Modeling and synchronizing chaotic systems from time-series data

Reggie Brown
Institute for Nonlinear Science, University of California, San Diego, Mail Code R-002, La Jolla, California 92093-0402

Nikolai F. Rulkov
Institute for Nonlinear Science, University of California, San Diego, Mail Code R-002, La Jolla, California 92093-0402

Eugene R. Tracy
Physics Department, College of William and Mary, Williamsburg, Virginia 23185
(Received 10 January 1994)

The problem of obtaining ordinary differential equations (ODE's) from chaotic time-series data is addressed. The vector fields for the ODE's are polynomials constructed from a basis set that is orthonormal on the data. The method for constructing the model is similar to the integration of ODE's using Adams predictor-corrector integration. The method is compared to the usual Euler model and is shown to be accurate for much larger sampling intervals. In addition, the method used to construct the model is capable of determining the optimal polynomial vector field for the given data. Finally, we demonstrate that it is possible to synchronize (in the sense of Fujisaka and Yamada [Prog. Theor. Phys. **69**, 32 (1983)] as well as Pecora and Carroll [Phys. Rev. Lett. **64**, 821 (1990); Phys. Rev. A **44**, 2374 (1991)]) the model to a time series. Synchronization is used as a nontrivial test to determine how close the model vector field is to the true vector field. Implications and possible applications of synchronization are discussed.

PACS number(s): 05.45.+b

I. INTRODUCTION

Experimental time series obtained from measurements of chaotic systems have been the subject of a great deal of research in the past decade. Much of this work has been motivated by the observation that many systems in the world around us appear to behave in a chaotic manner [1]. If one can learn how to detect and model such systems reliably (i.e., to learn the evolution rules), then the door is open to further applications such as short-term prediction and/or control. Thus, one of the goals of current research has been to model the time evolution of the system under investigation. In this paper we will focus on modeling dynamical systems that are continuous in time and are naturally modeled by global autonomous ordinary differential equations (ODE's), $d\mathbf{y}/dt=\mathbf{F}[\mathbf{y}]$. (Modeling discrete time systems was addressed in a previous paper [2].) Our goal is to build vector fields $\mathbf{F}$, which produce orbits whose evolution in time mimics the behavior of the time series. Furthermore, the only *a priori* information that will be used is the time series itself.

For most situations one is unable to measure simultaneously all of the variables necessary to describe completely the experiment under observation. Often one does not even know *a priori* how many variables to measure, much less the identity of these variables. Under these circumstances the phase space of the dynamics must be reconstructed. The difficulty of obtaining experimental data implies that one usually has a scalar time series which we denote by $s(n)=s(n\tau)$, $n=1,2,\ldots,N_D$, where $N_D \gg 1$. Next, some method is used to convert the $s(n)$'s into phase space vectors $\mathbf{y}(n)$, $n=1,\ldots,N$. The theorems of Mañé, Takens, and Sauer, Yorke, and Casdagli lead one to believe that phase space reconstruction can be performed under many circumstances [3–5].

Local methods for modeling ODE's have been developed by a variety of authors [6–10]. These techniques have been shown to be useful for predicting the evolution of initial conditions as well as the important problem of noise reduction and signal separation. The methods model both continuous and discrete time systems with discrete time mappings that are local in the phase space. Methods that are more global (but, they still make explicit reference to the data) have also bee investigated. These methods employ neural nets [11–13], radial basis functions [14,15], and exponential functions [16], as well as polynomials. Like the local methods, these methods model continuous time dynamics with discrete time mappings.

There have also been attempts to model continuous time systems with ODE's [17–21]. Most researchers attempted to model the dynamics by treating each data point as an initial value problem for the ODE. A shift of emphasis has led at least one group to consider modeling the dynamics as a two point boundary value problem [22]. In addition, there has been some interest in modeling nonstationary processes with ODE's [20]. We direct the reader who wishes a more complete discussion of any or all of these issues to the papers cited or to one of the reviews of these topics [23–25]. Our methods differ from the previous work in several ways. The remainder of this introduction is devoted to a discussion of these differences.

With the exception of Baake *et al.* [22], most of the techniques for modeling time-series data with ODE's used a Euler integration approach. This approach models the time evolution via

$$\mathbf{y}(n+1)=\mathbf{y}(n)+\tau\mathbf{F}[\mathbf{y}(n)]$$

or

$$\mathbf{y}(n+1)=\mathbf{y}(n)+\tau\mathbf{F}[\mathbf{y}(n+1)] ,$$

where $\mathbf{y}(n)=\mathbf{y}(n\tau)$. The first of these equations is the more common method and corresponds to explicit Euler integration. The second method corresponds to implicit Euler integration. Given these equations, the modeling problem reduces to finding a suitable vector field $\mathbf{F}$. These methods work well when the time interval between measurements, τ, is small. In many cases, τ must be a few hundred times smaller than the typical oscillations that occur in $\mathbf{y}$. For many experimental situations it is not feasible to use such a small sampling interval. It is therefore worthwhile to seek a method that is capable of accurately modeling $\mathbf{F}$ when τ is much larger (say, by an order of magnitude) than the Euler method would allow. Addressing this problem is one of the purposes of this paper.

Another problem associated with many of the models employed by previous researchers involves the question of truncating the model. Usually, one does not know the functional form of the vector field. Almost surely it cannot be written in closed form in terms of known functions. This is especially true in reconstructed phase spaces. Hence, the best that one can hope for is to model the vector field of the dynamics as a series expansion in some basis:

$$\mathbf{F}[\mathbf{z}]=\lim_{\mathbf{N}_p\to\infty}\sum_{\mathbf{I}=0}^{\mathbf{N}_p}\mathbf{p}^{(\mathbf{I})}\pi^{(\mathbf{I})}(\mathbf{z}) .$$

In this equation, $\pi^{(\mathbf{I})}[\mathbf{z}]$ denotes the set of basis functions and the $\mathbf{p}^{(\mathbf{I})}$'s are parameters whose values must be determined by the modeling procedure. $\mathbf{I}$ is a vector index that is used to identify a particular parameter or basis function. In this paper we will employ a class of basis functions that are constructed to be orthonormal on the data. Besides the obvious esthetic appeal of orthonormal functions, this basis set has been shown to be useful when the data sets are not large and the noise in the data is not small [2,26,27]. For numerical purposes the series must be truncated. The typically unanswered question is: given a basis set $\pi^{(\mathbf{I})}(\mathbf{z})$, at what order (what value of $\mathbf{N}_p$) should one truncate the series? This question is of practical importance and has not been well addressed in the literature [2,25]. We will use a fitting criterion that is capable of selecting the *optimal* order (value of $\mathbf{N}_p$) for the model, given the data.

The final point addressed is the question: How close is the model $\mathbf{F}$ to the true vector field in the phase space? To address this question, we study synchronization of the fitted models to data obtained from the observed system. As we will show, synchronization can be used as an indirect and subtle test of whether or not two dynamical systems are the same.

Two systems are synchronized if they follow the same trajectory. Since the systems we are investigating are chaotic, a first guess would lead one to think that the sensitive dependence on initial conditions would not allow two systems to synchronize. However, it has been clearly demonstrated by Fujisaka and Yamada [28] (FY) as well as Pecora and Carroll [29] (PC) that two "identical" models or two "identical" experiments can be made to synchronize [28,29]. The FY method of synchronization has been explored using several different dynamical systems [30–32]. The same can be said for the PC method of synchronization [33–36].

In this paper we will synchronize a model of the reconstructed dynamics to a time series. To accomplish this, we will use the data as the drive system and the model as the response system. To the best of our knowledge, this has not been previously reported. Note that FY have considered the synchronization of chaotic oscillators with mutual coupling. We modify this method for the case of the drive and response system. We will use both the modified FY and the PC method to accomplish the synchronization. (The papers of FY, originally published in 1983, do not appear to be well known to the research community. The method differs from the PC method; indeed, we will show that the PC method is a special case of the modified FY method.) Applications of this technique for nondestructive testing, performance monitoring, and fault detection will be presented in the conclusion.

An outline of the remainder of this paper is as follows: In Sec. II we present the method used to train our vector fields. The method is based on Adams predictor-corrector integration and is linear in the fitting parameters. It also allows one to determine the optimal order for the model. The modified FY and PC synchronization is also discussed in this section. In Sec. III we present the results of our numerical experiments. We have investigated four different systems. The first two systems are numerical models that have been previously investigated. The third and fourth systems involve experimentally recorded data from an electronic circuit [32,37] and the Belousov-Zhabotinskii reaction [38]. Finally, in Sec. IV we summarize our results, present our conclusions, and make some speculations.

II. MODELING F

In this section we present the formalism used to obtain approximations to the vector field $\mathbf{F}$. The discussion has been broken into three subsections. Each subsection deals with a different aspect of our modeling and testing procedures. Section II A presents the functional form of $\mathbf{F}$. It also introduces the integration model we will use and the method used to fit the coefficients. Section II B presents the minimum description length (MDL) criterion that we use to select the optimal model. In section II C we present a short discussion of the type of driving that we have used to synchronize our models.

We will assume that the data is in the form of a vector time series. Therefore, we have assumed that either all of the relevant phase space variables have been experimen-

tally measured, or the phase space has been reconstructed from a scalar time series. In either case we will denote the phase space vector at time $n\tau$ by $\mathbf{y}(n)$ where each $\mathbf{y}(n)\in\mathbb{R}^d$ and τ is the sampling interval.

A. The vector field and Adams integration

The vector field $\mathbf{F}$ will be written as a series expansion in a polynomial basis. The set of basis functions are constructed to be orthonormal on the attractor mapped out by the phase space vectors $\mathbf{y}(n)$.

It is useful to define a d dimensional index vector $\mathbf{I}$ that can be used to denote a particular basis function. The values taken by the elements of $\mathbf{I}$ are the positive integers and 0. In addition, the elements obey the rule $I_j \geq I_{j+1}$. The first component of $\mathbf{I}$ is I_1 and will always indicate the maximum order of the polynomial.

When modeling an experimental time series, the only piece of *a priori* information that is known is the experimental data. One can define an invariant density on the attractor by

$$\rho(\mathbf{z})=\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}\delta(\mathbf{z}-\mathbf{y}(n)) .$$

This formulation of the density is not unique, but it is the one most often used for modeling [39]. An orthonormal set of polynomial basis functions, $\pi^{(\mathbf{I})}[\mathbf{z}]$, can be constructed on this density via Gram-Schmidt orthogonalization. In brief, one initializes the Gram-Schmidt procedure by defining $\pi^{(\mathbf{0})}[\mathbf{z}]=1$, where $\mathbf{0}=(0,\ldots,0)$. Next, the orthonormality between two basis functions is defined by

$$\begin{aligned}\langle\pi^{(\mathbf{I})}|\pi^{(\mathbf{J})}\rangle &= \int d\mathbf{z}\rho(\mathbf{z})\pi^{(\mathbf{I})}[\mathbf{z}]\pi^{(\mathbf{J})}[\mathbf{z}]\\ &= \lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}\pi^{(\mathbf{I})}[\mathbf{y}(n)]\pi^{(\mathbf{J})}[\mathbf{y}(n)]\\ &= \delta_{\mathbf{IJ}} .\end{aligned}$$

Using the definition of $\rho(\mathbf{z})$, orthonormality, and the initialization $\pi^{(\mathbf{0})}[\mathbf{z}]=1$, one recursively constructs all of the higher order basis functions. A detailed discussion of how to construct this basis from data has been previously presented [2,27]. As such, the details will not be presented in this paper. Instead, we will assume that the necessary calculations have been performed and the polynomial basis set $\pi^{(\mathbf{I})}[\mathbf{z}]$ for large enough values of $\mathbf{I}$ is available.

As stated in Sec. I, we will model the dynamics by ODE's of the form

$$\begin{aligned}\frac{d\mathbf{y}}{dt} &= \mathbf{F}[\mathbf{y}]\\ &= \sum_{\mathbf{I}=\mathbf{0}}^{\mathbf{N}_p}\mathbf{p}^{(\mathbf{I})}\pi^{(\mathbf{I})}[\mathbf{y}] ,\end{aligned} \tag{1}$$

where $\mathbf{N}_p=(N_p,\ldots,N_p)$. The $\mathbf{p}^{(\mathbf{I})}$'s are parameters whose values are to be determined so that, when integrated, Eq. (1) carries $\mathbf{y}(n)$ into $\mathbf{y}(n+1)$ as accurately as possible. [Hence, we formulate the modeling problem as an initial value problem for Eq. (1).] To accomplish this task, we imagine a situation where the $\mathbf{p}^{(\mathbf{I})}$'s are known and one seeks to integrate Eq. (1) to determine the $\mathbf{y}$'s. When $\mathbf{F}$ is known (the $\mathbf{p}^{(\mathbf{I})}$'s are known), one possible method of integrating Eq. (1) is the Adams predictor-correct method. Under this method, one integrates Eq. (1) from an initial point $\mathbf{y}(n)$ to the next point $\mathbf{y}(n+1)$ in one time step of size τ via

$$\mathbf{y}(n+1)=\mathbf{y}(n)+\tau\sum_{j=0}^{M}a_j^{(M)}\mathbf{F}[\mathbf{y}(n+1-j)] . \tag{2}$$

The $a_j^{(M)}$'s are the implicit Adams predictor-corrector coefficients. M indicates the order of the corrector portion of the integration. The numerical values of the $a_j^{(M)}$'s are known for all values of j and order M [40]. The integration method is implicit since it involves evaluating $\mathbf{F}$ at the point $\mathbf{y}(n+1)$ and reduces to implicit Euler integration when $M=0$. A discussion of Adams predictor-corrector integration can be found in most numerical methods texts [40].

A reinterpretation of Eq. (2) reveals that if the $\mathbf{y}(n)$'s are known, then this equation can be employed as a means of finding an unknown function $\mathbf{F}$ (equivalently, the $\mathbf{p}^{(\mathbf{I})}$'s). Equation (1) indicates that $\mathbf{F}$ is linear in its unknown coefficients $\mathbf{p}^{(\mathbf{I})}$. A standard least squares minimization approach to finding these coefficients [which can be derived from the maximum likelihood (ML) principle of Gaussian noise [40]] involves selecting the $\mathbf{p}^{(\mathbf{I})}$'s so as to minimize

$$\chi^2_{\mathrm{ML}}=\frac{1}{2N\sigma^2}\sum_{n=1}^{N}\left|\mathbf{y}(n+1)-\mathbf{y}(n)-\tau\sum_{j=0}^{M}a_j^{(M)}\sum_{\mathbf{I}=\mathbf{0}}^{\mathbf{N}_p}\mathbf{p}^{(\mathbf{I})}\pi^{(\mathbf{I})}[\mathbf{y}(n+1-j)]\right|^2 . \tag{3}$$

In order to implement numerically the ML criterion, we should permit the values of the $a_j^{(M)}$'s to vary. Therefore, in principle one should minimize χ^2_{ML} over the $a_j^{(M)}$'s as well as the $\mathbf{p}^{(\mathbf{I})}$'s. It is important for our modeling procedure that χ^2_{ML} be quadratic with respect to the fitting parameters. Therefore, we will assume that the optimal values of the $a_j^{(M)}$'s are the ones originally dictated by the Adams integration method, and we will fix the $a_j^{(M)}$'s at these prescribed values. This still leaves the order of the Adams method, M, to be determined, as well as the order of the polynomial, N_p, and the coefficients $\mathbf{p}^{(\mathbf{I})}$.

The procedure we will use to minimize χ^2_{ML} is not the usual least squares procedure. (However, it can be shown to be equivalent to the usual least squares procedure [41].) Minimizing χ^2_{ML} involves choosing numerical values for $\mathbf{p}^{(\mathbf{I})}$ so as to obtain the smallest possible value for χ^2_{ML}. Clearly, for these values of $\mathbf{p}^{(\mathbf{I})}$, Eq. (2) will be nearly satisfied. We infer that choosing numerical values

for $\mathbf{p}^{(\mathbf{I})}$ so as to satisfy Eq. (2) will have the effect of minimizing χ^2_{ML}. In order to obtain the $\mathbf{p}^{(\mathbf{I})}$'s we insert Eq. (1) into Eq. (2) and project the result onto $\pi^{(\mathbf{J})}$. It is fairly straightforward to show that this results in

$$G_\beta[1;\mathbf{J}]=\sum_{\mathbf{K}=0}^{\mathbf{N}_p} p_\beta^{(\mathbf{K})}\left[\sum_{k=0}^{M} a_k^{(M)}\Delta[1,k;\mathbf{J},\mathbf{K}]\right], \tag{4}$$

where we have made the following definitions:

$$G_\beta[j,\mathbf{J}]=\lim_{N\to\infty}\frac{1}{\tau N}\sum_{n=1}^{N}[\mathbf{y}(n+1)-\mathbf{y}(n)]_\beta \times\pi^{(\mathbf{J})}[\mathbf{y}(n+1-j)] \tag{5}$$

and

$$\Delta[j,k;\mathbf{J},\mathbf{K}]=\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}\pi^{(\mathbf{J})}[\mathbf{y}(n+1-j)] \times\pi^{(\mathbf{K})}[\mathbf{y}(n+1-k)] . \tag{6}$$

In these equations the subscript β indicates the component of a vector. For example, $y_\beta(n)$ is the β component of the vector $\mathbf{y}(n)$. In addition, from the definition of orthonormality it is obvious that $\Delta[j,j;\mathbf{J},\mathbf{K}]=\delta_{\mathbf{JK}}$.

The term in square brackets in Eq. (4) can be defined as

$$X[\mathbf{J};\mathbf{K}]=\sum_{k=0}^{M} a_k^{(M)}\Delta[1,k;\mathbf{J},\mathbf{K}] . \tag{7}$$

If this definition is inserted into Eq. (4), we obtain the desired form of the minimization problem:

$$G_\beta[1;\mathbf{J}]=\sum_{\mathbf{K}=0}^{\mathbf{N}_p} X[\mathbf{J};\mathbf{K}]p_\beta^{(\mathbf{K})} . \tag{8}$$

The form of Eq. (8) suggests a matrix inversion problem. $X[\mathbf{J};\mathbf{K}]$ can be interpreted as the $(\mathbf{J},\mathbf{K})$ element of a matrix. The numerical values of the elements of this matrix are determined from the known data $\mathbf{y}(n)$, the known polynomials $\pi^{(\mathbf{I})}$, and the known values of $a_k^{(M)}$ [see Eqs. (6) and (7)]. $G_\beta[1;\mathbf{J}]$ can be interpreted as the $(\mathbf{J},\beta)$ element of a matrix whose numerical values can be determined [see Eq. (5)]. Finally, $p_\beta^{(\mathbf{K})}$ can be interpreted as the $(\mathbf{K},\beta)$ element of an unknown matrix. There are many methods for solving the problem $\mathbf{V}=\mathbf{XS}$ where $\mathbf{V}$ and $\mathbf{X}$ are known matrices and $\mathbf{S}$ is an unknown matrix. For the numerical experiments we present in Sec. III, we have used the single value decomposition.

B. The minimum description length criterion

Equation (3) is quadratic in the fitting parameters $\mathbf{p}^{(\mathbf{I})}$, so a unique solution exists which results in a unique model for the vector field $\mathbf{F}$. Clearly, by increasing the order of the polynomial (and/or the Adams corrector M), one can obtain as accurate a model for $\mathbf{F}$ as desired. In fact, if $M=0$ and the number of $\mathbf{p}^{(\mathbf{I})}$'s is equal to N, then it is easy to get $\chi^2_{\mathrm{ML}}=0$ [40]. For our purposes, N is a few thousand. In this case, having a polynomial with a few thousand coefficients (corresponding to N_p of a few hundred) is not useful and is not an interesting limit. Therefore, one desires an objective criterion for truncating the series given by Eq. (1). Whatever criterion is used, it should also be able to determine M, the order of the Adams predictor corrector. In a series of papers Rissanen has developed the minimum description length principle for truncating a model [42]. We present a brief discussion of this principle and direct the reader to the original papers by Rissanen for a complete discussion of the MDL principle [42].

The MDL principle is an extension of the maximum likelihood principle. Let Y denote the entire data set $\mathbf{y}(n)$, $n=1,\ldots,N$, and let Θ denote all of the parameters in the model. To be specific, $\Theta=(a_0^{(M)},\ldots,a_M^{(M)},\mathbf{p}^{(0)},\ldots,\mathbf{p}^{(\mathbf{N}_p)})$. We will use M_P to indicate the total number of nonzero components in the vector Θ. The ML principle selects a model by minimizing

$$\chi^2_{\mathrm{ML}}=-\log_2[P(Y|\Theta)] ,$$

where $P(Y|\Theta)$ is the conditional probability of obtaining Y *given* Θ [40]. The size of the model is determined, in part, by the number of components in the vector Θ. The ML principle assumes that the number of components of Θ is given. Therefore, it cannot be used to determine the optimal size of the model.

In contrast, the MDL principle selects a model by minimizing

$$\chi^2_{\mathrm{MDL}}=-\log_2[P(Y,\Theta)] ,$$

where $P(Y,\Theta)=P(Y|\Theta)P(\Theta)$ is the joint probability of obtaining Y *and* Θ, and $P(\Theta)$ is the probability of obtaining Θ. By using a joint probability, the MDL principle is capable of determining the optimal size of the model by determining the probability of Θ. Another interpretation of the MDL principle says that the optimal model [within the class given by Eqs. (1) and (2)] is the one that produces the smallest encoding for the data *and* the vector field.

In order to associate a rigorous definition of size to a model, one assumes that the parameters have finite precision. Hence, one has coarse grained the parameter space. Two parameter values are considered identical if they fall within the same box in the grid. In this way the entire parameter space can be covered with boxes and one can associate a unique integer to each box. A particular model can be associated with a particular value of Θ. The size of the model is the value of the integer that identifies the box in which Θ resides. The probability of the model, $P(\Theta)$, is the probability of the integer that is associated with the box that contains Θ.

An explicit formula for the function to be minimized can be obtained by optimizing the precision used to express the parameters. To see this, let Θ^* be the center of the box that holds Θ. Then, $\chi^2_{\mathrm{ML}}(\Theta)$ can be approximated by a Taylor series expansion which we write as

$$\chi^2_{\mathrm{ML}}(\Theta)\simeq\chi^2_{\mathrm{ML}}(\Theta^*)+\tfrac{1}{2}\|\Theta-\Theta^*\|^2 . \tag{9}$$

The norm in this equation is given by

$$\|\Theta\|^2=\sum_{j=1}^{M_P}\sum_{k=1}^{M_P}\Theta_j M_{j,k}\Theta_k ,$$

where Θ_j is the jth component Θ and $M_{j,k}$ is a metric tensor given by

$$\mathbf{M}=\begin{bmatrix} \partial^2\chi^2/\partial a_j^{(M)}\partial a_k^{(M)} & \partial^2\chi^2/\partial a_j^{(M)}\partial p_\alpha^{(\mathrm{I})} \\ \partial^2\chi^2/\partial p_\alpha^{(\mathrm{I})}\partial a_j^{(M)} & \partial^2\chi^2/\partial p_\alpha^{(\mathrm{I})}\partial p_\beta^{(\mathrm{J})} \end{bmatrix} .$$

(The metric tensor is sometimes called the first fundamental form [43]. For the standard Euclidean norm the metric tensor is just the identity matrix. Other, more complicated, forms of the metric tensor often appear in research involving general relativity.)

The second term in Eq. (9) can be used to define an ellipsoid in the parameter space. Let L_1 be some length; then $L_1=\|\Theta-\Theta^*\|^2$ defines an ellipsoid in the M_P dimensional parameter space. The integer associated with Θ is given by the ratio of the volume of the ellipsoid, $\|\Theta\|^2$, to the volume of the box that defines the grid. By optimizing the grid size, one obtains the following expression:

$$\chi^2_{\mathrm{MDL}}=\frac{1}{2\sigma^2 N}\sum_{n=1}^{N}\left|\mathbf{y}(n+1)-\mathbf{y}(n)-\tau\sum_{j=0}^{M}a_j^{(M)}\sum_{\mathrm{I}=0}^{\mathrm{N}_p}\mathbf{p}^{(\mathrm{I})}\pi^{(\mathrm{I})}[\mathbf{y}(n+1-j)]\right|^2 + \frac{M_P}{2N}[\ln(2\pi eN/M_P)+\ln(\|\Theta\|^2)] . \tag{10}$$

Minimizing Eq. (10) is the MDL criterion for fitting a model [42]. The first term in this equation is the usual χ^2_{ML} term. The second two terms in Eq. (10) arise from $P(\Theta)$ and are associated with the size of the model. Typically, as one adds more parameters, the first term decreases, but the sum of the second two terms increases. This implies a minima somewhere in the middle. One chooses the minima of χ^2_{MDL} as the optimal model. In practice it is usually the case that the χ^2_{ML} term dominates Eq. (10). Under these conditions, the procedure for minimizing Eq. (10) involves minimizing the first term in Eq. (10) for a particular pair of N_p and M values. Having thus found the $\mathbf{p}^{(\mathrm{I})}$'s, one calculates the second two terms and adds them to the first term. The minimum of Eq. (10) can be found by repeating this procedure for all values of N_p and M.

In order to evaluate Eq. (10), the square of the norm of the parameter vector, $\|\Theta\|^2$, must be calculated. It is tedious but straightforward to show that the norm of the parameter vector can be written in terms of $\Delta(j,k;\mathbf{J},\mathbf{K})$ and $\mathbf{G}(j,\mathbf{J})$ as

$$\|\Theta\|^2=\frac{6\tau^2}{\sigma^2}\sum_{j=0}^{M}\sum_{k=0}^{M}\sum_{\mathrm{J}=0}^{\mathrm{N}_p}\sum_{\mathrm{K}=0}^{\mathrm{N}_p}a_j^{(M)}a_k^{(M)}\mathbf{p}^{(\mathrm{J})}\cdot\mathbf{p}^{(\mathrm{K})}\Delta[j,k;\mathbf{J},\mathbf{K}] - \frac{2\tau^2}{\sigma^2}\sum_{j=0}^{M}\sum_{\mathrm{J}=0}^{\mathrm{N}_p}a_j^{(M)}\mathbf{p}^{(\mathrm{J})}\cdot\mathbf{G}[j;\mathbf{J}] . \tag{11}$$

At this point all of the necessary tools for generating an optimal model from an experimental data set have been presented. In order to carry out the modeling, one must select a range of values for M and N_p that will be investigated. In principle, one must choose the maximum and minimum values of these parameters so that the optimal solution is inside this range. In general, one does not have this information *a priori*. For our numerical experiments we have used $M\in[0,9]$ and $N_p\in[0,10]$. We find that for all the test cases we have investigated, the optimum was contained within these ranges. For real data one is attempting to determine a noisy value of $\mathbf{y}(n+1)$ from a noisy value of $\mathbf{y}(n)$. We suspect that one's ability to make this type of prediction will saturate at a relatively low order polynomial. This lack of increase in predictive ability as the order of the polynomial increases is one of the motivations behind the MDL principle and the penalty terms in Eq. (10). For these reasons we believe that the ranges we have used above will be appropriate for most situations.

The exact procedure we will implement to produce the models is as follows.

(1) Set the initial values of M and $\mathbf{N}_p$ to zero.

(2) Solve the matrix problem represented by Eq. (8) and evaluate χ^2_{ML}. (This is equivalent to performing a maximum likelihood calculation of minimizing χ^2_{ML}.)

(3) Calculate the norm of the parameter vector by evaluating Eq. (11).

(4) Calculate χ^2_{MDL} by evaluating Eq. (10).

(5) If M is less than its maximum value, then increase M by 1. If M is at its maximum value, then increase N_p by 1 and reset M to zero.

(6) Return to step (2).

After searching all of the values $M\in[0,9]$ and $\mathbf{N}_p\in[0,10]$, choose the model with the smallest value of χ^2_{MDL}.

Before leaving this section, we digress with a short discussion of the benefits associated with the basis functions $\pi^{(\mathrm{I})}$. The basis set we have used for the expansion of the vector field $\mathbf{F}$ is not the only one that could be used. Even with the class of basic functions called polynomials, we could have used many different basis sets. For example, in $d=2$ dimensions one is tempted to use the standard basis set given by $1,x,y,x^2,xy,y^2$, etc. For arbitrary dimension we could write $\mathbf{F}$ in terms of the standard basis as

$$\mathbf{F}[\mathbf{z}]=\sum_{\mathrm{I}=0}^{\mathrm{N}_p}\mathbf{B}^{(\mathrm{I})}\prod_{\beta=1}^{d}z_\beta^{I_\beta-I_{\beta+1}} . \tag{12}$$

The standard basis set is not orthonormal on the attractor given by $\rho[\mathbf{z}]$. It is known that numerical accuracy can be increased by performing the calculations with a basis that is orthonormal on the domain of interest [44]. For many situations one may also observe that certain parameters are "small." When this occurs, it is often desirable to set these parameter values to be exactly zero.

By doing so, one expects a slightly poorer fit to the dynamics, but this may be well worth the price if the model becomes sufficiently simple. We would argue that it is easier to determine the "small" parameters when the basis set is orthonormal. Examples of this benefit will be presented in Sec. III.

C. Synchronization

We now discuss synchronization of two chaotic systems. Synchronization involves using an output from one dynamical system (the driving system) as a driving input to another dynamical system (the response system). If **G** is the vector field of the driving system and **F** is the vector field of the response system, then the modified FY method of synchronization is given by

$$\begin{aligned} \frac{d\mathbf{y}}{dt} &= \mathbf{G}(\mathbf{y}) , \\ \frac{d\mathbf{z}}{dt} &= \mathbf{F}(\mathbf{z}) - \mathbf{E}\cdot(\mathbf{z}-\mathbf{y}) , \end{aligned} \tag{13}$$

where the matrix **E** denotes the coupling between the two systems. We have assumed that $\mathbf{y}\in\mathbb{R}^d$ and $\mathbf{z}\in\mathbb{R}^d$ denote locations in the phase spaces of drive and response systems. For our numerical experiments the fitted model **F** plays the role of the response system. The driving system **G** is the experimental system, which is assumed to be unknown.

An important special case is $\mathbf{F}=\mathbf{G}$. In this case the full system given by Eqs. (13) has an integral manifold

$$\mathbf{z}=\mathbf{y} . \tag{14}$$

Trajectories located on this manifold correspond to identical (synchronous) oscillations in drive and response systems. Synchronization of the response system $\mathbf{z}(t)$ with chaotic driving $\mathbf{y}(t)$ can be achieved if the coupling **E** stabilizes a chaotic limit set located on the manifold. Stability of synchronization is determined by the conditional Lyapunov exponents calculated from the linearized equations:

$$\frac{d\delta\mathbf{z}}{dt} = \{\mathbf{DF}[\mathbf{y}(t)]-\mathbf{E}\}\cdot\delta\mathbf{z} ,$$

where $\delta\mathbf{z}=\mathbf{z}-\mathbf{y}$ is the difference between states of the response and drive systems. The response system will reproduce the exact trajectory of the driving system only if all of the Lyapunov exponents of this equation are negative. We note that for this type of driving it is *always* possible to find a coupling **E** that will result in synchronization [28].

For some systems synchronization can be obtained with a coupling matrix **E**, which contains only one nonzero element, $E_{\alpha\alpha}$ [45]. When this is true, the driving is given by the scalar y_α. Assume that this type of driving results in synchronization for a particular vector field **F**. Using a $d=3$ dimensional example, we now show that this method of driving reduces to the one used by PC in the $E_{\alpha\alpha}\to\infty$ limit. The equations of the driving and response systems can be written in the form

$$\begin{aligned} \frac{dy_1}{dt} &= F_1[y_1,y_2,y_3] , \\ \frac{dy_2}{dt} &= F_2[y_1,y_2,y_3] , \\ \frac{dy_3}{dt} &= F_3[y_1,y_2,y_3] \end{aligned} \tag{15}$$

and

$$\begin{aligned} \frac{dz_1}{dt} &= F_1[z_1,z_2,z_3] , \\ \frac{dz_2}{dt} &= F_2[z_1,z_2,z_3] , \\ \frac{dz_3}{dt} &= F_3[z_1,z_2,z_3] - \epsilon(z_3-y_3) , \end{aligned} \tag{16}$$

where $\epsilon=E_{33}$ is the coupling parameter. All other elements of the matrix **E** are zero.

In the limit $\epsilon\to\infty$ the differential equation for z_3 can be reduced to the equation $z_3=y_3$, which describes a stable five-dimensional manifold of slow motion in the six-dimensional phase space $(\mathbf{y},\mathbf{z})$. (We have assumed that the systems synchronize, so for any initial condition the response system approaches the manifold after a short transient.) The motion of the response system along the manifold is given by

$$\begin{aligned} \frac{dz_1}{dt} &= F_1[z_1,z_2,y_3] , \\ \frac{dz_2}{dt} &= F_2[z_1,z_2,y_3] , \end{aligned}$$

which are the equations one would obtain from the PC method for synchronizing a response system by driving with y_3. Thus, PC synchronization is a special case of modified FY synchronization. FY and PC observed that under suitable conditions the variables $[z_1,z_2,z_3=y_3]$ follow the same orbit as the variables $[y_1,y_2,y_3]$ even when the dynamics of the driving system is chaotic and the initial conditions for the driving and response systems are different.

When $\mathbf{G}=\mathbf{F}$, synchronous oscillations in the drive and response systems correspond to trajectories located on the integral manifold given by Eq. (14). For our numerical experiments we will always be faced with $\mathbf{G}\neq\mathbf{F}$. The lack of equality arises because the vector field **F** is only an approximation to **G**, the true vector field that governs the dynamics. When $\mathbf{F}\neq\mathbf{G}$, synchronous oscillations correspond to an attractor which is no longer on the integral manifold given by Eqs. (14) (unless we have the odd case where the vector fields **F** and **G** are identical in the region covering the attractor, but differ in other parts of phase space).

If **F** and **G** differ slightly, then the dynamics of $\delta\mathbf{z}=\mathbf{z}-\mathbf{y}$ can be approximately described by the linearized equations

$$\frac{d\delta\mathbf{z}}{dt} = [\mathbf{DF}(\mathbf{y})-\mathbf{E}]\cdot\delta\mathbf{z} - [\mathbf{G}(\mathbf{y})-\mathbf{F}(\mathbf{y})] ,$$

where $\mathbf{y}(t)$ is the chaotic trajectory generated by the driv-

ing system, **G**. Using proper coupling, one can achieve near synchronization where the systems oscillate with $0<|\delta \mathbf{z}|\ll 1$. If the value of the coupling parameter **E** is small (about the same order as is necessary for the synchronization with $\mathbf{F}=\mathbf{G}$), then near synchronization can be found only if $|\mathbf{G}(\mathbf{y})-\mathbf{F}(\mathbf{y})|\ll 1$ when **y** is near the attractor. This property enable us to use synchronization as a nontrivial test for our models in the vicinity of the attractor. (If we can synchronize **F** to the time series, then we conclude $\mathbf{F}\simeq\mathbf{G}$.) As a final note, we point out that the form of driving given by Eqs. (13) permits us to check *each* component of the fitted vector field **F**. This is in contrast to PC type driving where one of the components of the vector field is always removed (see the $\epsilon\rightarrow\infty$ example above).

III. NUMERICAL EXPERIMENTS

This section of our paper reports the results of numerical experiments using the modeling techniques presented in Sec. II. The experiments addressed each of the three purposes specifically mentioned in the Introduction. Specifically, we will demonstrate the advantages of our method when modeling data whose sampling interval τ is not small. We will demonstrate that the MDL criterion [i.e., minimizing Eq. (10)] allows one to choose the optimal model [within the polynomial class of models given by Eqs. (1) and (2)]. Finally, we will demonstrate that it is possible to synchronize our models to a time series. Synchronization between a model and a time series is a very important phenomenon. It indicates that near the attractor the model is close to the true dynamical system that generated the time series.

The numerical experiments will employ four different dynamical systems. The first system we will investigate is the well-known Lorenz system of three coupled ODE's [46]:

$$\frac{dx}{dt}=s(y-x)\ ,$$

$$\frac{dy}{dt}=rx-y-xz\ , \tag{17}$$

$$\frac{dz}{dt}=-bz+xy\ ,$$

where $s=16$, $b=4$, and $r=45.92$. This system has become one of the canonical examples in dynamical systems theory. It was originally derived as a very crude model for convection in the atmosphere [46]. Since an exact representation in terms of polynomials is possible [when $\mathbf{y}(n)=(x(n\tau),y(n\tau),z(n\tau))$], Eqs. (17) represent the simplest and most direct test of our modeling methods.

The second system we will investigate is a theoretical model for chemical passivation of metal in an aqueous solution. Passivation (the loss of chemical reactivity as the surface of a metal becomes covered) is an important phenomenon that has been observed to exhibit a variety of nonlinear behavior. Period doubling [47] and multistability [48], as well as quasiperiodicity [49,50] and homoclinic Shilnikov chaos [51], have all been experimentally observed. The model we will study is given by the following coupled ODE's [52]:

$$\begin{aligned}\frac{dY}{dt}&=p(1-\theta_{\mathrm{O}}-\theta_{\mathrm{OH}})-qY\ ,\\ \frac{d\theta_{\mathrm{OH}}}{dt}&=Y(1-\theta_{\mathrm{O}}-\theta_{\mathrm{OH}})f(\theta_{\mathrm{OH}})\\ &\quad+2s\theta_{\mathrm{O}}(1-\theta_{\mathrm{O}}-\theta_{\mathrm{OH}})-[r+g(\theta_{\mathrm{OH}})]\theta_{\mathrm{OH}}\ ,\\ \frac{d\theta_{\mathrm{O}}}{dt}&=r\theta_{\mathrm{OH}}-s\theta_{O}(1-\theta_{\mathrm{O}}-\theta_{\mathrm{OH}})\ ,\end{aligned} \tag{18}$$

where $p=2\times10^{-4}$, $q=1\times10^{-3}$, $s=9.7\times10^{-5}$, and $r=2\times10^{-5}$. McCoy *et al.* [52] examined $f(\theta_{\mathrm{OH}})=1$ and $g(\theta_{\mathrm{OH}})=\exp[-b\Theta_{\mathrm{OH}}]$ with $b=5$, and we will do the same.

Even when

$$\mathbf{y}(n)=(Y(n\tau)\ ,\theta_{\mathrm{OH}}(n\tau)\ ,\theta_{\mathrm{O}}(n\tau))\ ,$$

this model is not a polynomial so it is impossible for our method to reproduce exactly Eqs. (18). As we argued in the Introduction, when one is forced to build models from experimentally measured time series, the vector field will not, in general, be a finite series of known functions. Hence, one will always be forced to use approximations when the models are being constructed from experimental data. For example, the basic form of the equations for many chemical reactions can often be determined from first principles. Indeed, this type of consideration lead to Eqs. (18) [52,53]. What is unknown is the functional form of f and g. Given a suitable experimental time series, our method could be used to determine f and g.

The construction of polynomials that are orthonormal on the attractor associated with Eqs. (18) is not an easy numerical task. In terms of the variables Y, θ_{O}, and θ_{OH}, the attractor makes very small excursions about the mean values $\bar{Y}$, θ_{OH}, and $\bar{\theta}_{\mathrm{O}}$. These variations are $\sim10^{-3}$. For variations that are this small, it is numerically difficult to obtain a basis set (the $\pi^{(1)}$'s) above $N_p\simeq2$. In order to overcome this limitation, we have performed a linear change of variables by removing the mean values and rescaling the fluctuations via $\alpha U=Y-\bar{Y}$, $\beta V=\theta_{\mathrm{OH}}-\bar{\theta}_{\mathrm{OH}}$, $\gamma W=\theta_{\mathrm{O}}-\bar{\theta}_{\mathrm{O}}$, and $Tx=t$. The mean values were determined from a numerically generated data set of 60 000 points and are $\bar{Y}=0.112\,85\ldots$, $\bar{\theta}_{\mathrm{OH}}=0.319\,13\ldots$, and $\bar{\theta}_{\mathrm{O}}=0.116\,61\ldots$. The rescaling parameters were chosen to be $\alpha=0.002$, $\beta=0.1$, $\gamma=0.0004$, and $T=1000$. The values of α, β, and γ are close to the standard deviations of the variables as determined from the data set.

In terms of the new variables, the passivation model is

$$\frac{dU}{dx}=100[A-5\bar{Y}]-U-10[V+0.004W]\ ,$$
$$\frac{dV}{dx}=10^4\left[A\bar{Y}-r\bar{\theta}_{\rm OH}\left[1-\frac{2\Gamma A}{\bar{\theta}_{\rm OH}}\right]\right]+20AU-10^3[\bar{Y}+r(1+2\Gamma)]V-4\left[\bar{Y}+2r\Gamma\left[1-\frac{A}{\bar{\theta}_{\rm O}}\right]\right]W$$
$$+2[U+0.4sW][V+0.004W]-10^3[V+10\bar{\theta}_{\rm OH}]E\exp(-V/2)\ , \tag{19}$$
$$\frac{dW}{dx}=50\bar{\theta}_{\rm OH}\left[1-\frac{\Gamma A}{\bar{\theta}_{\rm OH}}\right]+5[1+\Gamma]V+0.02\Gamma\left[1-\frac{A}{\bar{\theta}_{\rm O}}\right]W+0.002\left[\frac{s}{r}\right][V+0.004W]W\ ,$$

where $\Gamma=s\bar{\theta}_{\rm O}/r\sim\frac{1}{2}$, $E=\exp(-5\bar{\theta}_{\rm OH})$, and $A=1-\bar{Y}-\bar{\theta}_{\rm OH}-\bar{\theta}_{\rm O}$. Our numerical experiments will use data from Eqs. (18) that have been rescaled in the manner discussed $[\mathbf{y}(n)=(U(n\tau),V(n\tau),W(n\tau))]$. Therefore, we will actually be modeling Eqs. (19). (If necessary, one could always invert the change of variables to return to the original Y, $\theta_{\rm OH}$, and $\theta_{\rm O}$ variables.) It is worth noticing that the equation of motion for U is linear and the equation of motion for W has only weak nonlinearities of quadratic order. This fact will be important later in this section.

The final two systems we will investigate must be reconstructed from *experimentally measured* time series. The first system is an electronic circuit whose block diagram is shown in Fig. 1. This circuit has been previously investigated by one of the authors (Rulkov) [32,37] and has an attractor that is similar to the one associated with the so-called Chua circuit [54]. The circuit consists of a nonlinear amplifier N which transforms input voltage $x(t)$ into the output $\alpha f(x)$ [32]. The parameter α characterizes the gain of N around $x=0$. The nonlinear amplifier has linear feedback which contains a series connection to a low-pass filter (RC') and LC resonance. It has been shown that this circuit can exhibit a transition from periodic oscillations to chaos via period doubling cascades, intermittency, and crises of chaotic attractors [37]. In our experiment we investigated two different values of the parameter α. The values correspond to chaotic attractors which appear after a period doubling cascade. Each chaotic signal $x(t)$, measured from the capacitor C, has been amplified and then digitized with the sampling period 20 μsec. This amplified and digitized signal is the scalar time series $s(n\tau)$ we investigated. (We also rescaled time to $\tau=0.02$.) The α values we investigated are $\alpha=17.4$ and 18.9. Further details concerning this circuit and its attractor can be found in the papers referenced above.

The method of time delays was used to reconstruct phase space vectors from $s(n)$:

$$\mathbf{y}(n)=(s(n),s(n+T),\ldots,s[n+(d-1)T])\ .$$

Time evolution in the reconstructed phase space is given by $\mathbf{y}(n)\mapsto\mathbf{y}(n+1)$. In order to find the embedding delay T and the embedding dimension d, we used the methods of average mutual information and false near neighbors, respectively [55,56]. The results of the calculations for $\alpha=17.4$ are shown in Figs. 2 and 3. The Figures indicate that the correct embedding time is $T=10$ and the correct embedding dimension is $d=3$. The other value of R also indicated $T=10$ and $d=3$.

The second system we have investigated is the Belousov-Zhabotinski (BZ) reaction. (We are indebted to Dr. Nicolas Tufillaro for providing us with the experimental data set we have used.) The BZ reaction has become one of the canonical experimental systems for research into chaotic dynamics [57,58]. For this system we reconstructed the phase space in the manner employed by Mindlin *et al.* [58]. The embedding method is not time delay. Instead it is a $d=3$ dimensional combination of filtering and differential embedding. If the scalar time series is $s(n)$, the first thing we did was re-

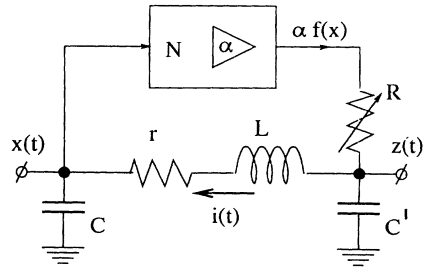

FIG. 1. A schematic diagram of the electronic circuit we investigated. For this circuit we collected data at R=3.38$k\Omega$, $L=145$ mH, $C=343$ nF, $C'=225$ nF, $r=347\Omega$, a sampling period of 20 μs, and $\alpha=17.4$ and 18.9.

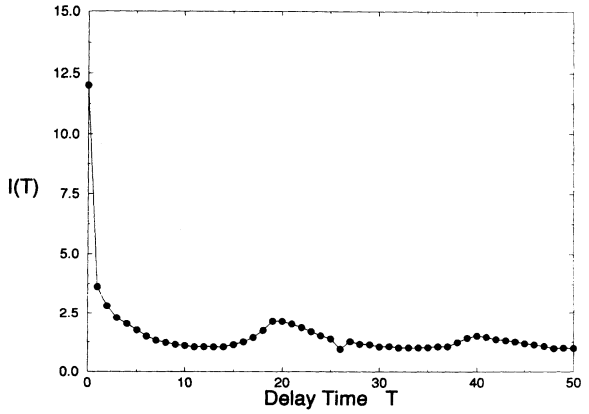

FIG. 2. The results of an average mutual information calculation for a time series taken from the electronic circuit shown in Fig. 1. The curve indicates that the optimal time delay is $T=10$.

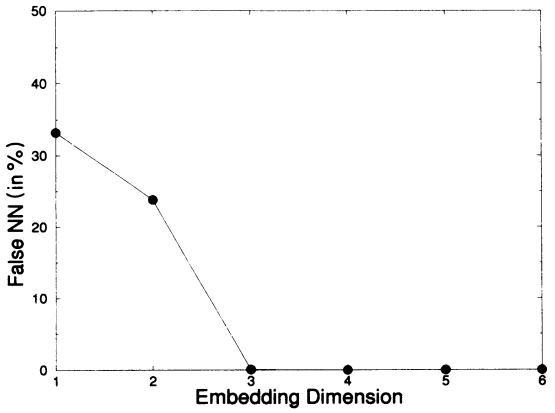

FIG. 3 The results of a false nearest neighbor calculation for a time series taken from the electronic circuit shown in Fig. 1. The graph indicates that the optimal embedding dimension is $d=3$.

move the mean value $s(n)\rightarrow s(n)-\bar{s}$. Next, we embedded the data via

$$z_1(n)=\sum_{j=1}^{n} s(n)\exp[-(i-j)/K] ,$$

$$z_2(n)=s(n) ,$$

$$z_3(n)=s(n)-s(n-1) .$$

For our numerical experiments we have used $K=100$. Finally, note that each component of **z** has a different scale. We found that for our data set the standard deviations of the components of **z** are $\sigma_1=33\,355$, $\sigma_2=1513$, and $\sigma_3=140$. We expect numerical problems if the components of the phase space vectors have such different scales. To overcome this, we identified $y_\alpha(n)=z_\alpha(n)/\sigma_\alpha$. [This rescaling is similar to the rescaling we performed on the passivation model, Eqs. (18).] It has been shown that this type of embedding is useful when constructing the template for this attractor. For our purposes this embedding emphasizes the fact that the world is not always time delay. Our procedure for modeling **F** does not assume a special form for the embedding and, thus, works for this embedding as well time delay embedding. We believe that this flexibility is an advantage that researchers should retain whenever possible.

The attractors associated with the Lorenz equations, the passivation model, the circuit, and the BZ reaction are shown in Figs. 4, 5, 6(a), 6(b), and 7. As one can see from Figs. 4–7, the data sets used to reconstruct the phase spaces are very clean. These figures show the typical data sets we will use to construct our models. For all systems the data sets provide a fair representation of their attractors, but the representations are by no means even close to complete.

When modeling a particular system, one should, in principle, utilize three different data sets. One data set should be used to construct the orthonormal basis functions $\pi^{(I)}$. A second data set should be used in the modeling procedure to obtain the optimal values for M, N_p, and the $\mathbf{p}^{(I)}$'s. A third data set should then be used to determine how well these optimal values model the dynamics.

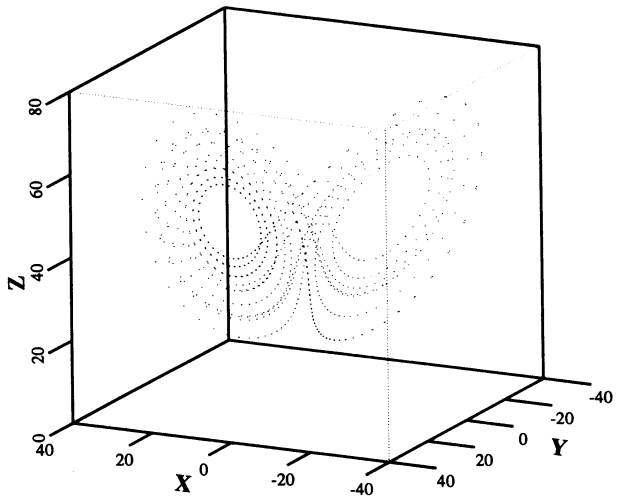

FIG. 4. The attractor associated with the Lorenz system given by Eqs. (17). In this figure we show $N=867$ data vectors. The sampling interval is $\tau=0.011\,53$ and is approximately the limit of our ability to determine accurately the correct vector field.

For our numerical experiments we have not done this. The following are the reasons for this decision.

The $\pi^{(I)}$'s are constructed from moments of the distribution [27,2]. In order to obtain the most accurate representation for the $\pi^{(I)}$'s, the first data sets should provide a picture of the attractor sufficiently complete that the moments can be accurately calculated. Clearly, this is not the case in Figs. 4 and 5. For example, the Lorentz data set that is shown favors the left lobe over the right lobe. It is quite likely that a different data set of this size will favor the right lobe over the left lobe. Thus, the $\pi^{(I)}$'s that are constructed on the data set shown in Fig. 4 will not be orthonormal on typical data sets of this size. A similar problem can be seen in Fig. 5 and is associated

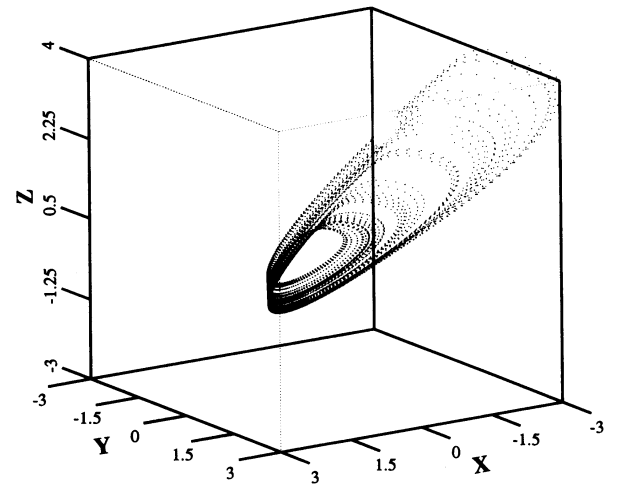

FIG. 5. The attractor associated with the passivation model given by Eq. (19). In this figure we show $N=5000$ data vectors. The rescaled sampling interval is $\tau=0.0085$.

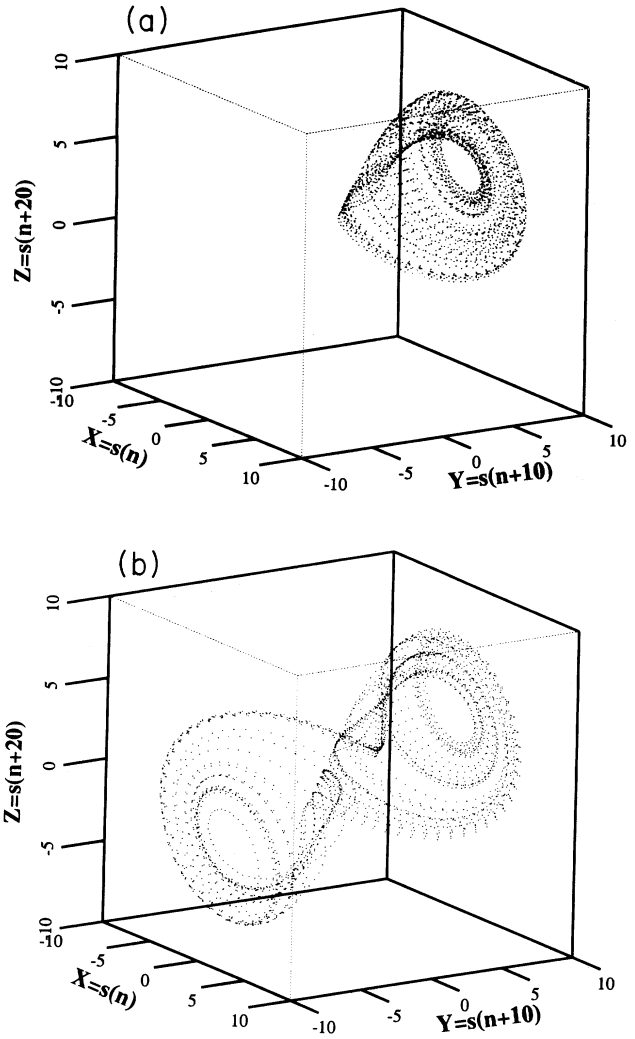

FIG. 6. The attractor obtained by embedding data from the electrical circuit shown in Fig. 1. We show $N=3000$ data vectors constructed using a time delay of $T=10$. (a) The resistance is $\alpha=17.4$. (b) The resistance is $\alpha=18.9$.

with the frequency with which an orbit explores the end of the "ear." This data set extends into the "ear" approximately 10 times. Other data sets that are this small may extend into the "ear" a different number of times. Poor results can be expected if the distribution of the data set used to calculate the $\pi^{(\mathrm{I})}$'s is reasonably different from the one used to calculate the $\mathbf{p}^{(\mathrm{I})}$'s. These difficulties vanish in the $N\rightarrow\infty$ limit, but we are using relatively small data sets so the difficulty must be addressed. By using the same data set to construct both the $\pi^{(\mathrm{I})}$'s and the $\mathbf{p}^{(\mathrm{I})}$'s, we can overcome this difficulty. To insure a fair test of the predictive ability of the fitted models, we have *always* used a different data set when determining how accurately the model predicts the dynamics.

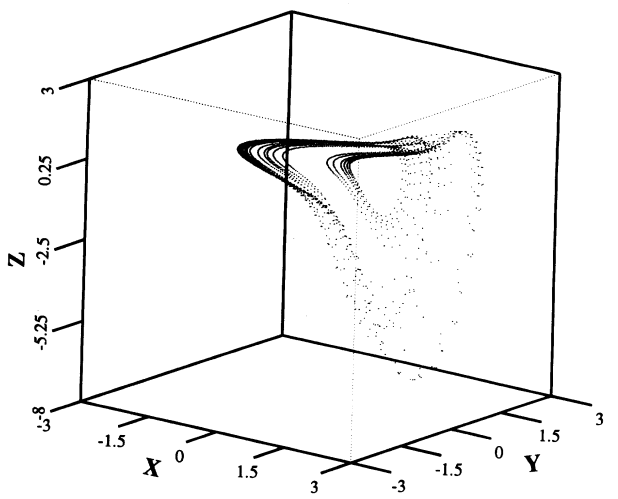

FIG. 7. The attractor obtained by embedding data from the BZ reaction. The embedding is the one used by Mindlin *et al.* [58] and $N=3000$ vectors are shown.

A. Prediction error versus sampling interval

In this subsection we examine the behavior of our modeling method as a function of τ, the interval between measurements, and M, the order of the Adams predictor corrector. In order to provide the best of all possible worlds, we have used all three components of the Lorenz equations, Eqs. (17), as our time series:

$$\mathbf{y}(n)=(x(n\tau),y(n\tau),z(n\tau)) .$$

At this time we are specifically not interested in the ability of the method to provide the optimal truncation in N_p. Rather, we are focusing on the behavior of the method as a function of τ and M. Therefore, we have taken the liberty of fixing $N_p=2$.

Several numerical data sets for the Lorenz system were generated by integrating Eqs. (17). Each data set used the same initial condition. The total interval of time represented by a data set is $T_T=N\tau$, where N is the number of points in the data set. For all of our test cases we always adjusted both N and τ so that T_T remained constant. As we alluded to above, the construction of the orthonormal polynomial basis set depends on the amount of the attractor covered by the time series. By keeping T_T fixed and using the same initial condition, we insure that for each tested value of τ the basis functions are always constructed from a data set with the same coverage on the attractor. In this way each test gets the same "look" at the attractor. (We found that the results were qualitatively the same if we did not insist on constant T_T.) Figure 4 shows the amount of coverage on the attractor that will be used for the numerical experiments.

We will report the following relative prediction error:

$$\chi_A=\left[\frac{1}{N}\sum_{n=1}^{N}\frac{\left|\mathbf{y}(n+1)-\mathbf{y}(n)-\tau\sum_{j=1}^{M}a_j^{(M)}\mathbf{F}[\mathbf{y}(n+1-j)]\right|^2}{|\mathbf{y}(n+1)-\mathbf{y}(n)|^2}\right]^{1/2}$$

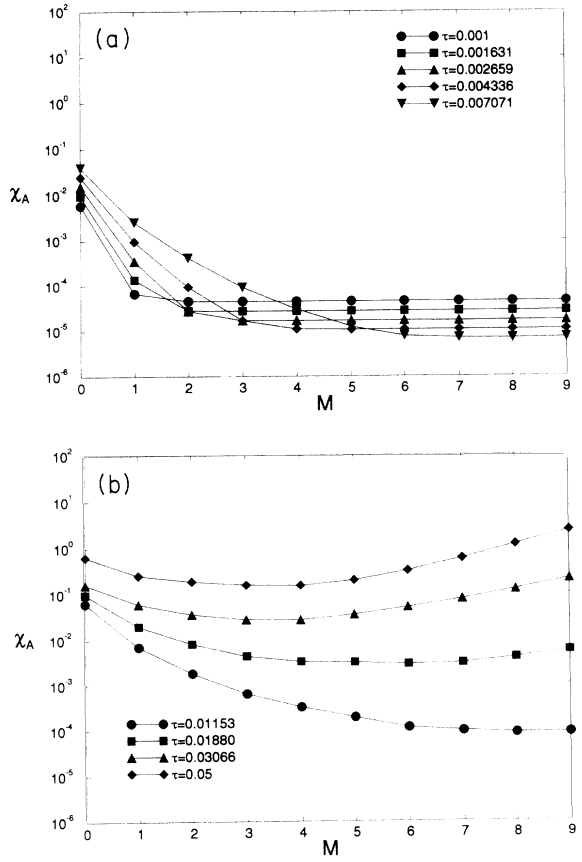

FIG. 8. The relative prediction error for the Lorenz system ($N_p=2$): (a) The circles, squares, up triangles, diamonds, and down triangles correspond to $\tau=0.001$, 0.001 631, 0.002 659, 0.004 336, and 0.007 171. (b) The circles, squares, triangles, and diamonds correspond to $\tau=0.011\,53$, 0.018 80, 0.030 66, and 0.05, respectively.

The numerator in this equation is the error between the actual position of the point $\mathbf{y}(n+1)$ and the predicted position of the point after one Adams integration step. The denominator normalizes the error to the size of the jump in phase space taken by the integration step. The normalization is necessary since a numerator of size 0.001 will probably be acceptable if the denominator is of size 1, but will probably be unacceptable if the denominator is of size 0.001.

Figures 8(a) and 8(b) indicate the results of numerical experiments on the Lorenz system. The error χ_A often decreases by several orders of magnitude when values of M greater than 0 are used. We remind the reader that $M=0$ corresponds to an implicit Euler integration method. The majority of the previous research on modeling numerical time series used an Euler integration method for constructing the model. Figure 8 clearly indicates that by using the Adams method can construct an ODE model (from a time series) that is more accurate then the Euler method allows. The benefit of this is that the sampling interval can be greatly increased (in this case by an order of magnitude) before the errors from the Adams method are comparable to those of the Euler method. To further emphasize this point, we present the coefficients obtained by the methods in Tables I and II. The Tables indicate that one is able to obtain much better estimates of the coefficients of **F**, the vector field, by using the Adams method (for $M>0$) instead of the implicit Euler method.

As a final comment on this experiment, we notice that none of the methods produced relative errors that were smaller than $\sim 10^{-5}$. The data used for the modeling only contained five decimals places of accuracy. Thus the errors in the data are $\sim 10^{-5}$. Given this, we see that for this test the errors in the Adams method are the same as the errors inherent in the data. We found similar behavior when the passivation model was examined. The Adams method permitted much larger step sizes than the Euler method.

B. Optimal modeling

In this subsection we will demonstrate that the MDL criterion provides a means of determining the optimal value for truncating the orders of the modeling procedure presented in Sec. II. In Sec. III A we investigated the behavior of our fitting procedure as a function of the sampling interval τ. In this subsection we will use a fixed

TABLE I. The polynomial coefficients obtained by modeling data from the true Lorenz equations with $\tau=0.001$. ($M=0$) The modeling method is equivalent to an implicit Euler integrator. ($M=2$) Except for the constant coefficient of the dz/dt equation (which should all be 0), this model finds the exact coefficients for the Lorenz system.

	$M=0$			$M=2$		
Index	B_1	B_2	B_3	B_1	B_2	B_3
0	0.000	−0.139	3.018	0.000	0.000	−0.001
1	−16.500	46.264	−0.042	−16.000	45.920	0.000
2	16.139	−1.427	0.043	16.000	−1.000	0.000
3	0.000	0.022	−4.233	0.000	0.000	−4.000
4	0.000	0.005	−0.003	0.000	0.000	0.000
5	0.000	−0.003	1.026	0.000	0.000	1.000
6	0.008	−1.010	0.000	0.000	−1.000	0.000
7	0.000	0.000	−0.017	0.000	0.000	0.000
8	0.000	0.014	0.000	0.000	0.000	0.000
9	0.000	0.000	0.004	0.000	0.000	0.000

TABLE II. The polynomial coefficients obtained by modeling data from the Lorenz system for $\tau=0.011\,53$. ($M=0$) The modeling method is equivalent to an implicit Euler integrator. The fitted vector field is not very close to the true vector field. ($M=9$) The coefficients are from the Adams modeling method. These values are much closer to the correct values than the ones produced by the Euler integrator.

	$M=0$			$M=9$		
Index	B_1	B_2	B_3	B_1	B_2	B_3
0	0.136	−4.144	33.92	0.000	−0.022	−0.009
1	−22.277	48.626	−0.113	−16.000	45.920	0.000
2	17.942	−5.152	0.248	16.000	−1.000	0.000
3	−0.019	0.419	−6.548	0.000	0.001	−4.000
4	−0.004	0.073	−0.003	0.000	0.000	0.000
5	0.003	−0.061	1.237	0.000	0.000	1.000
6	0.104	−1.090	−0.002	0.000	−1.000	0.000
7	0.000	0.005	−0.183	0.000	0.000	0.000
8	−0.009	0.144	−0.004	0.000	0.000	0.000
9	0.006	−0.011	0.039	0.000	0.000	0.000

sampling interval for the numerical tests.

In general, we expect that χ^2_{MDL} will be relatively insensitive to changes in M when compared to changes in N_p. This lack of sensitivity can be understood by examining Eq. (10). Changing M will cause only slight changes in $\mathbf{F}$ (via the $\mathbf{p}^{(\mathrm{I})}$'s) and χ^2_{ML} [the first term in Eq. (10)]. This will be especially true if $M \geq 4$. [See Fig. 8(a) for $\tau \leq 0.004\,336$ as an example.] The contribution from the second term is typically small when compared to χ^2_{ML} and will grow linearly with M as M_P increases. In addition, only small changes can be expected to occur in Θ since the changes in $\mathbf{p}^{(\mathrm{I})}$ are small. In contrast, changing N_p can be expected to cause major changes in $\mathbf{F}$ and χ^2_{ML}. In addition, a large jump can be expected to occur in M_P and Θ. This jump will cause a large change in the second term in Eq. (10).

A special case for the MDL criterion occurs when the true vector field in the phase space of the data is a polynomial. The Lorenz system with $\mathbf{y}=[x,y,z]$ is an example of this. As long as τ is not too large, the $\mathbf{p}^{(\mathrm{I})}$'s for $I_1 > 2$ will be very small. In principle they should be exactly zero. Numerically, all values of $\mathbf{p}^{(\mathrm{I})}$ below a certain tolerance should be interpreted as zero. For the Lorenz system we, arbitrarily, set that tolerance to 5.0×10^{-5}. Any $\mathbf{p}^{(\mathrm{I})}$'s whose absolute value is less than 5.0×10^{-5} will be interpreted as zero. The coefficients whose values are zero will not contribute to $\|\Theta\|^2$ and should not be included when calculating M_P. Thus we expect χ_{MDL} to reach a minimum for $N_p=2$ and remain there as N_p increases beyond 2. We hasten to point out that this behavior will *only* occur when components of the true vector field can be exactly represented by finite order polynomials. This type of behavior will not occur in general.

In Fig. 9 we present some of the values of χ_{MDL} that were obtained when using data from the Lorenz equations with a sampling interval of $\tau=0.004\,336$ and a data set of size $N=2306$. The figure indicates that χ_{MDL} saturates at $N_p=2$ for reasons that we have just discussed. This means that the optimal choice is $N_p=2$. Although it is difficult to see in the Figure, comparison of the actual numerical values indicates that the optimal value of M is $M=5$. (The actual values are $\chi_{\mathrm{MDL}}=1.45$, 0.7034, 0.6490, 0.6493, and 0.6542 for $M=3$–7.) Obviously, χ_{MDL} will be much larger than the plotted values for $N_p<2$ (the actual values are $\sim10^4$). From Fig. 8(a) we see that for $N_p=2$ and $M=5$, the errors associated with the prediction are the same order as the error inherent in the data.

A more realistic test for the optimization method occurs when modeling the rescaled data $[\mathbf{y}(n)=(U(n\tau),V(n\tau),W(n\tau))]$ from the passivation model. In this case we are attempting to model Eqs. (19) which clearly cannot be modeled as a finite polynomial. As stated above, there are numerical problems associated with calculating the $\pi^{(\mathrm{I})}$'s for large values of $N_p=I_1$. For the data shown in Fig. 5 we were able to find all polynomials up to and including $N_p=I_1=5$ before encountering these problems. We contend that a fifth order polynomial should be sufficient for many modeling purposes. The results of our attempt at optimal modeling are shown in Fig. 10. The optimal model is $N_p=5$ and $M=2$.

We have two comments about the vector field obtained from modeling numerically generated data from the passivation model. The MDL criterion can only indicate the

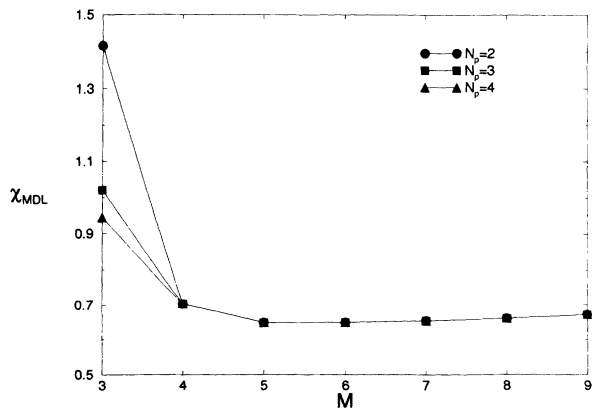

FIG. 9. The values of χ_{MDL} for the model that was trained on data from the Lorenz equations (17). The circles, squares, and triangles correspond to $N_p=2$, 3, and 4, respectively.

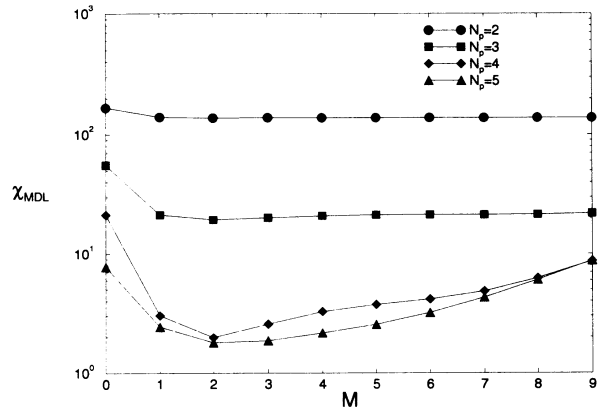

FIG. 10. The values of χ_{MDL} for the model that was trained on data from the passivation model [Eqs. (19)] with $\tau=8.5$. The circles, squares, diamonds, and triangles correspond to $N_p=2$, 3, 4, and 5, respectively.

best model among all of the models that are tested. There is no evidence that we have reached the point where larger values of N_p would not be useful. It could be that a model with $N_p=6$ would produce a lower value of χ_{MDL}. Our second comment is that the model selected by the MDL criterion is different, and therefore closer to optimal, than the one that would be produced by the standard Euler method.

The best test for the optimization method comes from modeling data obtained from the electronic circuit shown in Fig. 1. The same $N=3000$ reconstructed phase space vectors $\mathbf{y}(n)$ that are displayed in Fig. 6(a) were used to train the model. The rescaled sampling interval for the data was fixed at $\tau=0.02$. Although the vector field for this attractor is unknown, it is very unlikely that it can be written as a finite polynomial. Therefore, the $\mathbf{F}$ that we calculate will only be an approximation to the true vector field. The results of minimizing Eq. (10) are shown in Fig. 11. The Figure indicates a clear minimum at $N_p=4$ and $M=3$. This minimum represents the optimal model

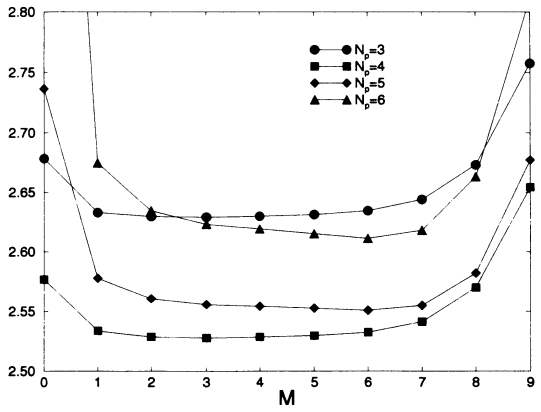

FIG. 11. The values of χ_{MDL} for the model that was trained on data from the electronic circuit for $\alpha=17.4$. The circles, squares, diamonds, and triangles correspond to $N_p=3$, 4, 5, and 6, respectively.

for the data set tested. (Notice that this optimal model could not have been found if an Euler integrator was used to model the data.) In order to model a data set, we are required to indicate a noise level σ. For the numerical examples previously investigated, it is easy to control the noise level by printing only a few digits to the data file. For experimental systems the noise level is harder to estimate, much less control. For the electronic circuit data we have estimated the noise level at $\sigma=0.01$.

At this point it has been demonstrated that the MDL principle is capable of determining an optimal model (within the class of models that are polynomials) for an experimental data set. By optimal we mean that the method is capable of indicating the value of N_p at which the order of the polynomial should be truncated. It can also indicate the value of M which produces the best set of $\mathbf{p}^{(I)}$'s. In Table III we show the optimal models for all of the dynamical systems we have tested. Except for the passivation model discussed above, we were always able to find the optimal model. (We could always construct $\pi^{(I)}$'s for values of $N_p=I_1$ that were beyond what was needed for optimal modeling.)

As we alluded to in Sec. II, there are many different basis sets that can be used for polynomial models. A natural question that arises is: What benefit is obtained by using a basis that is orthonormal on the attractor? To answer that question, we will plot the coefficients associated with the orthonormal basis set, the $\mathbf{p}^{(I)}$'s, and the coefficients associated with the standard basis set, the $\mathbf{B}^{(I)}$'s from Eq. (12).

Figures 12(a) and 12(b), respectively, show the $\mathbf{p}^{(I)}$'s and the $\mathbf{B}^{(I)}$'s associated with the $N_p=5$, $M=2$ model obtained from numerically generated data from the passivation model, Eq. (18). The circles, squares, and triangles are the coefficients of F_1, F_2, and F_3, respectively. In these figures a quadratic model is associated with $I<10$. There is a significant difference between the two figures. When the standard basis is used, many of the coefficients for $I>10$ are of comparable size. It is very difficult to determine which of the $\mathbf{B}^{(I)}$'s, if any, should be interpreted as zero. Yet an inspection of Eqs. (19) indicates that F_1 is linear and F_3 is only weakly nonlinear, while F_2 is strongly nonlinear. Furthermore, the nonlinearities in F_3 are only quadratic. Thus, *all* of the $B_1^{(I)}$'s and $B_3^{(I)}$'s for $I>10$ should vanish, while *none* of the $B_2^{(I)}$'s should vanish. This type of behavior would be very difficult to determine from Fig. 12(a). In contrast, Fig. 12(b) clearly indicates that the $p_1^{(I)}$'s and the $p_3^{(I)}$'s vanish for $I>10$. From this figure one can convincingly determine that F_1

TABLE III. Optimal models from the MDL criterion.

System	N_p	M	χ_{MDL}	σ
Lorenz equations	2	5	0.649	10^{-5}
Chemical equations	5	2	1.797	10^{-4}
Electronic circuit $\alpha=17.4$	4	3	2.528	10^{-2}
Electronic circuit $\alpha=18.9$	5	3	2.590	10^{-2}
BZ reaction	7	5	10.01	10^{-3}

and F_3 are at best quadratic.

The conclusion that one draws from Figs. 12 is that the orthonormal basis presents a much clearer picture of the relative sizes and importance of the various coefficients associated with a polynomial model of a vector field. If one actually needs to know the $\mathbf{B}^{(\mathbf{I})}$'s we suggest performing the calculations in the orthonormal basis. After obtaining the $\mathbf{p}^{(\mathbf{I})}$'s, determine which coefficients are zero and reset these $\mathbf{p}^{(\mathbf{I})}$'s to zero. Only then should the $\mathbf{B}^{(\mathbf{I})}$'s be evaluated. The formula necessary for transforming from the $\mathbf{p}^{(\mathbf{I})}$'s to the $\mathbf{B}^{(\mathbf{I})}$'s can be found in a previous paper [2].

C. Synchronization of models and time series

In their papers Pecora and Carroll point out that synchronization is structurally stable. They were able to synchronize two electrical circuits that were constructed using off the shelf components [29]. No two systems constructed in this manner are ever "identical." [Formally, if the parameters of the drive and response systems are in any way different, then the synchronized oscillations do not reside on the manifold of Eq. (14) and are not "identical" [32].] Structural stability implies that it should be possible to synchronize one of the models we have constructed to an experimentally measured time series. Our numerical experiments will attempt to synchronize F's trained using data sets from the passivation model, the electronic circuit, and the BZ reaction to time series from the respective dynamical systems.

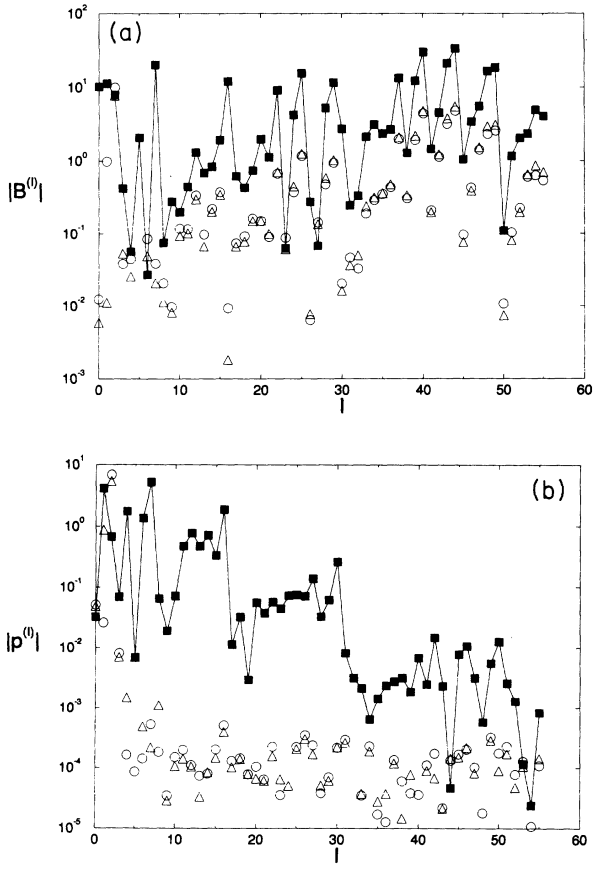

FIG. 12. The coefficients of the polynomial used to model the vector field F. The circles, squares, and triangles indicate the coefficients of F_1, F_2, and F_3, respectively. (a) The $\mathbf{B}^{(\mathbf{I})}$'s are the coefficients of F when it is written in the form of Eq. (12). (b) The $\mathbf{p}^{(\mathbf{I})}$'s are the coefficients of F when it is written in the form of Eq. (1). For this case the representation using $\mathbf{p}^{(\mathbf{I})}$'s more clearly indicates that F_1 and F_2 are basically linear.

Synchronizing a model to a time series requires that we integrate the ODE under conditions where the drive variable comes from the time series. For time series data the value of the drive variable is only known at fixed time intervals. We will investigate modified FY driving as well as PC driving. When using modified FY driving for large values of ϵ [see Eq. (16)], the equations of motion will be stiff since the time scale for the driven equation is different from that of the undriven equations. We will use a backwards difference integrator to integrate the equations of motion when modified FY driving is used [59]. When PC driving is used, we will use a variable step size Runge-Kutta integrator [40]. To obtain estimates of the driving variable at values of time not in the data set, we linearly interpolated between the vector just before and the vector just after the time in question. For all of our tests the data set used as the drive was not the data set used to train the $\pi^{(\mathbf{I})}$'s and the $\mathbf{p}^{(\mathbf{I})}$'s.

As a preliminary check of our models, F, we selected an initial condition on their respective attractors and integrated the fitted vector fields forward in time using both variable time step Runge-Kutta and variable time step, variable order Adams integration routines [59,40]. For all of our test systems (the passivation model, the electronic circuit at both values of R, and the BZ reaction), the fitted F's traced out attractors that were visually the same as those shown in Figs. 6 and 7.

Each numerical experiments involved three different orbits: $\mathbf{y}(n)$ denotes the orbit that was used as the drive. $\mathbf{w}(n)$ denotes the orbit that results when the initial condition $\mathbf{y}(1)$ is integrated forward in time. This orbit resides on the attractor. Initially, it is close to $\mathbf{y}(n)$ but soon diverges due to chaos. $\mathbf{z}(n)$ denotes the orbit that results when F is driven using either modified FY or PC driving. The initial condition we used for the $\mathbf{z}(n)$ orbit is not part of the $\mathbf{y}(n)$ data set.

In Figs. 13 we have plotted $D_z(n)=|\mathbf{y}(n)-\mathbf{z}(n)|$ and $D_w(n)=|\mathbf{y}(n)-\mathbf{w}(n)|$ for the electronic circuit and the BZ reaction. When modified FY driving is used, we set $\epsilon=20$. In all of the figures we have used y_3 as the driving variable. $D_w(n)$ (the dashed line) starts at zero but quickly grows. The oscillations occur because the two orbits, $\mathbf{y}(n)$ and $\mathbf{w}(n)$, are evolving independently on the same attractor. In contrast, $D_z(n)$ (the solid line) starts with a

TABLE IV. The results of our synchronization tests using both modified FY and PC driving on all four dynamical systems.

System	PC Driving			FY Driving		
Passivation model	No	Yes	No	No	Yes	No
Electronic circuit	No	No	Yes	No	Yes	Yes
$\alpha=17.4$						
$\alpha=18.9$	No	No	Yes	No	Yes	Yes
BZ reaction	No	No	Yes	No	No	Yes

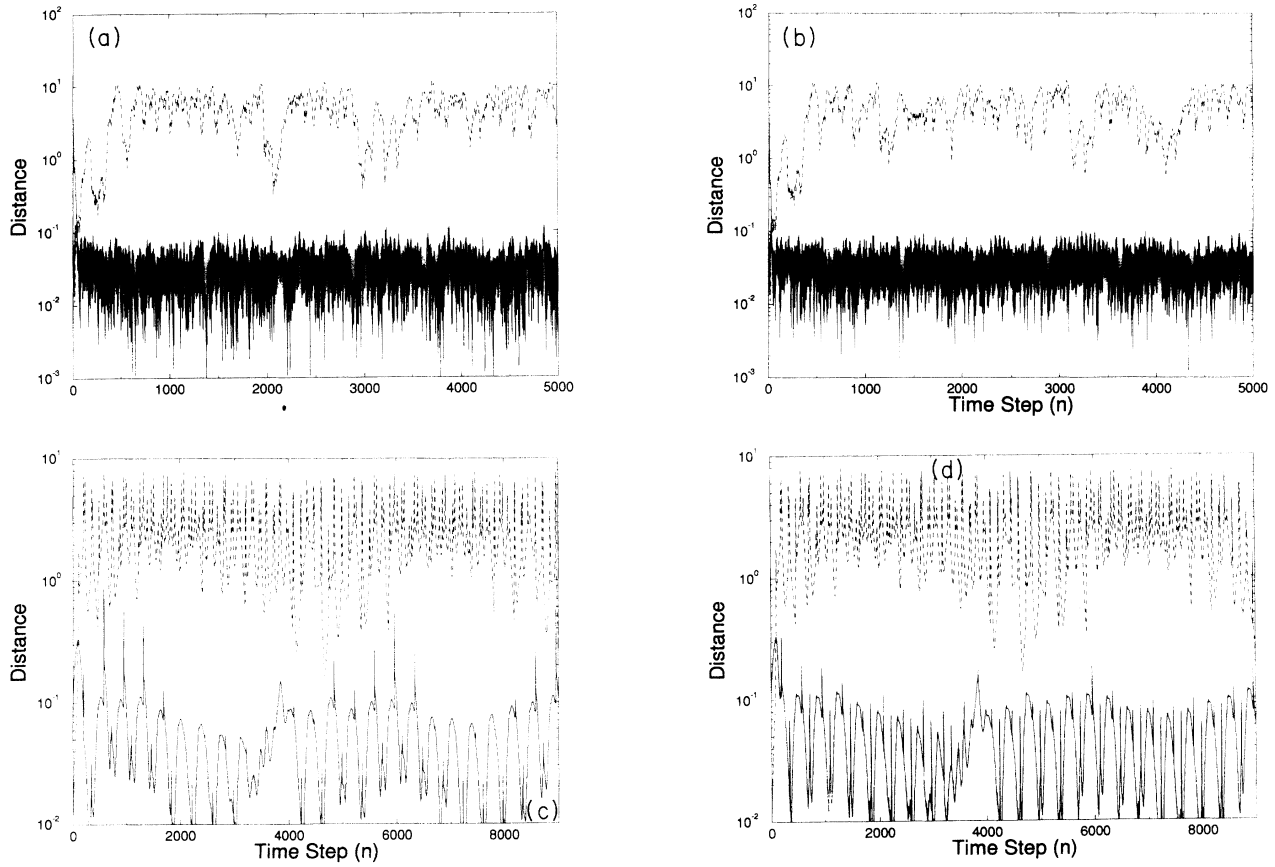

FIG. 13. The distances between the orbits $\mathbf{y}(n)$, $\mathbf{z}(n)$, and $\mathbf{w}(n)$. The dashed line is $D_z(n)=|\mathbf{y}(n)-\mathbf{z}(n)|$, while the solid line is $D_w(n)=|\mathbf{y}(n)-\mathbf{w}(n)|$. Synchronization between $\mathbf{y}(n)$ and $\mathbf{w}(n)$ is clearly demonstrated. (a) PC synchronization of the model for the electronic circuit and a data set. We used y_3 as the driving variable and the data is from $\alpha=17.4$. (b) Modified FY synchronization of the same model and data. We used $\epsilon=20$. (c) PC synchronization of the model for the BZ reaction and a data set. We used y_3 as the driving variable. (d) Modified FY synchronization of the same model and data. We used $\epsilon=20$.

finite separation that decreases to a small value. This implies that the driven orbit has become synchronized to the drive orbit.

In Table IV we present the results of our experiments using both modified FY and PC driving on all four dynamical systems. Each column represents the driving y_1, y_2, and y_3 respectively. One sees that for each system and each driving method we are able to find at least one means of synchronizing the optimal model **F** to a data set.

IV. SUMMARY AND CONCLUSIONS

In this, our final section, we will summarize the results, and present our conclusions. We shall also make a few speculations about possible applications for the methods we have developed.

Our goal was to develop a method for generating a vector field for an ODE whose dynamics mimicked the dynamics of the time series. We wrote the vector field of the ODE, **F**, as an expansion in terms of polynomials that are orthonormal on the attractor given by the data. These polynomials form a natural basis set in which to write a polynomial model [2]. The method we developed for training the expansion coefficients—we called them $\mathbf{p}^{(I)}$'s—involved modeling the dynamics as Adams predictor-corrector integration [see Eqs. (1) and (2)]. When the order of the Adams integration is $M=0$, this method reduces to implicit Euler integration. In Figs. 8(a) and 8(b) and Tables I and II we demonstrate that the Adams method allows one to use a much larger sampling interval than the Euler method. Thus the Adams method is an improvement over the Euler method, which was used by most previous researchers. The sampling interval that is needed if the Euler method is to produce good results is far too small for most experimental circumstances. The fact that the sample interval for the Adams method could be as much as an order of magnitude larger than the Euler method implies it is a much better modeling procedure for those people who wish to develop ODE models from experimental data.

We also demonstrated that the numerical values for the expansion coefficients are more reliable when calculated using the orthonormal basis than they are using the standard basis. In Figs. 12(a) and 12(b) it was demonstrated that it is sometimes difficult to decide which coefficients can be ignored, i.e., set to zero, when the standard basis set is used. In contrast, it is much easier to determine which coefficients can be ignored, i.e., set to zero, when

the orthonormal basis set is used.

The procedure we used to train the model involved minimizing Eq. (10). One of the major benefits of the training procedure we have used is associated with the second and third terms in Eq. (10). These terms are associated with the size of the vector field **F** used to model the data. This term grows as the number of nonzero parameters increases. By minimizing Eq. (10), we are able to obtain the *optimal* model, from the class of polynomial models, for the data this is being modeled. This ability to determine the optimal model is one of the benefits of our method. We have observed that for all cases the optimal model was not the one that was produced by the implicit Euler method. Thus it is the conjunction of the Adams method of Eq. (2) and the penalty terms of Eq. (10) that provides the improvement in modeling.

Once one obtains a vector field from the data, the natural question that must be addressed is: How good is the model? Put another way, if $\mathbf{G}(\mathbf{z})$ is the true vector field in the phase space, then how close is $\mathbf{F}(\mathbf{z})$ to $\mathbf{G}(\mathbf{z})$? One could just integrate **F** and look at the attractor to obtain a visual answer to this question. This is unsatisfactory if for no other reason than one cannot do this when the embedding dimension is greater than 3. In previous papers we examined the spectrum of Lyapunov exponents and the invariant density [16,2]. In this paper we used an indirect and subtle test. We attempted to synchronize the model **F** to a time series from the experiment. Synchronization will occur only if $[\mathbf{G}(\mathbf{z})-\mathbf{F}(\mathbf{z})]\simeq 0$ near the attractor (see Sec. II). We used two different types of synchronization procedure. We found that for *all* of the test cases, at least one variable could be used as a drive that would synchronize the optimal model to a time series. This occurred for both FY and PC driving. (There were instances where a particular variable would synchronize the model to the time series when modified FY driving was used but would not synchronize the model when PC driving was used. The converse was never observed.) Since we were always able to synchronize the optimal model to a time series, we conclude that in the vicinity of the attractor the optimal model **F** is very close to the true vector field.

We are now in a position to discuss possible applications. What does one do with a "synchronizer"? Such models can be used for many of the traditional tasks associated with modeling: detection and classification, for example. Suppose one has previously constructed a tailored synchronizing model for a system of interest. The modeler is then presented with a mystery signal and asked whether a signal from the system is present (possibly masked by noise). If the model synchronizes to the incoming signal (when noise is present, the model will only be able to synchronize to within the noise level), then the system of interest has been detected. Given several distinct systems, each with a tailored synchronizer, if the incoming signal synchronizes to one and not the others, this would constitute classification. Another potential area of application is in health monitoring or fault detection. As with detection, we start with a measured time series taken from the system of interest and construct a synchronized model. The numerical model is then run in tandem with the system as a real-time monitor of its operating state. If the system undergoes a change of state, possibly signaling impending failure, then the system could be modified or shut down to prevent damage. The same type of tandem operation could be done as part of a maintenance program to determine the level of wear and tear on a device. In short, a "synchronizer" answers the question: How do we know when a system, which normally oscillates chaotically, is behaving properly?

ACKNOWLEDGMENTS

This work was supported by grants from the Department of Energy (Grant No. DE-FG03-90ER14148), the University of California University Wide Energy Research Group, and the Physics Department of the College of William and Mary. We would also like to thank the research staff at INLS—in particular, Jesse Goldberg—for many helpful discussions.

[1] F. C. Moon, *Chaotic and Fractal Dynamics* (Wiley, New York, 1992); S. Vohra, M. Spano, M. Shlesinger, L. Pecora, and W. Ditto, *Proceedings of the 1st Experimental Chaos Conference* (World Scientific, New Jersey, 1992).

[2] R. Brown, INLS Report April 1, 1993 (unpublished).

[3] R. Mañé, in *Dynamical Systems and Turbulence, Warwick, 1980,* edited by D. A. Rand and L. S. Young, Lecture Notes in Mathematics Vol. No. 898 (Springer, Berlin, 1981).

[4] F. Takens, in *Dynamical Systems and Turbulence, Warwick, 1980,* edited by D. A. Rand and L. S. Young, Lecture Notes in Mathematics Vol. No. 898 (Springer, Berlin, 1981).

[5] T. Sauer, J. A. Yorke, and M. Casdagli, J. Stat. Phys. **65**, 579 (1991).

[6] J. D. Farmer and J. J. Sidorowich, Phys. Rev. Lett. **59**, 845 (1987).

[7] S. M. Hammel, Phys. Lett. A **148**, 421 (1990).

[8] E. Kostelich, Physica D **58**, 138 (1992).

[9] T. Sauer, Physica D **58**, 193 (1992).

[10] J. Jiménez, J. A. Moreno, and G. J. Ruggeri, Phys. Rev. A **45**, 3553 (1992).

[11] A. S. Lapedes and R. M. Farber, in *Neural Information Processing Systems,* edited by D. Z. Anderson (AIP, New York, 1988).

[12] D. S. Broomhead and D. Lowe, Complex Syst. **2**, 321 (1988).

[13] A. S. Weigend, B. A. Huberman, and D. E. Rumelhart, Int. J. Neur. Syst. **1**, 193 (1990).

[14] J. L. Hudson, M. Kube, R. A Adomaitis, I. G. Kevrekidis, A. S. Lapedes, and R. M. Farber, Chem. Eng. Sci. **45**, 2075 (1990).

[15] M. J. D. Powel, University of Cambridge Tech. Report, 1985 (unpublished).

[16] H. D. I. Abarbanel, R. Brown, and J. B. Kadtke, Phys. Rev. A **41**, 1782 (1990).

[17] J. P. Crutchfield and B. S. McNamara, Complex Syst. **1**, 417 (1987).

[18] T. Eisenhammer, A. Hübler, N. Packard, and J. A. S. Kelso, Biol. Cybernetics **65**, 107 (1991).
[19] J. L. Breeden and A. Hübler, Phys. Rev. A **42**, 5817 (1990).
[20] J. S. Brush and J. B. Kadtke, in ICASSP-92, *1992 IEEE International Conference on Acoustics, Speech, and Signal Processing* (IEEE, New York, 1992).
[21] G. Gouesbet, Phys. Rev. A **43**, 5321 (1991); **46**, 1784 (1992).
[22] E. Baake, M. Baake, H. G. Bock, and K. M. Briggs, Phys. Rev. A **45**, 5224 (1992).
[23] H. D. I. Abarbanel, R. Brown, J. J. Sidorowich, and L. S. Tsimring, Rev. Mod. Phys. **65**, 1331 (1993).
[24] M. Casdagli, Physica D **35**, 335 (1989).
[25] M. Casdagli, D. Des Jardins, S. Eubank, J. D. Farmer, J. Gibson, N. Hunter, and J. Theiler, in *Applied Chaos,* edited by J. Kim (Addison-Wesley, Reading, MA), 1992.
[26] R. Brown, Phys. Rev. E **47**, 3962 (1993).
[27] M. Geona, F. Lentini, and V. Cimagalli, Phys. Rev. A **44**, 3496 (1991).
[28] H. Fujisaka and T. Yamada, Prog. Theor. Phys. **69**, 32 (1983).
[29] L. M. Pecora and T. L. Carroll, Phys. Rev. Lett. **64**, 821 (1990); Phys. Rev A **44**, 2374 (1991).
[30] V. S. Afraimovich, N. N. Verichev, and M. I. Rabinovich, Radiophys. Quantum Electron. **29**, 747 (1986).
[31] A. R. Volkovskii, and N. F. Rulkov, Pis'ma Zh. Tekh. Fiz. **15**, 5 (1989) [Sov. Tech. Phys. Lett. **15**, 249 (1989)].
[32] N. F. Rulkov, A. R. Volkovskii, A. Rodríguez-Lozano, E. Del Río, and M. G. Velarde, Int. J. Bif. Chaos **2**, 669 (1992).
[33] J. M. Kowalski, G. L. Albert, and G. W. Gross, Phys. Rev. A **42**, 6260 (1990).
[34] P. Badola, S. S. Tambe, and B. D. Kulkarni, Phys. Rev. A **46**, 6735 (1992).
[35] M. de Sousa Vieira, A. L. Lichtenberg, and M. A. Lieberman, Phys. Rev. A **46**, R7359 (1992).
[36] R. He and P. G. Vaidya, Phys. Rev. A **46**, 7387 (1992).
[37] A. R. Volkovski and N. F. Rulkov, Pis'ma Zh. Tekh. Fiz. **14**, 1508 (1988) [Sov. Tech. Phys. Lett. **14**, 656 (1988)].
[38] J. C. Roux, R. H. Simoyi, and H. L. Swinney, Physica D **8**, 257 (1983).
[39] J.-P. Eckmann and D. Ruelle, Rev. Mod. Phys. **57**, 617 (1985).
[40] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes,* 2nd ed. (Cambridge University, New York, 1992).
[41] J. Theiler (private communication).
[42] J. Rissanen, Ann. Stat. **11**, 416 (1983); Automatica **14**, 465 (1978); *Stochastic Complexity in Statistical Inquiry* (World Scientific, New Jersey, 1989).
[43] E. Kreyszig, *Differential Geometry and Riemannian Geometry* (University of Toronto, Toronto, 1968).
[44] L. Fox and D. F. Mayers, *Computing Methods for Scientists and Engineers* (Clarendon, Oxford, 1968).
[45] N. F. Rulkov and A. R. Volkovskii, in *Chaos in Communications,* edited by Louis M. Pecora, SPIE Proc. Vol. 2038 (SPIE, Bellingham, WA, 1993).
[46] E. N. Lorenz, J. Atmos. Sci. **20**, 130 (1963).
[47] F. N. Albahadily, J. Chem. Phys. **90**, 822 (1989).
[48] Y. Xu and M. Schell, J. Phys. Chem. **94**, 7137 (1990).
[49] M. R. Bassett, and J. L. Hudson, J. Phys. Chem. **92**, 6963 (1988).
[50] N. Fetner and J. L. Hudson, J. Phys. Chem. **94**, 6506 (1990).
[51] M. R. Bassett and J. L. Hudson, J. Phys. Chem. **93**, 2731 (1989).
[52] J. K. McCoy, P. Parmanada, R. W. Rollins, and A. J. Markworth, J. Mater. Res. **8**, 1858 (1993).
[53] L. E. Reichl, *A Modern Course in Statistical Physics* (University of Texas, Austin, 1980).
[54] L. O. Chua, M. Komuro, and T. Matsumoto, IEEE Trans. Circ. Syst. **33**, 1073 (1982).
[55] A. M. Fraser, and H. L. Swinney, Phys. Rev. A **33**, 1134 (1986); A. M. Fraser, IEEE Trans. Info. Theory **35**, 245 (1989).
[56] M. B. Kennel, R. Brown, and H. D. I. Abarbanel, Phys. Rev. A **45**, 3403 (1992).
[57] D. P. Lathrop and E. J. Kostelich, Phys. Rev. A **40**, 4028 (1989).
[58] G. B. Mindlin, H. G. Solari, M. A. Natiello, R. Gilmore, and X.-J. Hou, J. Nonlinear Sci. **1**, 147 (1991).
[59] J. R. Ford and W. R. Borland, Los Alamos National Laboratory, Common Los Alamos Mathematical Software Compendium, Document No. CIC 148, 1988 (unpublished).