true in our case of predicting one physical variable from another (see text below and Fig. 3). So the map  $z(n)=g(\mathbf{y}(n))$  has five-dimensional vectors  $\mathbf{y}(n)=(x(n-2T), x(n-T), x(n), x(n+T), x(n+2T))$  as its argument. For test purposes we split our data set of  $N=10\,000$  ( $x, z$ ) pairs into a training set of  $N_T=9000$  and a recovery set of  $N_R=1000$  pairs. First we build a  $kd$  tree in the five-dimensional space. For each  $x(j)$  from the recovery set we construct a vector  $\mathbf{y}(j)$  and find the  $N_B$  nearest neighbors  $\{\mathbf{y}^{(k)}(j), k=1, \dots, N_B\}$  among the training set. Using the  $\mathbf{y}^{(k)}(j)$  and their images  $z^{(k)}(n)$ , we fit the coefficients of the local map (5). Finally we substitute our test vector  $\mathbf{y}(n)$  into (5) and obtain the recovered value  $z_r(j)$  of the variable  $z(j)$ .

We varied the time delay  $T$ , always keeping  $d_E=5$ , and determined how well our reconstructed values  $z_r(j)$  match the known values of  $z(j)$ .  $T=1$  is used in Fig. 2(a), where we have the values of the direct error  $z(n)-z_r(n)$  over the recovery set of 1000 points. Note the change in scale as compared to Fig. 1. Figure 2(b) shows the recovery error for  $T=2$ ; note the substantial improve-

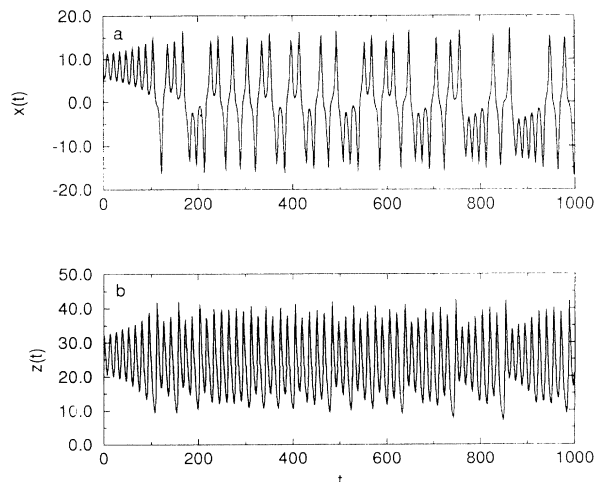


FIG. 1. Time series for two variables from the Lorenz system (9);  $\tau_s=0.02$  s: (a)  $x(t)$ , and (b)  $z(t)$ .

ment in recovery with a larger  $T$ . Big glitches in the prediction error of Fig. 2(a) are clearly correlated with the points where the local derivative  $\partial z/\partial x$  becomes large. This effect is especially significant near  $x=0$ . In these points small variations of  $x$  are accompanied by large variations of  $z$ , which is known to cause a noise amplification phenomenon [16]. As one can see from Fig. 2, the best way to diminish this effect is to choose the optimal (non-small) value of time delay. At  $T=5$ , as shown in Fig. 2(c), we have still better predictability—note the change in scale. In Figs. 2(d) and 2(e), the values of  $T$  are  $T=7$  and  $12$ , respectively, and we see the quality of the local map predicting  $z(n)$  degrade slowly. In a sense this is another method for determining a time delay for the reconstructions of the state space. It rests on a quite different aspect of the data from mutual information, and while good predictability sounds attractive, it seems to us to have little fundamental meaning in terms of the properties of dynamical systems.

In Fig. 3 the root-mean-squared error  $\sigma$  averaged over all data points

$$\sigma^2 = \frac{\langle (z - z_r)^2 \rangle - \langle (z - z_r) \rangle^2}{\langle z^2 \rangle - \langle z \rangle^2}, \quad (9)$$

is shown for a fixed time delay,  $T=5$ —the one which worked best—as a function of the embedding dimension. We show this result both for the local quadratic map just discussed and for a local linear map. For each there is a

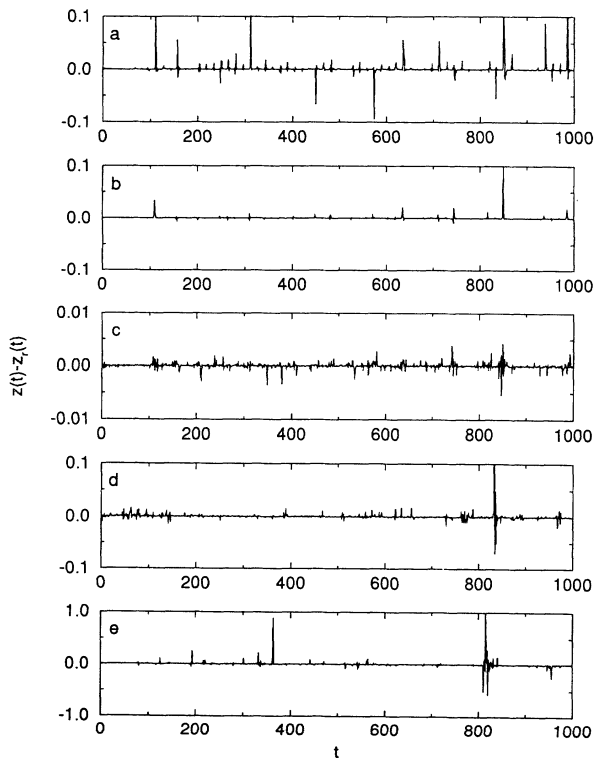


FIG. 2. Recovery error for modeling the  $z$  variable of the Lorenz system in a phase space composed of vectors built from time delays of the  $x$  variable. Each analysis used  $d_E=5$  and a local quadratic fit. The time delay used in the phase-space vectors was varied in these calculations: (a)  $T=1$ ; (b)  $T=2$ ; (c)  $T=5$ ; (d)  $T=7$ ; and (e)  $T=12$ .

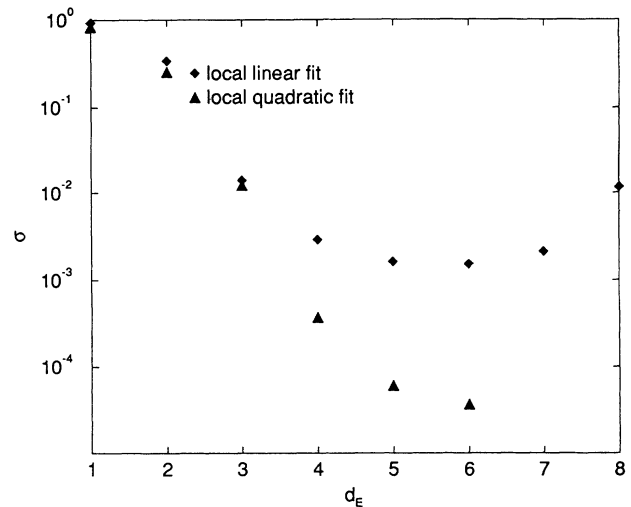


FIG. 3. Root-mean-square (rms) error from modeling  $z(n)$  on  $x(n)$  using data from the Lorenz system as a function of the embedding dimension  $d_E=5$ . In each analysis  $T=5$ . Results are shown for both local linear and local quadratic maps.

sharp decrease in  $\sigma$  as we pass through  $d_E=3$ , as is consistent with the false nearest-neighbor determination of  $d_E$ . For the local quadratic map the error continues to decrease with  $d_E$ , until we run out of data points in higher dimensions with which to determine a good map  $y(n) \rightarrow z(n)$ . Figure 4 looks at the method from the point of view of variation in time delay for fixed  $d_E$ . A minimum in  $\sigma$  is seen with both linear and quadratic local models. The rms error in each is quite small, and the message is that while quadratic models provide a substantial improvement in predictability, for many purposes they may not be required if prediction at the 1% level is sufficient.

It is noteworthy that the inverse problem, namely reconstructions of  $z(n)$  from  $x(n)$ , is impossible for Lorenz system because of its symmetry  $x \rightarrow -x$ ,

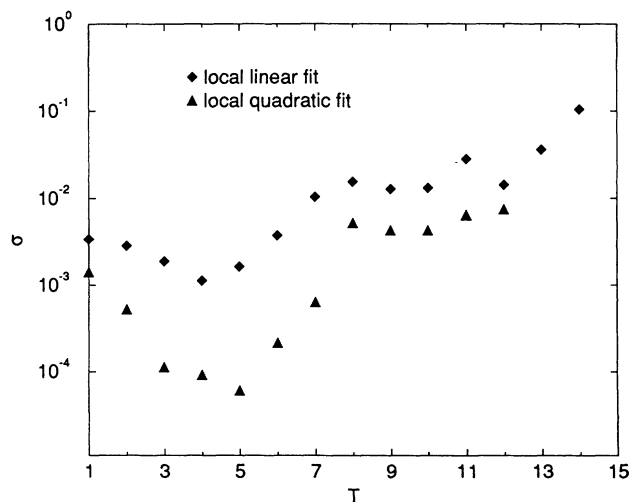


FIG. 4. rms error from modeling  $z(n)$  on  $x(n)$  using data from the Lorenz system as a function of the time delay  $T$ . In each analysis  $d_E=5$ . Results are shown for both local linear and local quadratic maps.

$y \rightarrow -y, z \rightarrow z$ . Indeed, each  $z$  can correspond to two different values  $x$  and  $-x$ , so there is no way to determine the sign of  $x$  from the knowledge of  $z$ . However, we believe that this sort of symmetry is rather exceptional, and in most cases the reconstruction is possible in both directions.

### III. HYSTERETIC CIRCUIT

#### A. Description of the circuit

The hysteretic circuit used for our first set of experimental data has been described earlier [17] and consists in essence of an unstable second degree oscillator which operates in the range of several hundred Hz connected to a hysteretic element. Diagrams for the circuit can be found in the earlier paper. The circuit is similar to some described by Newcomb and co-workers [18].

The hysteretic circuit can be modeled by three ordinary differential equations:

$$\begin{aligned} \frac{dx_1(t)}{dt} &= x_2(t) + \gamma x_1(t) + cx_3(t), \\ \frac{dx_2(t)}{dt} &= \omega x_1(t) - \delta_2 x_2(t), \\ \epsilon \frac{dx_3(t)}{dt} &= [1 - x_3(t)^2][Sx_1(t) - D + x_3(t)] - \delta_3 x_3(t), \end{aligned} \quad (10)$$

with  $\gamma=0.2$ ,  $c=2.2$ ,  $\delta_2=0.001$ ,  $\delta_3=0.001$ ,  $\epsilon=0.3$ ,  $\omega=10.0$ ,  $S=1667.0$ , and  $D=0.0$ . The constants in the equation for  $x_3(t)$  are set to give a square hysteretic loop. We present these equations for illustration purposes only; the data we will be using was not generated from the equations, but taken directly from the experimental circuit. Two voltages were observed in the experiment,  $V_A(t)$  and  $V_B(t)$ . The voltage  $V_A(t)$  corresponds to the variable  $x_1(t)$  above, and  $V_B(t)$  is described by  $x_2(t)$ . Both voltages are in the order of 0–10 V, and were digitized at  $\tau_s=0.1$  ms. Samples of digitized and rescaled voltages are shown in Figs. 5(a) and 5(b). As always in an experiment, there is some noise present in the data. In this particular case the noise level was quite low, on the order of a few mV; that is, about 0.1% noise.

#### B. Analysis of the data: two observed scalars

We have previously reported some analysis of one of the voltages from the hysteretic circuit [5], and repeat part of that for completeness. The Fourier power spectra for these data are not displayed here both because they appear elsewhere and little is learned from them except that the spectrum is broad in each set. The two data sets have been called  $V_A(n)$  and  $V_B(n)$ ; each was 64 000 points in length. In Fig. 6 we show the results of computing the average mutual information in each data set as a function of time lag; the physical units of the sampling time are  $\tau_s=0.1$  ms. We see that each data set has its first minimum at  $T=6$ , and we use that value throughout our analysis of these data. In Fig. 7 we present the result of the global false nearest-neighbor computation for each

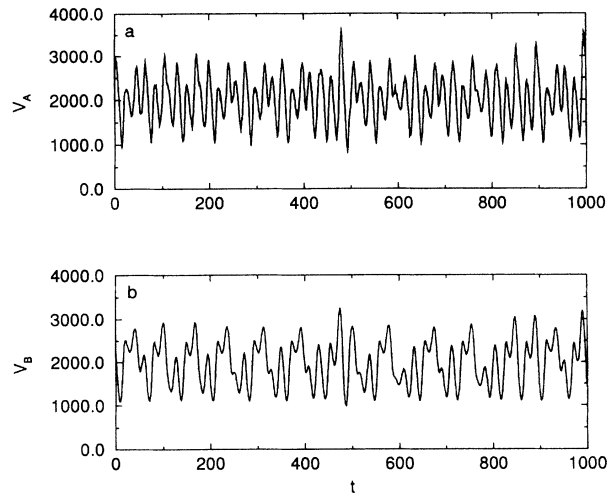


FIG. 5. Two voltages  $V_A$  and  $V_B$  taken at the same time from the hysteretic circuit described in the text. Here and in all other data sets both voltages were offset and normalized to obtain a better resolution upon digitizing with a 12-bit analog-to-digital converter (ADC), so the relative scales are not preserved (for our predicting algorithm it does not matter). The sampling time  $\tau_s=0.1$  ms.

data set. The  $V_B(n)$  data are certainly embeddable in  $d_E=3$ , while  $V_A(n)$  data appear to require  $d_E=5$ .

This result in itself is quite illuminating. We see that using different coordinates can result in rather different choices for global embedding dimension. This is consistent with the embedding theorem, which gives only a sufficient condition for the global embedding dimension  $d_E$  in terms of the fractal (box counting) dimension  $d_A$ ; namely the integer  $d_E$  must be  $d_E > 2d_A$ . In the present case we will argue below that the fractal dimension (to be precise the Lyapunov dimension) of this system is about 2.4, so an integer embedding dimension greater than twice this, namely  $d_E=5$ , is sufficient. The false nearest-

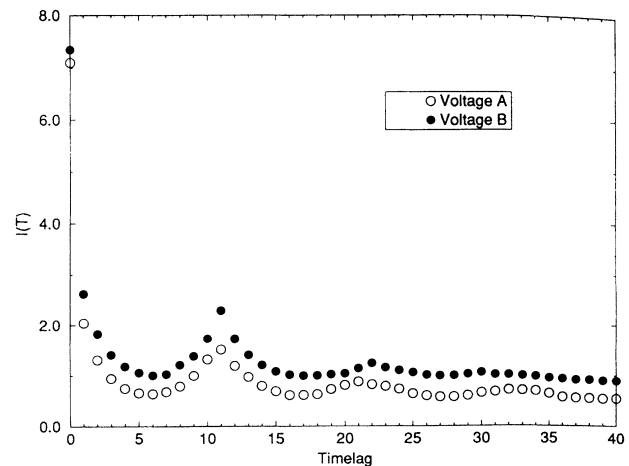


FIG. 6. Average mutual information as a function of time delay for signals  $V_A$  and  $V_B$  from the hysteretic circuit. 64 000 data points were used. The units of the time lags are 0.1 ms. Each signal has its first minimum at  $T=6$ .

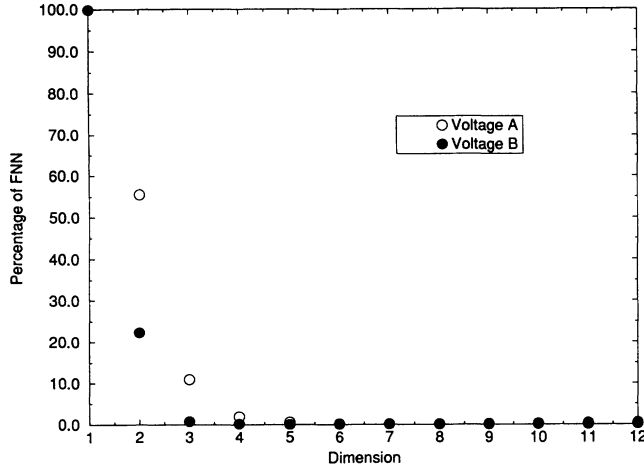


FIG. 7. Percentage of global false nearest neighbors as a function of the embedding dimension for the signals  $V_A$  and  $V_B$  from the hysteretic circuit,  $T=6$ , 47 000 points.  $V_A$  requires dimension  $d_E=5$  to unfold the attractor, while  $V_B$  only requires  $d_E=3$ .

neighbor method determines for a given data set what the *necessary* embedding dimension must be for time-delay reconstruction of the attractor using a given set of time-delayed measurements. It need not be the same necessary embedding dimension for all choices of reconstruction coordinates, since one set is some unknown nonlinear transformation from another, and this can twist the attractor in such a fashion as to require a higher dimension to unfold it via time-delay coordinates. The embedding theorem tells us with certainty when we can stop unfolding, as additional dimensions could not possibly achieve any further unfolding of the attractor.

In Figs. 8 and 9 we show the result of evaluating *local*

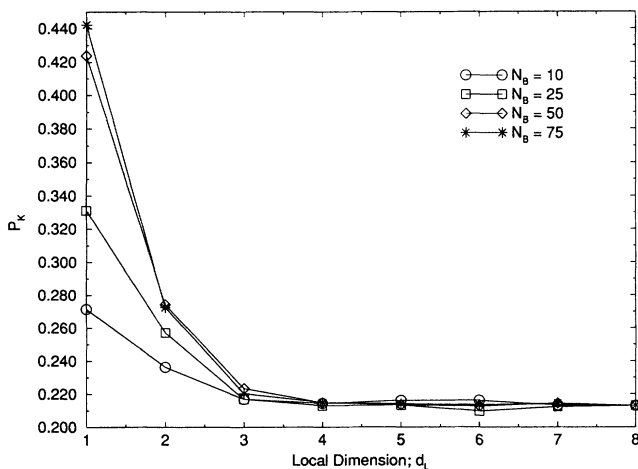


FIG. 8. Local false nearest-neighbor test for  $V_A$  from the hysteretic circuit. Here  $P_K$  is the percentage of bad predictions, and  $N_B$  is the number of neighbors used for local predictions. We used 60 000 data points, time delay  $T=6$  and  $\beta=0.3$ , where  $\beta$  is the fraction of the attractor size two neighboring points must separate in defining a bad prediction (Ref. [7]). It is important that  $d_L=3$  is identified here as in Fig. 9.

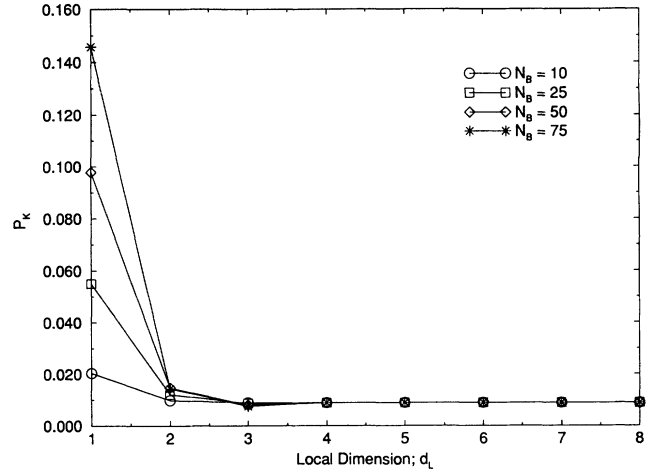


FIG. 9. Local false nearest-neighbor test for  $V_B$  from the hysteretic circuit, the same parameter values as in Fig. 8.  $d_L=3$  is identified here as in Fig. 8.

*false nearest neighbors* [7] for these two data sets. The local dimension must be the same for each data set regardless of which global embedding dimension is chosen. This local dimension is the (integer) number of degrees of freedom active in the orbits making up the attractor. Whatever the global twisting of the attractor in one coordinate system or another, we must have the same local dynamical dimension. This number also determines what dimension model will be required locally as well as the number of Lyapunov exponents are to be considered true and not artifacts of the choice of global embedding. We see in these two figures that both  $V_A(n)$  and  $V_B(n)$  indicate that  $d_L=3$  for this system. Using data from the  $V_B(n)$  data set, we have evaluated the average local Lyapunov exponents [19,20] in global dimensions  $d_E=4, 5$ , and  $6$ , with  $d_L=3$  in each case. The analyses are consistent regardless of  $d_E$ . It is from this that we are able to deduce that the Lyapunov dimension is  $D_L \approx 2.4$ . One of the local exponents goes rapidly to zero, with the number of time steps along the attractor indicating that ordinary differential equations are the appropriate model for the dynamics producing these observations. This, of course, is totally in accord with earlier studies of the hysteretic circuit.

### C. Modeling $A$ on $B$ and $B$ on $A$

Now we turn to the discussion of how well we can predict the evolution of one of the data sets from making local polynomial models for the variable we wish to predict based on a phase space reconstructed using the other variable. First let us see how well we can do using  $V_A(n)$  to create the state space. We reconstructed the phase space with time lags of  $V_A(n)$  and then made *local linear models* of  $V_B(n)=g(y_A(n))$  using various values of the

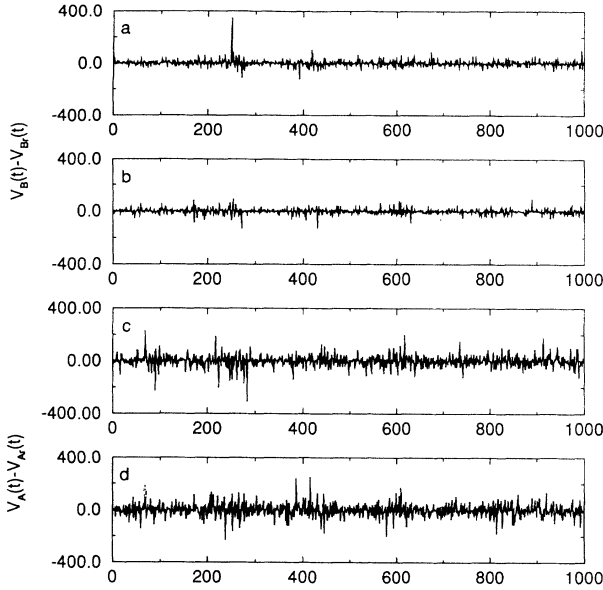


FIG. 10. Data from the hysteretic circuit. Recovery errors for modeling  $V_B$  using  $V_A$ : (a) and (b). Recovery errors for modeling  $V_A$  using  $V_B$ : (c) and (d). (a)  $d_E=6, T=6$ ; (b)  $d_E=7, T=6$ ; (c)  $d_E=3, T=6$ ; and (d)  $d_E=6, T=6$ . We used 50 000 points in the training set and employed local linear fits.

dimension of the state space of the  $\mathbf{y}_A(n)$ . All our analysis is done at  $T=6$ . The training data consisted of  $N_T=50\,000$ , and  $N_R=1000$  points were recovered. In Figs. 10(a) and 10(b) we show the  $V_B(n) - V_{B_r}(n)$  errors in the recovery set for  $d_E=6$  and 7, respectively. Although the overall performance in both cases is similar, using a higher embedding dimension removes large deviations in recovery which are probably related to regions in the phase space where we do not have unfolding in  $d_E=6$  but do have it in  $d_E=7$ .

Turning to our ability to model the behavior of  $V_A(n)$  from observations of  $V_B(n)$ , we show in Figs. 10(c) and 10(d) the results of making local linear models for  $V_A(n)$  in terms of  $V_B(n)$  in  $d_E=3$  and 6. The ability to reconstruct the attractor in  $d_E=3$  as revealed by the global false nearest-neighbor test now shows up in our improved ability to predict  $V_A$  from observations of  $V_B$ . In  $d_E=6$  we see little improvement over the results in  $d_E=3$ ; indeed, the effects of limited, possibly noisy data show up here.

#### IV. OTHER CIRCUITS

##### A. Four-dimensional

##### Van der Pol–Duffing oscillator: Circuit 1

The data for this circuit are from two points on a circuit used to study the synchronization of cascaded chaotic systems [21]. A circuit diagram can be found in the original paper. This circuit is described by four differential equations:

$$\begin{aligned} \frac{dx_1(t)}{dt} &= -1098[1.466x_1(t) - 2.466x_2(t) \\ &\quad + x_3(t) - x_4(t) + q_1(x_4(t))], \\ \frac{dx_2(t)}{dt} &= -10\,980x_1(t), \\ \frac{dx_3(t)}{dt} &= -4972x_2(t), \\ \frac{dx_4(t)}{dt} &= -10\,980 \left[ \frac{-100\,000}{R_v}x_1(t) + A_4x_4(t) \right. \\ &\quad \left. + q_2(x_4(t)) \right], \\ q_1(x) &= 0.5(|x - 2.5| - |x + 2.5|), \\ q_2(x) &= 0.5x + 0.164(|x - 1.3| - |x + 1.3|) \\ &\quad + 0.361(|x - 2.6| - |x + 2.6|). \end{aligned} \quad (11)$$

The parameters are set to  $A_4=0.5$  and  $R_v=79\,000\ \Omega$ ;  $R_v$  is the control parameter in this system. The functions  $q_1(x)$  and  $q_2(x)$  provide the nonlinear portions of the circuit. They are piecewise linear and are designed to make the circuits easy to match, so that they may be used in studies of chaotic synchronization. The parameters in the circuits were set so that a single well chaotic attractor was produced. For different settings, it was also possible to see a chaotic attractor that hopped between two centers. The data set  $V_A(t)$  contains the signal  $x_1(t)$ , digitized with 12-bit resolution at  $\tau_s=50\ \mu\text{s}$ . The data file  $V_B(t)$  contains the signal  $x_4(t)$ , digitized at the same rate.

Now we examine the observed  $V_A(t)$  and  $V_B(t)$  data. The time series are seen in Figs. 11(a) and 11(b). The average mutual information for both data sets is seen in Fig. 12, where a minimum at  $T=8$  (or 0.4 ms in dimensional units) is suggested for  $V_A(t)$  and  $T=7$  (0.35 ms) for  $V_B(t)$ . Using these values for the time delay, we eval-

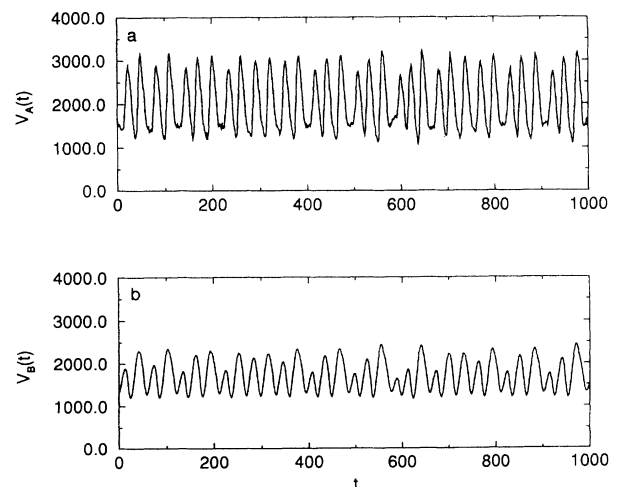


FIG. 11. Two voltages  $V_A$  and  $V_B$  taken at the same time from circuit 1 as described in the text. They were digitized with 12-bit resolution.  $\tau_s=50\ \mu\text{s}$ .

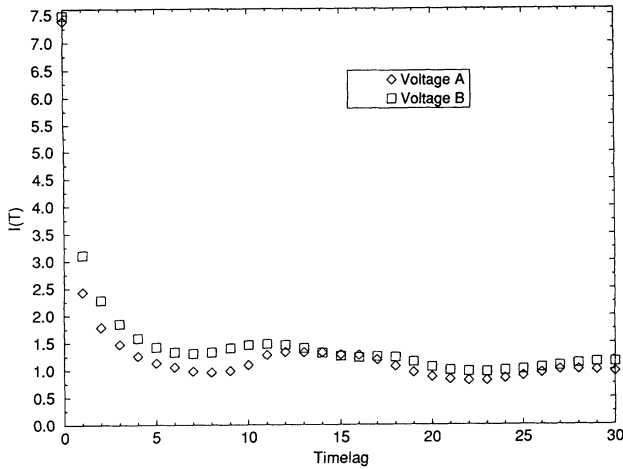


FIG. 12. Average mutual information as a function of time delay for signals  $V_A$  and  $V_B$  from circuit 1. 64 000 data points were used. The units of the time lags are  $50 \mu s$ .  $V_A$  has its first minimum of  $I(T)$  at  $T=8$ , while  $V_B$  has its first minimum at  $T=7$ .

uated the global false nearest-neighbor percentage as a function of embedding dimension, and found in Fig. 13 that  $d_E=4$  is clearly suggested for  $V_A(t)$ , and possibly for  $V_B(t)$  as well, although in the latter case  $d_E=3$  may do. To further examine this we look at the local false nearest neighbors which are displayed for  $V_A(t)$  in Fig. 14 for the data read forward in time, and in Fig. 15 for the data read backward in time. Each clearly distinguishes  $d_L=4$ , consistent with the known circuit equations.

For the data from  $V_B(t)$ , looking at the local false nearest neighbors results in Figs. 16 and 17, where  $d_L=4$  is selected, in accord with the analysis of the voltage  $V_A(t)$  just given.

Now working in  $d_E=5$  and using  $T=8$  for the time delay, we first constructed a phase space based on voltage  $V_A(t)$  and then created a local linear map

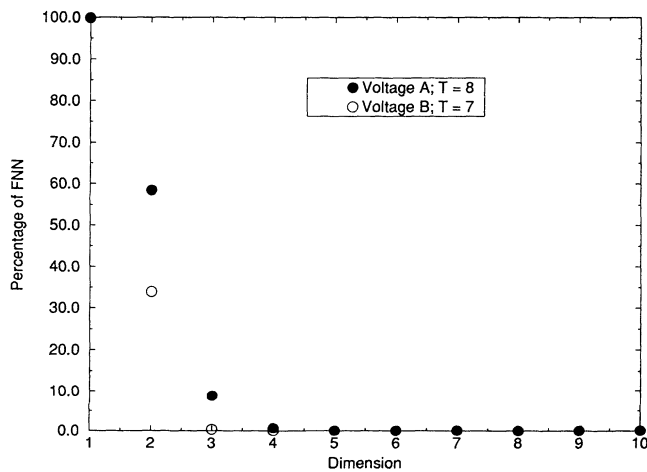


FIG. 13. Percentage of global false nearest neighbors as a function of the embedding dimension for the signals  $V_A$  and  $V_B$  from circuit 1, 37 000 data points.  $V_A$  requires dimension  $d_E=5$  to unfold the attractor, while  $V_B$  only requires  $d_E=4$ .

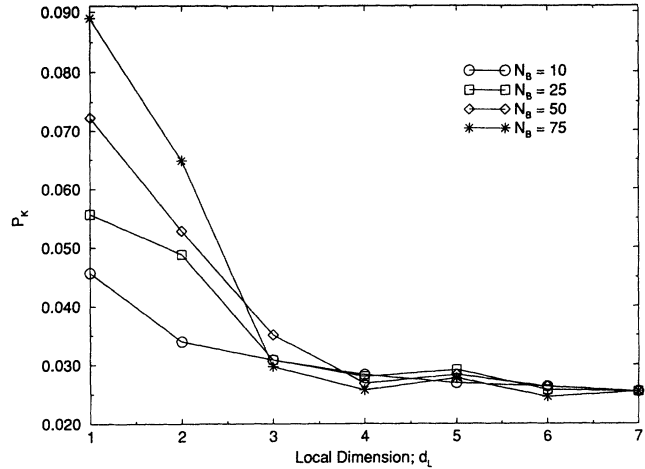


FIG. 14. Local false nearest neighbors for  $V_A$  from circuit 1,  $T=8$ ,  $\beta=0.4$ , and  $N=60\,000$ . It is important that  $d_L=4$  is identified here as for  $V_B$ .

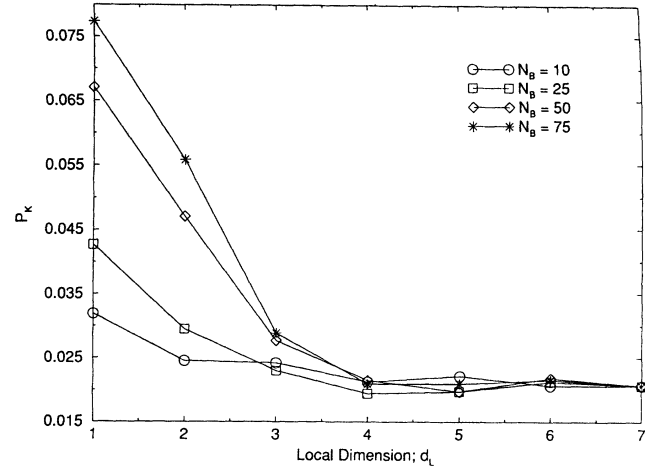


FIG. 15. The same as Fig. 14, but with the data read backward in time.

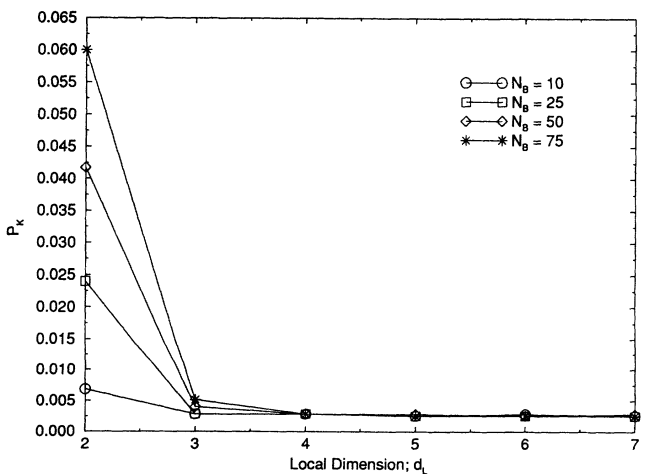


FIG. 16. Local false nearest neighbors for  $V_B$  from circuit 1,  $T=7$ ,  $\beta=0.1$ , and  $N=60\,000$ .  $d_L=4$  is identified here as for  $V_A$ .



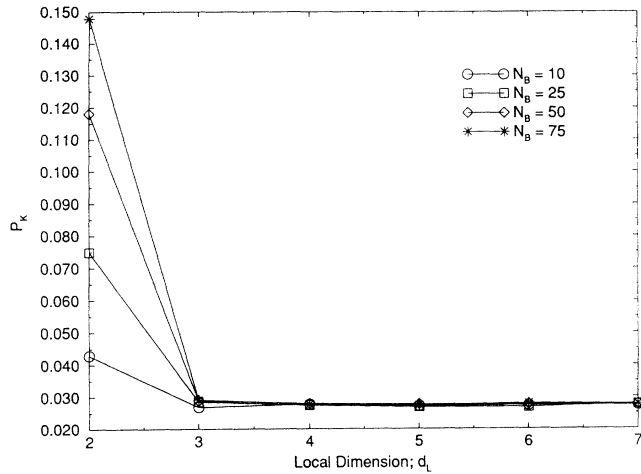


FIG. 17. The same as in Fig. 16, but with the data read backward in time.

$V_B(n) = g(V_A(n-2T), V_A(n-T), V_A(n), V_A(n+T), V_A(n+2T))$  which allows us to predict the values of voltage  $B$  when we observe only voltage  $A$ . Similarly we then used voltage  $B$  to create a phase space of dimension  $d_E = 5$ , and made a local predictor for voltage  $A$  in that space. In Fig. 18 we see the results of making these predictors a local linear map. In Tables I and II we present a summary of all our studies of circuit 1 associated with recovery of voltage  $A$  from voltage  $B$  and  $B$  from  $A$  with different time delays and embedding dimensions. The last column of both tables shows the rms error of recovery calculated by (9).

**B. Single well duffing circuit: Circuit 4**

A single well Duffing equation circuit was used to generate the data for voltages  $V_A(t)$  and  $V_B(t)$  in this exam-

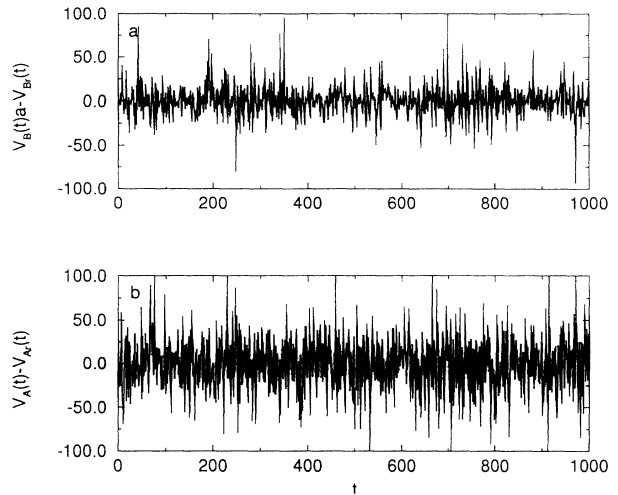


FIG. 18. Data from circuit 1. Recovery errors for modeling  $V_B$  using  $V_A$  (a), and  $V_A$  using  $V_B$  (b). We used 50 000 points in the training set and employed local linear fits. In each case  $d_E = 5, T = 8$ .

ple. This Duffing circuit used analog multiplier chips for the cubic part of the Duffing equations. The output signals for  $x$  and  $y$  are contained in files  $V_A(t)$  and  $V_B(t)$ , each digitized at  $\tau_s = 20 \mu s$  (see Fig. 19). The equations describing this circuit are

$$\begin{aligned} \frac{dx(t)}{dt} &= 100\,000y(t), \\ \frac{dy(t)}{dt} &= 100\,000[B \cos(\omega t) - 0.256y(t) - x(t)^3], \end{aligned} \tag{12}$$

where the driving frequency  $f$  ( $\omega = 2\pi f$ ) was 769 Hz, and the driving amplitude was  $B = 6.0$  V. Part of the dynamics in this circuit comes from the fact that the cubic function, which is executed by analog multiplier chips, satu-

TABLE I. A summary of the various local models made to recover  $V_B$  from  $V_A$  using data from circuit 1. It is clear that the method is rather robust under changes of the embedding dimension  $d_E$  and time delay  $T$ , in the vicinity of the values selected by global false nearest neighbors and average mutual information, respectively.

Time delay	rms errors		
	Circuit 1		
	Recovering voltage $B$ from voltage $A$		
	Embedding dimension	Order of local maps	rms error $B_r - B$
1	5	Linear	0.504
3	5	Linear	0.105
5	5	Linear	0.0528
7	5	Linear	0.0545
9	5	Linear	0.0477
11	5	Linear	0.0830
13	5	Linear	0.111
9	1	Linear	1.11
9	3	Linear	0.163
9	7	Linear	0.0324
9	9	Linear	0.0294
9	5	Quadratic	0.0499

TABLE II. A summary of the various local models made to recover  $V_A$  from  $V_B$  using data from circuit 1. It is clear that the method is rather robust under changes of the embedding dimension  $d_E$  and time delay  $T$ , in the vicinity of the values selected by global false nearest neighbors and average mutual information, respectively.

rms errors Circuit 1 Recovering voltage $A$ from voltage $B$			
Time delay	Embedding dimension	Order of local maps	rms error $A_r - A$
1	5	Linear	0.0515
3	5	Linear	0.0498
5	5	Linear	0.0527
7	5	Linear	0.0533
9	5	Linear	0.0544
11	5	Linear	0.0477
13	5	Linear	0.0566
9	1	Linear	1.114
9	3	Linear	0.0643
9	7	Linear	0.0473
9	11	Linear	0.0456
9	5	Quadratic	0.0526

rates at the power supply voltage of 15 V, producing a clipping effect.

The first step is to establish the properties of each data set, and in Fig. 20 we present the average mutual information for voltages  $V_A(t)$  and  $V_B(t)$  for circuit 4. Each has its first minimum at  $T=5$  or  $100 \mu s$ . Figure 21 shows that each data set can be embedded in  $d_E=4$ , while the local false nearest-neighbor plot in Fig. 22 clearly establishes that the local dimension of the dynamics is  $d_L=3$ , totally consistent with the circuit equations suggested for the data. Data from  $V_B$  read forward and backward in time are consistent with this. It is important to examine the data both forward and backward in time to confirm the value for  $d_L$ . This seems to be a general property

when the fractal dimension of the attractor is close to an integer, and was seen in the original analysis of the Lorenz model [7]. The reason is clearly numerical, as an attractor with fractal dimension close to an integer may appear to have that integer as its local dimension when real, and thus inaccurate or noisy data, are used to establish  $d_L$ .

Turning to the main subject of this paper, we now examine how well we can do using local linear maps to determine one measured voltage in terms of the other. First we question our ability to model the behavior of voltage  $V_B(n)$  in a state space constructed from time lags of voltage  $A$ . Using a six-dimensional embedding space, we constructed a function

$$V_B(n) = g(\mathbf{y}(n))\mathbf{y}(n) = (V_A(n-3T), V_A(n-2T), V_A(n-T), V_A(n), V_A(n+T), V_A(n+2T))$$

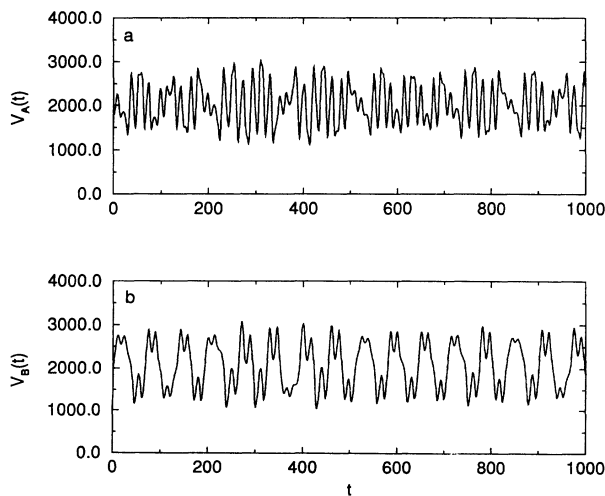


FIG. 19. Two voltages  $V_A$  and  $V_B$  taken at the same time from circuit 4 as described in the text. They were digitized with 12-bit resolution.  $\tau_s = 20 \mu s$ .

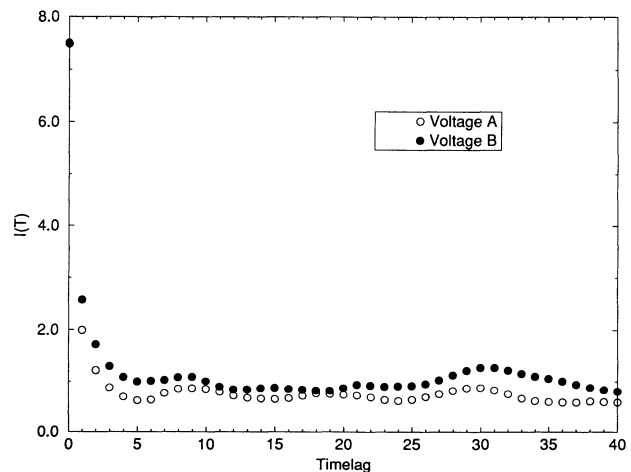


FIG. 20. Average mutual information as a function of time delay for signals  $V_A$  and  $V_B$  from circuit 4. 64 000 data points were used. The units of the time lags are  $20 \mu s$ . Each signal has its first minimum at  $T=5$ .

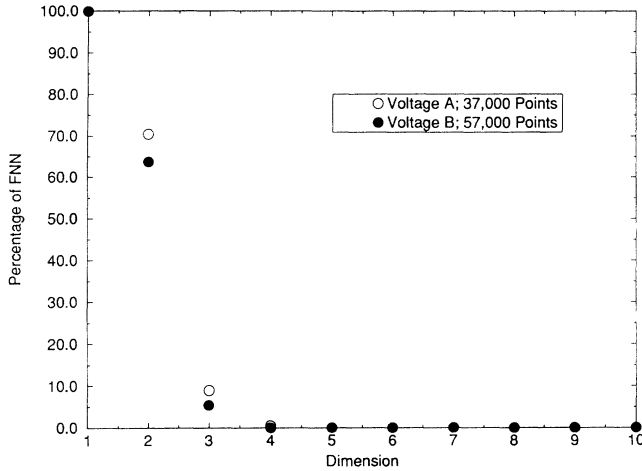


FIG. 21. Percentage of global false nearest neighbors as a function of embedding dimension for the signals  $V_A$  and  $V_B$  from circuit 4,  $T=5$ . Both require dimension  $d_E=5$  to unfold the attractor.

with  $T=5$ . Using 50 000 of the total of 64 000 measured data points as the training set, we predicted the values of  $V_B(n)$  for 1000 points in the recovery set. In Fig. 23(a) we show the result of this operation, using a local linear fit for the local maps. Errors in the recovery of the voltage  $V_B(n)$  of order 0.5% are seen. Turning the problem around—to recover  $V_A(n)$  from measurements of  $V_B(n)$ —in Fig. 23(b) we see a similar substantial success of the local mapping method. The errors rise to about 0.7% in this case but are still quite acceptable. Note that the noise level in the recovered series is a few times greater than in the measured signal (0.1%), which can easily be understood. Indeed, we use time-delayed values of one signal to recover another, and since the noise is amplified due to the largest positive and negative Lyapunov exponents, small errors in the last components  $V_A(t \pm d_E T/2)$  of the embedded vector produce an amplified effect on the predicted value  $V_B(t)$ , and vice versa.

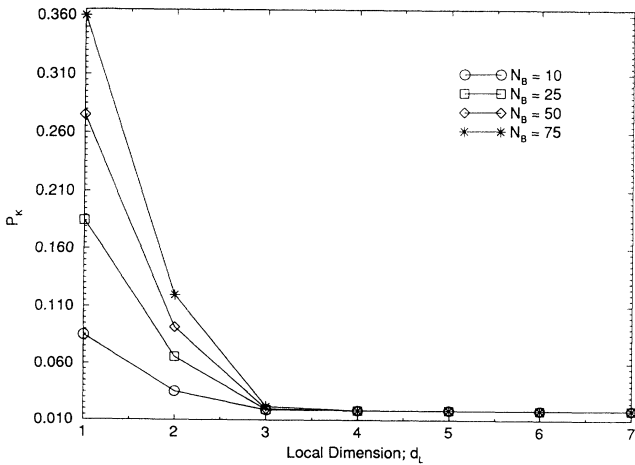


FIG. 22. Local false nearest neighbors for  $V_A$  from circuit 4,  $T=5$ ,  $\beta=0.25$ , and  $N=60\,000$ .  $d_L=3$  is identified here as for  $V_B$ .

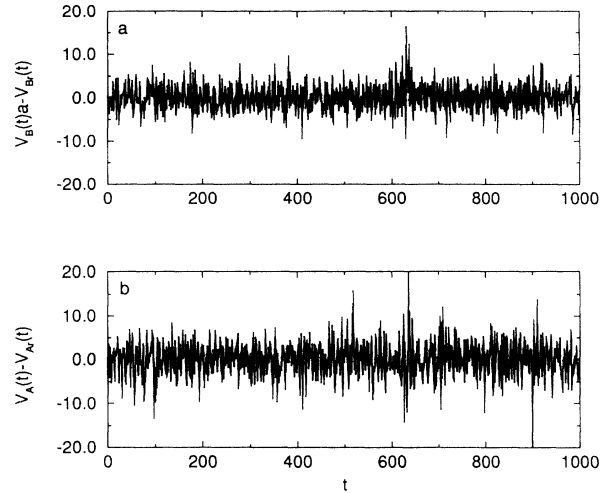


FIG. 23. Data from circuit 4. Recovery errors for modeling  $V_B$  using  $V_A$  (a), and  $V_A$  using  $V_B$  (b). We used 50 000 points in the training set and employed local linear fits. In each case  $d_E=6$ ,  $T=5$ .

### C. Square-loop hysteretic circuit: circuit 5

A different type of hysteretic circuit generated the data in this example [22]. This circuit is very similar to the first hysteretic circuit, except that the hysteresis loop is much sharper. This simple circuit is actually not simple to model, so for illustrative purposes we provide a set of equations that reproduces the approximate behavior of this circuit. Most of the parameters are the same as in the circuit, but the damping factors have been increased to produce chaotic behavior. The values of  $x$  and  $y$  in the equations exceed those in the circuit by factors of 3 or 4. The discrepancy between the actual circuit and the equations that describe it may be caused by capacitor leakage or inductance in the resistors used to provide damping in the integrator portions of the circuit. The equations that have been used for modeling the circuit are

$$\begin{aligned} \frac{dx(t)}{dt} &= -1000[0.67x(t) - 1.77y(t) \\ &\quad - 3.9g(y_n, y_{n-1}) - 0.77x(t)], \\ \frac{dy(t)}{dt} &= -100[x(t) + 0.675y(t)], \end{aligned} \quad (13)$$

where  $g(y_n, y_{n-1}) = -2.0 \operatorname{sgn}(y_n)$  if  $\operatorname{abs}(y_n) \geq 2.0$ ; otherwise it is equal to 2.0 if  $y_{n-1} \geq 0.0$ , and  $-2.0$  if  $y_{n-1} < 0.0$ . The function  $g(y_n, y_{n-1})$  is used to approximate a square hysteresis loop produced by a multivibrator in the circuit. The real circuit is a flow, but the numerical integration routines are maps, so the notation  $y_n$  and  $y_{n-1}$  refers to time steps in the numerical integration routine, and is used to provide a history for the hysteresis loop. The hysteresis loop provides the third degree of freedom necessary for chaos in a flow. The  $x(t)$  variable corresponds to  $V_A(t)$ , while  $y(t)$  corresponds to  $V_B(t)$ . The measured voltages seen in Fig. 24 were digitized at a rate of  $\tau_s = 20 \mu\text{s}$ .

The average mutual information for the two measured

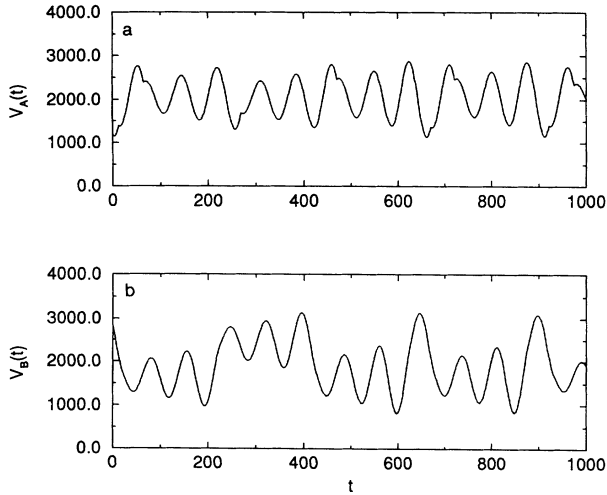


FIG. 24. Two voltages  $V_A$  and  $V_B$  taken at the same time from circuit 5 as described in the text. They were digitized with 12-bit resolution.  $\tau_s = 20 \mu s$ .

voltages is displayed in Fig. 25, where close inspection reveals that the first minimum is achieved for  $T=21$  (or  $420 \mu s$ ) for voltage  $V_A(n)$ , and at  $T=22$  for  $V_B(n)$ . Global false nearest neighbors shown in Fig. 26 demonstrates that each time series can be embedded by time delays in  $d_E=3$ . This is amply confirmed by the local false nearest-neighbor calculations shown in Figs. 27–29. For oversampled data sets like this one, one should be careful implementing a global false neighbor search, because nearest neighbors which are also close in time must be discarded. This requirement is similar in spirit to the observations of Theiler [23] in a related context, and form a standard part of our algorithms.

Working now in  $d_E=5$  with  $\mathbf{y}(n) = [V(n-2T), V(n-T), V(n), V(n+T), V(n+2T)]$  and using  $T=22$ , we make models of one voltage deter-

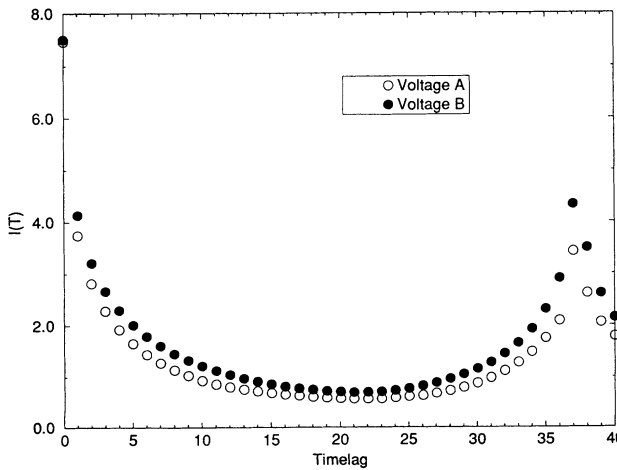


FIG. 25. Average mutual information as a function of time delay for signals  $V_A$  and  $V_B$  from circuit 5. 64 000 data points were used. The units of the time lags are  $50 \mu s$ .  $V_A$  has its first minimum of  $I(T)$  at  $T=21$ , while  $V_B$  has its first minimum at  $T=22$ .

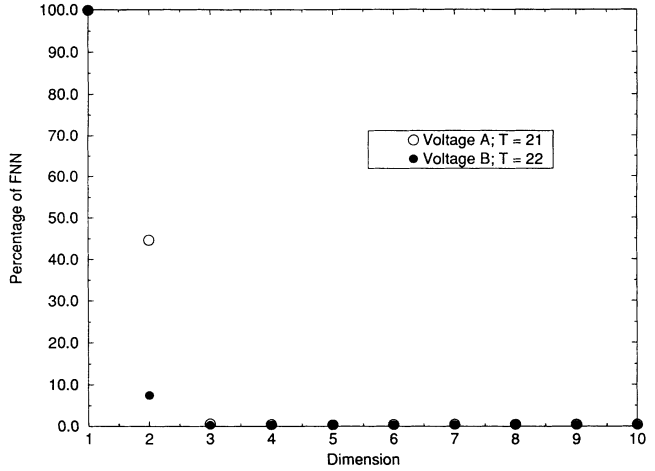


FIG. 26. Percentage of global false nearest neighbors as a function of the embedding dimension for signals  $V_A$  and  $V_B$  from circuit 5. Both require dimension  $d_E=3$  to unfold the attractor.

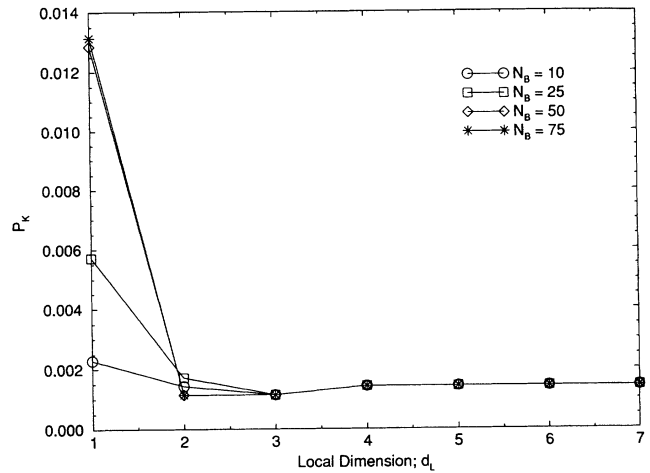


FIG. 27. Local false nearest neighbors for  $V_A$  from circuit 5,  $T=21$ ,  $\beta=0.4$ , and  $N=60\,000$ .  $d_L=3$  is identified here as for  $V_B$ .

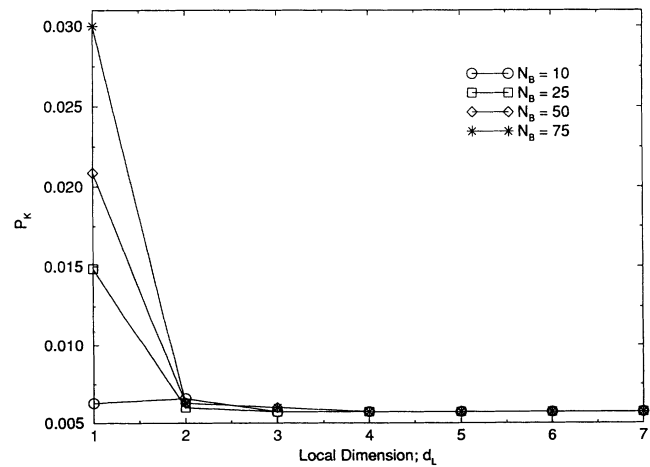


FIG. 28. Local false nearest neighbors for  $V_B$  from circuit 5, the same parameters are used as in Fig. 27.  $d_L=3$  is identified here as for  $V_A$ .

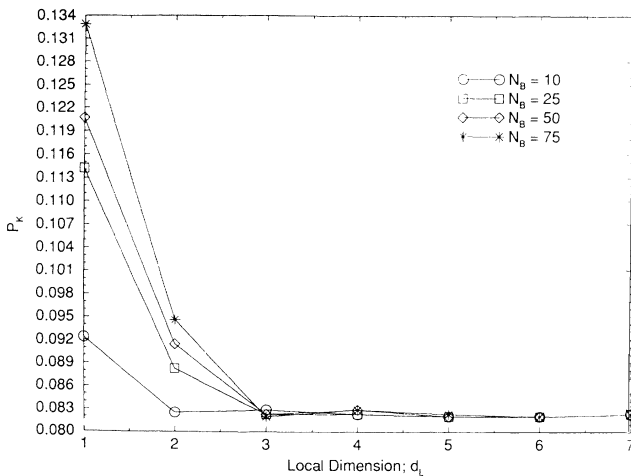


FIG. 29. The same as in Fig. 28, but with the data read backward in time. Here we see the local dimension is  $d_L = 3$ .

mined by time delays of the other. In Fig. 30(a) we have an error for  $V_B(n)$  modeled on voltage  $A$  with local linear fits. As usual the training set is 50 000 points, and the recovery set is taken to be 1000 points. The errors in this representation of  $V_B(n)$  are about 0.5%. In Fig. 30(b) we use local linear models to predict the other way around:  $V_A(n)$  within a time delay space of  $V_B(n)$ . The results are comparable to those just shown. In Fig. 30(c) we increase the order of the local model to quadratic, thus requiring more neighbors and larger neighborhoods for the local models, but we see no substantial increase in our ability to predict. Local linear models seem to work with great accuracy in these examples.

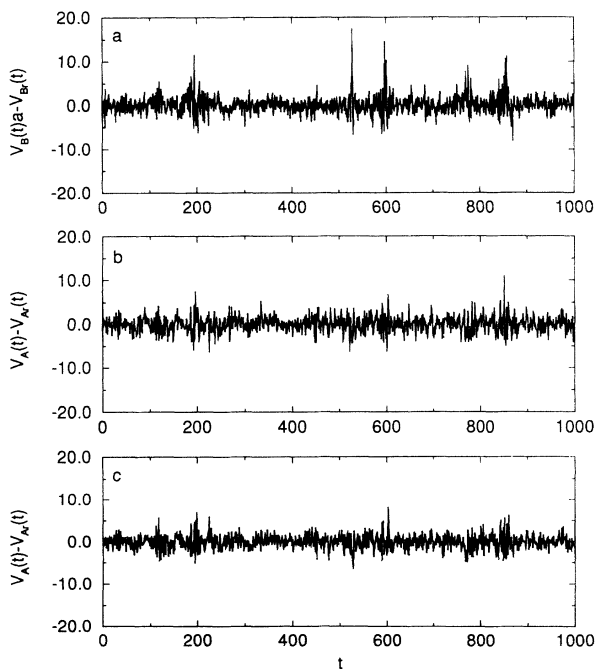


FIG. 30. Data from circuit 5, recovering errors. We used 50 000 points in the training set and employed local linear fits. (a) Recovering  $V_B$  and  $V_A$ . (b) Recovering  $V_A$  using  $V_B$ ,  $d_E = 5$ ,  $T = 22$ , and local linear fit. (c) The same as (b), using a local quadratic fit.

## V. SUMMARY

In this paper we have extended the usual nonlinear modeling of the evolution of a measured scalar variable  $x(n)$  through a function in  $d$ -dimensional space— $\mathbf{y}(n) = (x(n), x(n+T), \dots, x(n+(d-1)T)) \rightarrow F(\mathbf{y}(n)) = \mathbf{y}(n+1)$ —to the representation of a second physical variable observed in the same dynamical system within that reconstructed state space. We have demonstrated this recovery of one physical variable from the measurement of another in the Lorenz model and four examples of nonlinear electronic circuits in our laboratories.

In the process of constructing our examples, we started from the scalar observations and carefully examined the phase space to be reconstructed before carrying out the local mappings required. In each case we considered, computer generated data from the Lorenz model or experimental data from real circuits, we were able without knowledge of the equations of motion to verify the number of dynamical variables requires and to accurately model the evolution of one variable on another.

In a sense all we have accomplished is a demonstration that the embedding theorem is correct: once one has properly captured the state space, then all variables of a nonlinear system coupled to the variable from which the coordinates of the state space are built can be recovered in that space. The importance of the construction extends beyond some kind of indirect verification of the embedding theorem. Indeed, the motivation for this work was to discover how we can work with experimental data on a strange attractor and recover the equivalent of the differential equations we usually depend upon to tell us how one dynamical variable is linked to another. Consider the familiar case of an incompressible fluid, where knowledge of the fluid velocity allows the recovery of the pressure by the solution of the Poisson equation. Our procedure is the direct analog of that construction, but carried out without foreknowledge of the equations of motion, and done directly on the attractor.

We anticipate that the method we have described and demonstrated will be of use in the analysis of chaotic observations when one variable is relatively easy to measure but it is the detailed behavior of another variable which is of direct physical interest. One would proceed in that case to carry out experiments under a situation in which both variables could be measured, and deduce the dynamical rules—of the form  $z(n) = g(\mathbf{y}(n))$  as done here—which allow the evaluation of the second variable in another setting where only the variable  $x(n)$  is accessible.

There are some difficulties and unsolved problems in using this recovery method in practical applications. First of all, as we saw in the example of the Lorenz attractor, there are cases when due to some specific properties of the system at hand (symmetries, etc.), the recovery of one variable from another is impossible in principle. It is not clear at the moment how common these counterexamples are. Nevertheless, all the experimental data series we dealt with allowed the reconstruction in both directions, therefore suggesting that generically recovery is possible. Another interesting question is about the na-

ture of prediction errors we observed. The time series of differences between the original signal and its recovered counterpart reveal some well-separated sharp peaks. The occurrence of these peaks is certainly related to some dynamical properties of the system under investigation, and can possibly be explained by the noise amplification phenomenon [16]. One more problem which deserves systematic study is the influence of the dynamic or measurement noise on one's ability to recover physical variables via time-delay embedding.

Note that since we are making local models of the dynamics in phase space, we need not worry excessively about whether we are using polynomial basis functions, radial basis functions, or other richer sets of functions. Local linear and local quadratic models work exceedingly well for the goals we set. Of course, one can be more ambitious and seek global models for the same purpose of representing one observable scalar variable in terms of another, but this would take us far afield of the framework in this study, and its success would rest heavily on our ability to choose a set of basis functions properly tuned to the global dynamics of the system being observed. Local polynomial dynamics is more accurate since it is less clever: it is universal as long as the vector field governing the evolution is differentiable, namely in essentially every physical setting. Indeed, it works even

when the proposed equations of motion are not formally differentiable, as in several of the examples we presented here.

The unattractive feature of local models is that they are numerical "look up" or interpolation devices which embody the evolution in local regions of phase space. They are not global by construction and may not allow generalization of how differential equations in high-dimensional spaces collapse down to much smaller dimensional attractors. Recognizing their limitations in this way, they certainly provide a powerful tool for the analysis and prediction of actual physical experiments, as we have demonstrated.

#### ACKNOWLEDGMENTS

We thank the members of INLS for numerous discussions on this subject. This work was supported in part by the U.S. Department of Energy, Office of Basic Energy Sciences, Division of Engineering and Geosciences, under Contract No. DE-FG03-90ER14138, and in part by the Army Research Office (Contract No. DAAL03-91-C-052), and by the Office of Naval Research (Contract No. N00014-91-C-0125), under subcontract to the Lockheed/Sanders Corporation.

- 
- [1] J.-P. Eckmann and D. Ruelle, *Rev. Mod. Phys.* **57**, 617 (1985).
  - [2] R. Mañé, in *Dynamical Systems and Turbulence, Warwick 1980*, edited by D. Rand and L. S. Young, *Lecture Notes in Mathematics* Vol. 898 (Springer, Berlin, 1981), p. 230.
  - [3] F. Takens, in Ref. [2], p. 366.
  - [4] T. Sauer, J. A. Yorke, and M. Casdagli, *J. Stat. Phys.* **65**, 579 (1991).
  - [5] H. D. I. Abarbanel, R. Brown, J. J. Sidorowich, and Lev Sh. Tsimring, *Rev. Mod. Phys.* **65**, 1331 (1993).
  - [6] A. M. Fraser and H. L. Swinney, *Phys. Rev.* **33A**, 1134 (1986); A. M. Fraser, *IEEE Trans. Inf. Theory* **35**, 245 (1989); *Physica* **34D**, 391 (1989).
  - [7] H. D. I. Abarbanel and M. B. Kennel, *Phys. Rev. E* **47**, 3057 (1993).
  - [8] M. Casdagli and S. Eubank, *Nonlinear Modeling and Forecasting* (Addison-Wesley, Redwood City, CA, 1992).
  - [9] J. D. Farmer and J. J. Sidorowich, *Phys. Rev. Lett.* **59**, 845 (1987).
  - [10] R. F. Sproull, *Algorithmica* **6**, 579 (1991).
  - [11] M. Giona, F. Lentini, and V. Cimagalli, *Phys. Rev. A* **44**, 3496 (1991).
  - [12] R. Brown, Institute of Nonlinear Science Report, 1993 (unpublished).
  - [13] M. Casdagli and A. S. Weigend, in *Time Series Prediction: Forecasting the Future and Understanding the Past*, edited by A. S. Weigend and N. A. Gershenfeld (Addison-Wesley, Redwood City, CA, 1993).
  - [14] Matthew B. Kennel, R. Brown, and H. D. I. Abarbanel, *Phys. Rev. A* **45**, 3403 (1992).
  - [15] E. N. Lorenz, *J. Atmos. Sci.* **20**, 130 (1963).
  - [16] M. Casdagli, S. Eubank, J. D. Farmer, and J. Gibson, *Physica D* **51**, 52 (1991).
  - [17] T. L. Carroll and L. M. Pecora, *IEEE Trans. Circuits Syst.* **38**, 453 (1991).
  - [18] R. W. Newcomb and S. Sathyan, *IEEE Trans. Circuits Syst. CAS-30*, 54 (1983); R. W. Newcomb and N. El-Leithy, *Circuits Syst. Signal Process.* **5**, 321 (1986).
  - [19] H. D. I. Abarbanel, R. Brown, and M. B. Kennel, *J. Nonlinear Sci.* **1**, 175 (1991).
  - [20] H. D. I. Abarbanel, R. Brown, and M. B. Kennel, *J. Nonlinear Sci.* **2**, 343 (1992).
  - [21] T. L. Carroll and L. M. Pecora, *Int. J. Bifur. Chaos* **2**, 659 (1992).
  - [22] J. Neff and T. L. Carroll, *Sci. Am.* **269** (2), 120 (1993).
  - [23] J. Theiler, *Phys. Rev. A* **34**, 2427 (1986).